



***SOBRE LA INTELIGENCIA ARTIFICIAL Y LA TRANSGRESIÓN DE DERECHOS
HUMANOS EN MÉXICO Y JAPÓN***

**UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD CUAJIMALPA
DIVISIÓN DE CIENCIAS SOCIALES Y HUMANIDADES**

**SOBRE LA INTELIGENCIA ARTIFICIAL Y LA TRANSGRESIÓN DE
DERECHOS HUMANOS EN MÉXICO Y JAPÓN**

TESIS DE DOCTORADO
PARA OPTAR POR EL GRADO DE
DOCTOR EN CIENCIAS SOCIALES Y HUMANIDADES

PRESENTA:

MTRO. JORGE LUIS MORTON GUTIÉRREZ

DIRECTORA DE TESIS

DRA. MARIANA MORANCHEL POCATERRA

MIEMBROS DEL COMITÉ TUTORAL

DR. JORGE LIONEL GALINDO MONTEAGUDO

DRA. ESTHER MORALES FRANCO

Ciudad de México, 2025.

**En memoria de mi gran amigo, el ingeniero Salvador López López
(1985-2023)**

“El valor no necesita ser recordado,
porque nunca se olvida”.
The Legend of Zelda, 2017.

AGRADECIMIENTOS

Esta tesis ha sido posible gracias al invaluable apoyo de mi directora, la Dra. Mariana Moranchel Pocaterra, cuya confianza en este proyecto desde sus inicios y su guía experta fueron fundamentales para su desarrollo. Extiendo mi profundo agradecimiento a los miembros de mi comité tutorial, la Dra. Esther Morales Franco y el Dr. Jorge Lionel Galindo Monteagudo, cuyos comentarios y aportes enriquecieron significativamente este trabajo. Asimismo, expreso mi gratitud a los miembros de mi jurado de examen de candidatura, el Dr. Carlos Toxtli Hernández y el Dr. Carlos Lucatero, cuyas contribuciones fueron de gran ayuda, así como por su participación en el jurado de mi examen de grado.

Además, es importante destacar las contribuciones directas e indirectas que han fortalecido esta investigación. Los comentarios de la Dra. Sofia Trejo, la motivación brindada por la Dra. Huma Shah de Coventry University, así como mi participación en el proyecto CSI-COP como científico ciudadano, han sido especialmente significativos y se reflejan en el desarrollo de esta investigación. Reconozco también el apoyo constante de la Dra. Saiph Savage, quien no solo recomendó a un destacado académico como miembro del jurado, sino que ha demostrado ser una persona brillante y dedicada en el campo de la IA.

En cuanto a otros académicos, amigos y personas que han contribuido a este trabajo, quiero expresar mi profundo agradecimiento al Dr. Bolaños, al Dr. Maximiliano Martínez y, en particular, a la Dra. Akuavi Adonón por su enorme apoyo, así como al Dr. Barbosa, en un momento crucial de mi vida, y al difunto Latour cuya inspiración colocó los cimientos de este trabajo. Agradezco especialmente a mis amigos y colegas, incluyendo a la Licenciada Moreno, a Agus Rivero, a la maestra Ani Frere, al maestro Daniel Romero y a la niña de la profecía, a David y a mi hermano Diego, cuyo apoyo y compañerismo han sido fundamentales en este viaje académico.

Quiero expresar mi profundo agradecimiento y destacar la paciencia de amiga y persona tolerante, la maestra Claudia Acosta. También agradezco a mis compañeros de semanario Erick, Paty y aquel compañero malvado imaginario que

espero algún día conoceremos. Y reconocer la contribución de mi primo Carlos Antonio que fue como una linterna en la oscuridad, y a mi prima Mónica Macarena.

En este sentido, debo destacar el apoyo incondicional de mi hermano en momentos difíciles, el ingeniero Carlos Mejía, quien estuvo a mi lado durante los momentos más complejos de mi vida. Y finalmente, quiero agradecer a mi hermano, amigo y persona valiente, quien nunca me abandonó incluso en mis peores momentos, el Ingeniero Salvador López López. Sé que nos encontraremos de nuevo cuando pida el deseo a la trifuerza; mientras tanto, venceré a Ganon tantas veces como sea necesario en tu honor.

Por último, deseo agradecerme a mí mismo por el impulso vital que me permitió lograr un objetivo fundamental: dedicarme a la investigación. Querido niño que siempre estuvo solo en el recreo, que enfrentó el acoso de tus compañeros, el desprecio de tus maestros y la humillación de tu familia, quiero decirte que cumplí nuestra promesa. Logré uno de tus sueños y convertir todo tu dolor en la tinta de una pluma, de la cual surgieron mil ideas. Ahora puedes descansar tranquilo, cuidé de ti como pude y plasmé melodías de mil sueños con nuestra voz. Así que esta noche, duerme con una sonrisa, porque hoy todo está bien.

Índice

Glosario de términos y abreviaturas	6
INTRODUCCIÓN	13
1. Problematicación	13
2. Planteamiento del problema	17
A. Pregunta general	21
B. Preguntas particulares	21
3. Hipótesis de trabajo	22
4. Justificación	23
A. ¿Por qué México y Japón?	26
I. El sistema jurídico mexicano	29
II. El sistema jurídico japonés	33
B. Un breve contraste sobre mirada tecnológica de Japón	41
C. Perspectivas sociales de México y Japón en torno a la IA	47
D. Los aspectos ético-culturales de México y Japón con respecto a la IA	56
E. La sociedad mexicana y japonesa en torno a la IA	59
5. Metodología	63
A. Análisis de constelaciones	64
B. Análisis de constelaciones y el objeto de contorno	66
C. Ruta metodológica	69
D. Los problemas sociales en México en Japón y la IA	71
CAPÍTULO PRIMERO	73
MARCO TEÓRICO: UNA REFLEXIÓN SOCIAL EN TORNO A LA IA	74
1. Inteligencia Artificial	76
2. Transgresión, sistemas y riesgos	78
3. La teoría de actor red: Latour	83
	6

4. El modelo de equivalencia: Boltanski y Thévenot	89
5. La sociología de la capacidad crítica y modelo de equivalencias	91
6. El principio de equivalencia	92
7. La discreta armonía entre Boltanski y Latour	100
8. Velíz, los zombis morales, y la responsabilidad en la IA	102
CAPÍTULO SEGUNDO	107
LA TRAYECTORIA DE LA IA	108
1. Tres canciones de una máquina pensante: Lovelace, Babbage y Turing	110
2. McCarthy, el padrino de la IA	118
3. De los inviernos de la IA a la primavera de los objetos	123
A. El perceptrón de Rosenblatt	123
B. El reporte Lighthill	126
C. Tercera parte, del segundo invierno a la reivindicación de Rosenblatt	132
D. Canadá, Hinton y al aprendizaje de máquina	134
E. Watson	138
F. AlphaGo	143
4. Del laboratorio al mundo cotidiano: IA en redes sociales y economía gig	147
A. TikTok	153
B. Facebook	157
C. Didi y el aprendizaje de máquina en la economía gig	161
CAPÍTULO TERCERO	164
TAXONOMÍA Y PROBLEMAS DE LA IA	165
1. El aprendizaje de máquina	166
A. Redes neuronales	169
B. Tipos de aprendizaje de máquina	173
C. Aprendizaje de máquina profundo	179

D. Aplicaciones del aprendizaje de máquina	182
E. Futuras aplicaciones	192
F. La caja negra	195
2. Los falsos profundos y los algoritmos adversarios	198
3. El aprendizaje de máquina y el trabajo	203
4. Facebook y el uso de los algoritmos para la prevención del suicidio	206
CAPÍTULO CUARTO	211
TRAYECTORIA DE LOS PROBLEMAS ÉTICOS DE LA IA Y SU RELACIÓN CON LOS DERECHOS HUMANOS	212
1. Problemas éticos en torno a la IA	218
A. Recomendación sobre la ética de la IA de la UNESCO	220
B. Estudio particular de principios éticos de la IA	225
I. El problema de la discriminación	227
II. Privacidad	232
III. Responsabilidad	234
2. La Taxonomía de riesgos de la IA: por una IA confiable	236
A. Más allá de los valores y regulación occidental de la IA	246
B. La Declaración Universal de los Derechos Humanos y la IA	249
3. Marco regulatorio de la Unión Europea en torno a la IA	257
4. La regulación de México en torno a la IA	267
A. El contexto de la IA en México	268
B. La regulación de la IA en México	274
5. Japón en torno a la IA	282
A. La sociedad 5.0	283
B. Plan de gobernanza de la IA en Japón versión 1.0 y 1.1	287
C. Los derechos humanos desde el sistema jurídico japonés	292

CAPÍTULO QUINTO	295
ESTUDIOS DE CASO: LA IA FRENTE AL ESPEJO DE MÉXICO Y JAPÓN	295
1. Controversias legislativas globales y locales	295
2. Padrón de datos personales de telefonía móvil: privacidad y derechos humanos	302
3. Los enfoques de I + D de la IA en México y Japón	310
4. Alphabet-Google: la IA y los derechos humanos en México y Japón	315
A. Política ética de la IA en Google	317
B. Alphabet y su ecosistema de recolección de datos: cookies y perfiles	322
5. Uber y los grandes problemas de la IA en México	342
A. Las controversias de Uber y Uber eats a nivel global	344
B. Políticas de privacidad de Uber	349
C. Uber en México	353
D. Uber en Japón	359
E. Uber y teoría de actor red y la teoría de las capacidades	362
6. Sistemas metropolitanos de reconocimiento facial metropolitano	366
A. Introducción a las problemáticas de los sistemas de reconocimiento facial	370
B. Japón: Muéstrame tu cara y paga	375
C. México: El C5 y la central de abastos	380
D. Reflexiones finales sobre los sistemas de reconocimiento facial metropolitanos en México y Japón desde la mirada de la teoría de actor red	398
CONCLUSIONES	405
1. Conclusión primera: ¿Cómo informar a la sociedad?	410

2. Conclusión segunda: ¿Cómo generar marcos de transparencia y privacidad en torno a los sistemas y algoritmos de IA en México y Japón?	414
3. Conclusión tercera: ¿Cómo armonizar la implementación y desarrollo de la IA más allá de un modelo para prevenir transgresiones?	419
Uso de la IA generativa en la tesis	423
FUENTES DE REFERENCIA	424
1. Bibliografía	425
2. Hemerografía	454

Glosario de términos y abreviaturas

A

Actante: “es algo que actúa sobre cómo esa actividad es otorgada por otro” (Latour).

AIFA: Aeropuerto Internacional Felipe Ángeles.

AMP: Aprendizaje de Máquina Profundo.

Animismo: Es la práctica donde se veneran a los espíritus de la naturaleza. Por ejemplo, las culturas que creen que los ríos tienen alma o que veneran a la madre

naturaleza. Por consiguiente, las religiones animistas le dan un carácter de agencia a los objetos, a las cosas; a consecuencia, consideran que todo tiene alma, en especial, aquello que forma parte de la naturaleza.

Anonimización: Procesos, herramientas y técnicas que protegen la identidad de las personas que proporcionan sus datos. Entre estas destacan herramientas para evitar procesos de reidentificación a partir de datos falsos o ruido en la base de datos, entre otras.

Análisis de constelación: Es una herramienta metodológica interdisciplinaria en la que se trazan los elementos más importantes para hacer una investigación centrada en la trayectoria de un objeto o fenómeno tecnológico. Dentro de la constelación se definen las áreas de conocimiento, los conceptos a tratar, y los casos empíricos.

API (Interfaz de Programación de Aplicaciones): Es un conjunto de reglas y protocolos que permite que diferentes programas interactúen entre sí.

Aprendizaje automático: Es una rama de la IA que se dedica al estudio y desarrollo de sistemas capaces de aprender de manera automática a partir de datos.

B

Big Data: Enormes cantidades de datos que se han producido gracias a los avances en las tecnologías de información y comunicación de 1990-2021...

Bot: Herramienta que emplea algoritmos matemáticos o sistemas de aprendizaje profundo. Puede servir para ayudar a quitar correo no deseado, asistentes virtuales, sugerencias de rutas, etc.

C

C2: Centro de Comando y Control.

C5: Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano.

CEDA: Central de Abasto.

Chip 86x: Es una arquitectura de procesadores especializada en hacer tareas de alta densidad y demanda. Este tipo de chips son los que tradicionalmente se han usado en las computadoras personales. Y fueron impulsados por la compañía Intel.

Chip ARM: Es una arquitectura de procesadores especializada en hacer tareas de baja densidad y demanda.

Ciencia de datos: es el término para describir un área interdisciplinaria donde se extraen, se analiza y emplean datos. Muchas de las aplicaciones de la ciencia de datos se aplican en ramas como el aprendizaje de datos, la IA, y la clasificación de datos relevantes de los no relevantes (los que producen ruido en cuando su importancia).

Co-bots: Robots diseñados para desempeñarse en tareas donde tienen que convivir con humanos: robots que comparten tareas en fábricas y oficinas, robots de compañía, robots como tutores, y robots para el cuidado de personas de la tercera edad, entre otros.

Crowdsourcing: cuando una plataforma funciona a partir de información alimentada por el público o los usuarios. Por ejemplo, reseñas de comida, ubicación en tiempo real, entre otras.

Cuasiobjeto: Un objeto que ejerce actividad sobre otro objeto.

D

Data Brokers: Son actores que se dedican a comprar y vender datos que se extraen principalmente de plataformas digitales.

Datos no estructurados: Datos que no están etiquetados o catalogados bajo un orden alfanumérico. Por ejemplo, se pueden considerar como datos no estructurados a las fotografías, archivos de voz y sonidos musicales.

Deep Fake: Es una técnica de IA que se utiliza para generar contenido multimedia falsificado que simula a una persona o un evento real.

E

Entendible: Es el principio que supone que todas las plataformas, algoritmos e insumos de la IA deben ser comprensibles para los seres humanos. Esto es debido a que cada vez las plataformas de la IA toman decisiones de forma tan compleja que hace casi imposible que incluso sus diseñadores sepan cómo llegaron a esos resultados.

Estocástico: algo que está sometido al azar.

G

GPU: Tarjetas de procesamiento gráfico usadas principalmente para aplicaciones de vídeo, pero con gran utilidad para realizar tareas de aprendizaje de máquina.

H

Hot-spot: Una zona geográfica o un grupo de personas que experimentan una mayor incidencia de un fenómeno o evento determinado.

I

I + D: Investigación y Desarrollo.

I/O: Internet of things o internet de las cosas. La capacidad de que cada vez más objetos que antes no contaban con capacidad de cómputo se conecten a internet como los refrigeradores, relojes inteligentes, drones, entre otros.

Imparcialidad: La IA no debe de discriminar. Por ello se deben crear mecanismos para evitar la discriminación aprendida; aquella influenciada por los datos con los que se entrena. Además, se debe evitar la discriminación basada en la toma de decisiones para llegar al mejor resultado: negarle un crédito a una persona que este pasa por un mal momento financiero con base en sus estados de cuenta.

INAI: Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales.

INEGI: Instituto Nacional de Estadística y Geografía.

Inteligencia Artificial (IA): Es una rama de la informática que se dedica al estudio y desarrollo de sistemas capaces de realizar tareas que normalmente requieren

inteligencia humana, como el aprendizaje automático, el reconocimiento de patrones y el razonamiento.

IA acotada o sesgada: solamente puede hacer una tarea o especializarse en una sola tarea.

IA general: aquella que puede hacer una variedad de tareas y que se acerca al nivel cognitivo de los seres humanos.

J

JDI: Japan Display Inc. (Empresa japonesa especializada en pantallas, involucrada en tecnologías biométricas).

JR: Japan Railways (Ferrocarriles de Japón).

JR Central: Central Japan Railway Company (Compañía de Ferrocarriles del Centro de Japón)

JR East: East Japan Railway Company (Compañía de Ferrocarriles del Este de Japón).

L

LLM (Large Language Model): Los modelos grandes de lenguaje son redes neuronales artificiales diseñadas para imitar la comunicación humana mediante el procesamiento de enormes volúmenes de datos textuales. Actualmente, existen diversas versiones de estos modelos, muchas de ellas enfocadas en mejorar el razonamiento y la generación de respuestas cada vez más coherentes y sofisticadas.

M

Machine Learning: Es un conjunto de técnicas y algoritmos que permiten a las computadoras aprender automáticamente a partir de datos, sin que sean necesarios programas específicos para cada tarea a realizar.

Metadatos: Son los datos de los datos. En otras palabras, es la información que ayuda a identificar algún tipo de rasgo dentro de la gran masa de información.

METI: Ministry of Economy, Trade and Industry (Ministerio de Economía, Comercio e Industria).

MIC: Ministry of Internal Affairs and Communications (Ministerio de Asuntos Internos y Comunicaciones).

Modelo de difusión: Es una herramienta matemática que se utiliza para describir y predecir cómo una idea, un producto o una tecnología se propaga a través de una población.

Modelo generativo: Cuando la máquina puede hacer tareas como escribir música, escribir artículos periodísticos y pintar imágenes digitales a partir de puntos de referencia (Gillies, en Coursera, s.f).

MOJ: Ministry of Justice (Ministerio de Justicia, encargado de la regulación de datos biométricos en inmigración y seguridad).

N

Neurona convolucional: Se da cuando el resultado de un filtro crea otro filtro. Ese nuevo filtro se alimenta a otro filtro, a otra variable, lo que crea características más complejas. “Una neurona convolucional es básicamente un filtro de Photoshop que transforma una imagen” (Gillies, en Coursera, s.f). Otro ejemplo son los filtros de imágenes populares entre las aplicaciones.

Neurona artificial: Es una unidad de una red neuronal. Es decir, un elemento dentro de una función algorítmica. Por ejemplo, la edad de alguien, la forma de una nariz, entre otros.

NISC: National center of Incident readiness and Strategy for Cybersecurity (Centro Nacional de Preparación y Estrategia para la Ciberseguridad).

NPA: National Police Agency (Agencia Nacional de Policía).

O

Objeto de contorno: “Un objeto de contorno es un artefacto, documento o incluso una idea que ayuda a personas de diferentes comunidades a construir un entendimiento compartido” (Winget, 2008).

R

Razonamiento: Es el proceso mental mediante el cual se llega a una conclusión a partir de un conjunto de premisas.

Reconocimiento de patrones: Es una técnica de IA que se utiliza para analizar y clasificar datos a partir de patrones previamente identificados.

Redes naturales de lenguaje: Son sistemas de procesamiento de lenguaje que se basan en modelos matemáticos para entender y generar textos escritos en lenguaje natural.

Responsabilidad: Las plataformas deben de hacer un uso responsable de los datos. Asimismo, deben de tomar decisiones sin afectar a los usuarios en aras de mejorar un servicio para darle ventaja competitiva a las compañías. Por ejemplo, una empresa debe evitar problemas relacionados con la creación de perfiles de usuarios para que estos sean usados de manera nociva; hay que destacar que la IA ya puede deducir rasgos como la orientación sexual, preferencia política y estado emocional.

Rinri: Es la palabra que se usa para describir la ética japonesa (Kitano, 2007). El Rinri se basa en el principio de la armonía social. Esa armonía social incluye la relación de las personas con los objetos.

S

Sedena: Secretaría de la Defensa Nacional.

Shinto: Tradición espiritual y sistema de creencias japonés basado en la veneración de los kami (espíritus o deidades), la relación armoniosa con la naturaleza y los rituales de purificación.

T

TIC: Tecnologías de la Información y la Comunicación.

Transparencia: Todo modelo, plataforma y algoritmo debe de ser entendible. Para ello, los usuarios deben de conocer grosso modo la forma en que la IA toma decisiones y cómo usa su información y datos.

Valle inquietante: El valle inquietante es el momento en que la ilusión de familiaridad de un modelo no humano se rompe: Una muñeca con forma de niña se convierte en un objeto que nos mira con una forma humana pero no humana, una voz sintetizada suena tenebrosa y amenazante, y un robot humanoide crea un espejo fantasmagórico que se asemeja a lo humano pero que no actúa como humano.

U

UAM: Universidad Autónoma Metropolitana.

UNAM: Universidad Nacional Autónoma de México.

W

Wa: Concepto japonés que representa la armonía social, la cooperación y el equilibrio dentro de un grupo, priorizando el consenso y la estabilidad sobre el conflicto o la confrontación.

É

Ética en la IA: Son los principios generales para hacer uso y desarrollo responsable de la IA. Estos principios han sido planteados por autores y desarrolladores como el Doctor Ng y la Doctora van Dijck (aplicados especialmente a las plataformas).

新聞 **Shinbun:** Periódico o diario japonés.

INTRODUCCIÓN

1. Problematicación

La Inteligencia Artificial (IA) está modificando de manera acelerada la sociedad, impactando la manera en que nos relacionamos, tomamos decisiones e incluso en cómo se establecen y regulan los valores. La Inteligencia Artificial no solo influye en la configuración de valores, sino que también genera fenómenos que los transforman directamente. Esto es evidente en desarrollos recientes, como los grandes modelos de lenguaje (LLM por sus siglas en inglés), que impactan las emociones de los usuarios (Morton, 2022), y en la forma en que los algoritmos de la economía de plataformas moldean el comportamiento de sus trabajadores mediante prácticas de vigilancia, supervisión y presión (Morton, 2020). Por lo tanto, esta tesis busca analizar los efectos sociales de la implementación de la IA de Alphabet-Google, Uber y los sistemas de videovigilancia (C5) en relación con

los derechos de privacidad, acceso a la información y no discriminación en México y Japón desde una perspectiva, así como los problemas de *explicabilidad*, rendición de cuentas, responsabilidad y transparencia.

La IA es una herramienta cada vez más presente en nuestra vida cotidiana, y con ella surgen nuevos desafíos éticos y sociales. En este sentido, esta tesis busca analizar los efectos sociales de la implementación de la IA de Alphabet-Google, Uber y los sistemas de videovigilancia (C5) en relación con las nociones de privacidad, acceso a la información y no discriminación en México y Japón. El objetivo es comprender cómo estas tecnologías están transgrediendo la sociedad y sus valores, y cómo debemos abordar estos cambios de manera ética y responsable.

El presente estudio se realizará desde una perspectiva de derechos humanos y de los principios éticos que deben prevalecer en todas las etapas del ciclo de vida de la IA. Aunque no existe una definición universalmente aceptada de inteligencia artificial (IA) debido a su naturaleza dinámica y en constante evolución, es posible articular un concepto organizador a partir de sus elementos de innovación y desarrollo. En este sentido, la inteligencia artificial se puede definir como un área de la ciencia de la computación que busca imitar capacidades humanas como el razonamiento, el reconocimiento de imágenes y el procesamiento del lenguaje, entre otras.; sin embargo, en la actualidad, su enfoque principal es la creación de sistemas computacionales capaces de automatizar tareas y procesos, especialmente aquellos repetitivos, mediante el análisis avanzado de datos.

Ahora bien, es fundamental aclarar un aspecto clave en torno a la definición operativa de la inteligencia artificial: la distinción entre modelos y sistemas de IA. Aunque los modelos de inteligencia artificial son componentes esenciales de los sistemas, no constituyen por sí mismos sistemas completos (E Silva, 2024). Para funcionar como tales, requieren integrarse con otros elementos, como interfaces de usuario, dispositivos de despliegue, infraestructuras de datos y algoritmos complementarios, que conforman una red dinámica de actores interconectados.

Un modelo de IA, por ejemplo, puede ser Chat GPT, BERT, DALL-E u otros sistemas de procesamiento del lenguaje natural o generación de imágenes, que actúan como "motores" dentro de una red más amplia. Estos modelos no operan de forma aislada, sino que se entrelazan con otros actores —humanos y no humanos— para desplegar su funcionalidad. Desde la perspectiva de la teoría del actor-red, el modelo es un nodo clave dentro de esta red, donde elementos como interfaces, datos, usuarios, dispositivos e incluso normativas sociotécnicas configuran las conexiones necesarias que dan vida operativa y significado al sistema en un contexto específico.

Así, mientras que un modelo de IA puede entenderse como el "motor" detrás de aplicaciones como el reconocimiento facial en cámaras o los sistemas de recomendación en plataformas como Amazon, un sistema de IA es la red completa de asociaciones entre modelos, actores humanos y objetos tecnológicos. Esta red permite que la IA realice tareas específicas, como identificar patrones en grandes volúmenes de datos, o multitareas si se integran múltiples modelos especializados. En resumen, la inteligencia artificial no es un ente autónomo, sino un conjunto de relaciones mediadas por tecnología, contexto y actores sociales, donde cada componente contribuye a la funcionalidad global del sistema.

Asimismo, es evidente la necesidad de la IA. Esta regulación debe garantizar los derechos humanos, específicamente el bienestar y la dignidad de las personas, en consonancia con valores y principios éticos. Por lo que estos principios deben incorporarse en las distintas plataformas que utilizan la IA para garantizar que los intereses y derechos de todas las partes interactuantes sean protegidos.

Específicamente, se considerarán los problemas de transparencia *explicabilidad*¹, entendidos como “la inteligibilidad de la entrada, salida y funcionamiento de cada componente algorítmico y cómo contribuye a los resultados de los sistemas”

¹ De acuerdo con la UNESCO (2021) en Zárate Alcarazo: “la explicabilidad «hace referencia al hacer inteligible los resultados de los sistemas de IA. La XAI también hace referencia a la comprensibilidad de los datos, procesos y comportamientos de los distintos bloques algorítmicos y como cada uno de ellos contribuye al resultado del sistema. Así, la explicabilidad está estrechamente relacionada con la transparencia, ya que los procesos y sub-procesos que conducen a los resultados deberían ser comprensibles y trazables, apropiados para el contexto»

(UNESCO, 2022), así como la rendición de cuentas, la responsabilidad, y el derecho a la privacidad y protección de datos. Estos principios son fundamentales para garantizar que la IA se utilice de manera ética y para proteger los derechos humanos. Todo esto se realizará en el contexto del método de análisis de constelaciones y un enfoque empírico para investigar cómo estas plataformas pueden vulnerar los derechos humanos.

En este sentido, se considera importante mencionar que el estudio se llevará a cabo en torno a México y Japón, ya que ambos países tienen un contexto cultural y social distinto, lo que permite analizar cómo la implementación de la IA puede variar en función de estos factores. La investigación se basará en un enfoque cualitativo y cuantitativo, para recopilar información mediante la revisión de documentos y estudios existentes; y en el caso de Google, como plataforma utilizada en ambos países, se analizarán sitios de internet. Este enfoque metodológico permitirá obtener una visión más completa de los efectos sociales de la IA en México y Japón.

Algunos autores mencionan que la IA no es un agente moral, es decir, un actor que surge a partir de una red de asociaciones entre actores humanos y no humanos, y de donde adquiere sesgos en torno a estas asociaciones; no obstante, como actor, o actante, la IA empuja programas y crea anti programas. Pero como plantean Boltanski y Thevenot, los actores humanos sí son agentes que tienen la capacidad de demostrar resistencia; en términos de Latour, los actores humanos resisten ya sea un programa o antiprograma al crear nuevos actores dentro de la red de asociaciones. Estos programas son un resultado deseable: generar texto con base en una entrada, identificar un tumor, evaluar profesores, entre otros; y los antiprogramas son los resultados no deseados como el no respeto a la privacidad, la discriminación y la opacidad en los sistemas.

De igual forma, la presente investigación analizará los efectos sociales de la implementación de la IA de Alphabet-Google, Uber y los sistemas de videovigilancia (C5) en relación con las transgresiones a los derechos humanos de privacidad, acceso a la información y derecho a no ser discriminado en México y

Japón. El objetivo es comprender cómo estas tecnologías están violando derechos humanos y cómo se debe abordar esta problemática desde una perspectiva ética y responsable. Conforme la evolución de la tecnología y sus herramientas, surgen problemas éticos sobre el uso y la protección de los datos personales, ya que prácticamente toda la información la recolectan y usan empresas privadas, cuyo control con los marcos regulatorios existentes es poco eficaz.

Es importante destacar algunas problemáticas identificadas en el desarrollo de la presente investigación, especialmente en relación con las respuestas proporcionadas por el Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la CDMX (C5) relativas al uso de herramientas de IA en sus cámaras. El C5 no ofreció una claridad adecuada sobre qué entidad custodia la información y hubo grandes dificultades para obtener información sobre sus proveedores. En vista de esta falta de transparencia en la respuesta de esta autoridad de la Ciudad de México, nuevamente, el 22 de marzo del 2024 se presentó una nueva solicitud de información para abordar diversas temáticas que no se habían respondido en la primera solicitud.

A través del portal de transparencia, presenté una solicitud de información detallada sobre la implementación de sistemas de reconocimiento facial e IA en las cámaras de vigilancia operadas por el C5 en la Ciudad de México. Esta solicitud abarco diversos aspectos, entre los que se incluyeron, entre otros, las siguientes cuestiones:

- Descripción de la tecnología y software utilizados.
- Protocolos de privacidad y protección de datos personales.
- Estadísticas de eficacia y errores en la identificación de individuos.
- Medidas de supervisión y regulación para prevenir abusos.
- Información sobre la cooperación con otras entidades gubernamentales o privadas en relación con estos sistemas.

2. Planteamiento del problema

Los sistemas de inteligencia artificial han transformado significativamente la vida cotidiana, con un impacto que va más allá de las advertencias de figuras públicas como Elon Musk, quien ha señalado a la IA como un riesgo para la humanidad (Clifford, 2018). Más allá de estos pronunciamientos, el debate central radica en cómo la IA ha permeado y alterado las dinámicas diarias de las personas. Antes de 2021, la IA era un concepto distante para muchos; sin embargo, en solo tres años, su presencia se ha expandido a casi todos los aspectos de la vida de quienes utilizan un teléfono inteligente (Furnham, 2019).

En este sentido, como plantea Furnham (2019), es necesario considerar algunos aspectos clave de la inteligencia artificial (IA) y los conceptos fundamentales que siguen vigentes. Así pues, la IA puede definirse como un campo que abarca diversos enfoques y técnicas, desde algoritmos tradicionales hasta modelos de aprendizaje automático capaces de mejorar a partir de datos sin intervención humana constante. Además, su impacto en la vida cotidiana es cada vez más evidente, con aplicaciones en sectores como la medicina, las finanzas y la tecnología. No obstante, su desarrollo plantea importantes desafíos éticos, incluyendo preocupaciones sobre la privacidad, el empleo y la toma de decisiones autónoma. Por ello, es fundamental reflexionar sobre el futuro de la IA y su potencial para transformar la sociedad.

Además, el auge de herramientas como Chat GPT, que se popularizó a finales de 2022, reveló al público general la capacidad transformadora de la IA. Además, la denominada “carrera de la IA” entre Microsoft y OpenAI contra Alphabet (Google), junto con el desarrollo de modelos más ligeros como DeepSeek en China, ha acelerado el despliegue de grandes modelos de lenguaje (LLMs) sin los contrapesos regulatorios ni la autogestión ética por parte de estas empresas. Esta rápida expansión ha generado preocupaciones incluso entre pioneros del campo, como Geoffrey Hinton, una de las figuras clave en el desarrollo de redes neuronales artificiales, quien en 2023 renunció a Alphabet (Google) y advirtió: “Esta tecnología tiene un enorme potencial, pero es fundamental que el mundo

invierta de manera urgente en su seguridad y control” (Kleinman & Vallance, 2023).

Esto plantea desafíos, ya que la IA ya se utiliza en sistemas de monitoreo y vigilancia por su capacidad para identificar y distinguir rostros humanos (Rosenberg, 2018). Además, la IA se utiliza para estudiar y aprender sobre nuestros hábitos de consumo y para vender y crear productos mejores basados en nuestras preferencias individuales y colectivas (Bughin et al., 2018). Y en algunos casos, que cada vez son más comunes, la IA comienza a mediar nuestros procesos sociales (Singer, 2018).

La implementación de la IA en la vida cotidiana plantea desafíos éticos importantes relacionados con los derechos humanos, especialmente en lo que respecta a la transparencia y *explicabilidad*, rendición de cuentas y responsabilidad, y derecho a la intimidad y protección de datos personales en cuanto a cómo estas tecnologías están tomando decisiones y actuando. Por lo tanto, es crucial analizar cómo la IA afecta estos derechos en diferentes contextos culturales y sociales. En este sentido, esta tesis busca realizar un análisis comparativo entre México y Japón para comprender cómo la IA se regula y se construyen valores éticos en relación con sus sistemas regulatorios y los derechos humanos. El objetivo es proponer marcos jurídicos adecuados que aborden estos problemas éticos y garanticen que la IA se utilice de manera responsable y respete los derechos humanos.

Además, aunque no sea el centro de atención de este trabajo, es importante mencionar que los valores con los que se despliegan las políticas regulatorias pueden ser influenciados por su contexto social. Esto es especialmente relevante en los países del Pacífico de Asia, como China, Corea y Japón, donde las nociones de ética y la relación con herramientas y tecnologías creadas por humanos son distintas a las del occidente (Kitano, 2017). Por ejemplo, en los productos culturales de estos países (cine, teatro, cómics, entre otros) se ve a la máquina como algo que forma parte de la sociedad, por lo que se podría pensar que la construcción de políticas públicas en torno a la ética y el desarrollo de la IA

en esas naciones se basa en la trayectoria de objetos de contorno y los programas y antiprogramas que estos encierran.

Así pues, esta tesis tiene como objetivo analizar el impacto social de la IA en México y Japón, considerando las posibles violaciones a los derechos humanos de privacidad, acceso a la información y no discriminación. Para ello, se empleará el método de análisis de constelaciones, que permite explorar cuatro trayectorias clave: la trayectoria de la IA, la trayectoria de la declaración de los derechos humanos y las leyes sobre propuestas regulatorias de la IA en la Unión Europea, y las trayectorias específicas de México y Japón en cuanto a sus leyes sobre derechos humanos y propuestas regulatorias sobre IA. Además, se analizarán casos empíricos del impacto social de esta tecnología en ambos países. Se adoptará una perspectiva de derechos humanos, ya que tanto México como Japón tienen códigos jurídicos que se basan en estos derechos.

Para ello, se utilizará el método de análisis de constelaciones, ya que permite rastrear las interacciones entre actantes (actores humanos y no humanos), conceptos y trayectorias tecnológicas, analizando sus asociaciones e impactos a través de una red de conexiones dinámicas. Este enfoque se enmarca en la teoría del actor-red, que constituye la base del análisis de constelaciones y que, como plantea Latour, es más una “asociología” que una sociología convencional, ya que se centra en el papel de los objetos dentro de los entornos tecnológicos. Finalmente, este marco se complementará, de manera mesurada, con el modelo de equivalencias de Callon, que permitirá establecer relaciones comparativas entre distintos elementos de la red, facilitando una comprensión más profunda de las transformaciones y dinámicas que emergen en torno a la inteligencia artificial.

De esta manera, se analizará la trayectoria de las propuestas regulatorias sobre la IA en México y Japón, como una constelación que integra los análisis de las tres trayectorias anteriores. La tesis tendrá un enfoque retrospectivo, al estudiar las trayectorias de la IA, la ética y la ética artificial, y un enfoque prospectivo, al proponer regulaciones en México para abordar los problemas sociales

identificados en el análisis de los casos empíricos: Uber y el sistema de videovigilancia C5.

Ahora bien, antes de analizar las trayectorias, hay que incluir un capítulo que explique los conceptos técnicos del lenguaje utilizado en la IA de manera accesible para las ciencias sociales y las humanidades. Luego, se analizará el contenido de los marcos regulatorios en torno a la IA en México y Japón utilizando el método de análisis de constelaciones junto con la teoría de actores red y el objeto de contorno, así como la relación de la población con respecto a esta tecnología, con el fin de conocer el impacto social de la IA en la sociedad de ambos países y los anti programas que han desarrollado para hacer frente a estos problemas a través de la regulación ética.

A. Pregunta general

¿En qué medida los sistemas de IA en plataformas digitales de Alphabet-Google, Uber y los sistemas de videovigilancia (C5) que utilizan IA en México y Japón podrían transgredir los *derechos humanos* relacionados con la *privacidad y acceso a la información*?

- ¿Cómo operan los sistemas de inteligencia artificial de Alphabet-Google, Uber y los sistemas de videovigilancia C5 en la toma de decisiones y en la gestión de datos?
- ¿Qué diferencias y similitudes existen en la forma en que estas plataformas operan y son reguladas en México y Japón?

B. Preguntas particulares

1. Aplicación del marco teórico al análisis de IA

- a. ¿Cómo pueden aplicarse la teoría del actor-red de Latour y la teoría de la equivalencia de Boltanski al estudio de la IA y su relación con los problemas sociales derivados de su desarrollo?

- i. ¿De qué manera los conceptos de programa y antiprograma permiten analizar los conflictos entre los objetivos de las plataformas digitales y los intereses de los usuarios y reguladores en el uso de la IA?
- ii. ¿Cómo la noción de actante ayuda a comprender el papel de los sistemas de IA como agentes que configuran prácticas sociales, laborales y de vigilancia en plataformas digitales y sistemas de reconocimiento facial?
- iii. ¿Qué rol juega la traducción en la forma en que la IA reconfigura las relaciones de poder y las responsabilidades entre los distintos actores en México y Japón?

2. Trayectoria e impacto social de la IA

- a. ¿Cuál ha sido la trayectoria de la IA y qué impactos específicos ha generado en ámbitos como el trabajo, la privacidad y la seguridad en México y Japón?

3. Problemáticas éticas y efectos sociales

- a. ¿Cuáles son los principales dilemas éticos asociados al desarrollo y despliegue de la IA, y cómo han repercutido en la vida cotidiana, las relaciones laborales y la gobernanza digital?

4. Regulación existente

- a. ¿Qué leyes y propuestas regulatorias han sido desarrolladas en México y Japón para abordar los riesgos de la IA en plataformas digitales y sistemas de videovigilancia con reconocimiento facial?

i. Posibles iniciativas y regulaciones futuras:

- 1. ¿Qué otras propuestas regulatorias o iniciativas podrían desarrollarse para mitigar los riesgos de la IA? Por ejemplo, programas de alfabetización digital para que las personas protejan sus datos personales o modelos de IA diseñados para prevenir la vigilancia algorítmica en plataformas digitales de trabajo.

3. Hipótesis de trabajo

A partir del estudio comparativo de las plataformas Alphabet-Google, Uber y los sistemas de reconocimiento facial en áreas metropolitanas de México y Japón, se plantea que los riesgos que la inteligencia artificial representa para los derechos humanos de privacidad, acceso a la información, transparencia y no discriminación varían en función de la interacción entre factores tecnológicos (como la automatización y los sesgos algorítmicos), regulatorios (como la existencia o ausencia de normativas específicas) y socioculturales (como las percepciones y resistencias hacia la IA).

Asimismo, desde la teoría del actor-red, estos sistemas de IA pueden analizarse como actantes que reconfiguran las relaciones de poder en estos entornos, donde los conceptos de programa, antiprograma y traducción permiten entender los conflictos y adaptaciones entre reguladores, plataformas y usuarios.

4. Justificación

La IA puede vulnerar diversos derechos humanos de privacidad, acceso a la información y no discriminación. Los sistemas o plataformas con IA pueden recopilar, analizar y comercializar tus datos sin siempre explicar claramente cómo lo hacen. Además, pueden adquirir sesgos a partir de los datos con los que se entrenan o del propio diseño de la plataforma, lo que puede llevar a la discriminación por género, tono de piel y preferencia e identidad sexual. Un ejemplo documentado de esto es el sistema de contratación de personal de Amazon, que favorecía a los candidatos hombres y discriminaba a las mujeres (Trejo et al, 2020).

Ahora bien, es necesario responder algunas preguntas respecto al alcance de este trabajo con respecto a la pertinencia de la teoría, la relevancia social y la factibilidad. En cuanto a la relevancia teórica y metodológica, los elementos en torno a la teoría de la red de actores y el método de análisis de constelaciones

ayudan a crear una perspectiva de análisis interdisciplinario, agrupando las complejas trayectorias disciplinarias y casos de estudio en torno a constelaciones que convergen alrededor de un corpus de análisis más trazable. Esto es útil, sobre todo cuando se trata de un problema de estudio tan complejo como la IA y los derechos humanos.

En este sentido, la teoría de actor-red y sus diferentes herramientas y conceptos permiten identificar tantos actantes, integrantes humanos y no humanos de una red de asociaciones; mapear el rol de estos actantes dentro de la red; y analizar el programa o la intención de sus miembros. A su vez, esta perspectiva posibilita reconocer la existencia de anti programas, es decir, resistencias o dinámicas que desafían las intenciones iniciales de los sistemas de IA y de sus diseñadores. Finalmente, el concepto de traducción permite rastrear cómo estos programas y anti programas se modifican en función de interacciones tecnológicas, regulatorias y socioculturales, dando lugar a nuevas configuraciones en el uso y regulación de la IA.

Además, es importante destacar cómo las constelaciones de actores y las trayectorias disciplinares convergen en este análisis, permitiendo una visión más completa y trazable de los desafíos planteados por la IA en el contexto de los derechos humanos. Con esto, se permite ver los momentos de tensión e inestabilidad dentro de la red, donde los diferentes actantes—humanos y no humanos—negocian, resisten o refuerzan ciertos programas de acción. Estos momentos de tensión revelan los puntos críticos en los que las decisiones tecnológicas, regulatorias y socioculturales configuran el impacto de la IA en derechos como la privacidad, la transparencia y la no discriminación; no obstante, también permiten identificar momentos críticos intervenciones que mitiguen los riesgos asociados a estas tecnologías.

Además, realizar esta tesis fue factible. Se aprovechó el tiempo destinado por el posgrado para examinar a profundidad las referencias teóricas y la parte empírica de esta investigación. Aunque el objeto de este estudio está en constante cambio por los avances de la IA, se pudo generar un entorno que permitió estudiar y

actualizarse respecto a los aspectos más importantes y pertinentes de la IA para esta tesis.

Con respecto a la privacidad, esta puede estar en peligro debido a las personas que utilizan directa o indirectamente las plataformas con IA. Estas plataformas necesitan un flujo constante de datos para funcionar y para generar productos viables a nivel comercial mediante la intervención de los llamados *data brokers* o comerciantes de datos. Además, existe el constante riesgo de que los mecanismos de seguridad de la plataforma sean vulnerados por *hackeos*, lo que expondría la privacidad de las personas.

La privacidad es un derecho fundamental que puede verse amenazado por la IA, ya que esta tecnología requiere de grandes cantidades de datos para funcionar (ONU, 1948). Esto puede facilitar la intervención de intermediarios o comerciantes de datos, así como generar vulnerabilidades en los sistemas de seguridad que pueden exponer la información personal de las personas (Constitución Mexicana, 1917). Tanto la Declaración de Derechos Humanos de la ONU (ONU, 1948) como la Constitución Mexicana (Constitución Mexicana, 1917) reconocen la privacidad como un derecho humano. Lo mismo ocurre con la legislación japonesa, que incluye la privacidad en sus propuestas regulatorias sobre la IA (Japón, 2020).

El derecho a acceso a la información también se reconoce en el corpus jurídico de los derechos humanos en México (Comisión Nacional de los Derechos Humanos, 2020). Según la Comisión Nacional de los Derechos Humanos (CNDH), “toda persona, sin necesidad de acreditar interés alguno o justificar su utilización, tendrá acceso gratuito a la información pública, a sus datos personales o a la rectificación de éstos” (CNDH, 2020). Este derecho es fundamental para conocer cómo operan y funcionan las plataformas de IA y para entender por qué pueden transgredir los derechos humanos, ya que se necesita conocer los datos con los que se han entrenado los sistemas de IA, entre otros aspectos. Sin embargo, para el caso de la información y los datos que manejan las empresas privadas que funcionan con plataformas IA, no existe obligación de transparentar su funcionamiento.

Esto tiene enorme relación con el valor de explicabilidad, Van Dijck et al (2018) argumentan que la explicabilidad es importante en el contexto de la asimetría de información entre empresas de alta tecnología y consumidores. Según los autores, "con tanta asimetría de información, se puede argumentar que las empresas de alta tecnología les deben a los consumidores una explicación de cómo funciona el esquema de monetización de una plataforma. Su oscuridad dificulta seriamente la comprensión de los servicios en línea por parte del consumidor...". Por lo tanto, los autores también sugieren que el flujo de datos debe ser transparente para que los ciudadanos que utilizan las plataformas y sistemas de IA entiendan quién es el dueño de los datos, quién los usa y para qué los usa.

Hay que resaltar que la Guía de Interacción Humana con la IA de Microsoft (2019; 2024) señala la importancia de explicar cómo funcionan las plataformas de IA. Según la guía, se deben seguir dos pasos: primero, se debe explicar qué es lo que el sistema puede hacer; segundo, se debe dejar claro cómo el sistema puede hacer lo que hace y explicar que el sistema puede cometer errores. Así pues, no sólo se debe velar por el acceso a la información y también por un tipo de derecho a explicar cómo funcionan los sistemas que usan esa información; además, se deben de dar pasos importantes en hacer accesible la investigación y el desarrollo de la inteligencia artificial es un paso crucial para diversificar el campo (Responsible AI Transparency Report How we build, support our customers, and grow Contents, 2024).

Es importante señalar que el problema de la responsabilidad en la IA es identificar quién tiene la responsabilidad de los problemas sociales causados por la IA. Este es un problema complejo, ya que muchos actores humanos y no humanos participan en la toma de decisiones de la IA y de los sistemas y plataformas que la utilizan. Según Van Dijck et al (2018), "una sociedad que se aferra a los principios de igualdad, inclusión y trato justo sin favoritismo ni discriminación debe poder juzgar las decisiones automatizadas tomadas por algoritmos que ahora están ocultos al escrutinio público".

Así, hay que saber quiénes son los responsables de los problemas sociales de la IA. Y también promover y exigir sistemas con esta tecnología que sean auditables y entendibles, no solamente por expertos, sino también por usuarios promedio. Y, por consiguiente, se deben de examinar los problemas sociales que estas plataformas han ocasionado en escenarios jurídicos que funcionan como espejos en torno a México y Japón.

A. ¿Por qué México y Japón?

Hay que examinar el impacto de la IA en diferentes contextos geográficos y culturales. En este sentido, Japón y México son países interesantes de analizar por varias razones (Ohmae, 1986; Okazaki, 2020; Valenzuela, 2006):

1. Las características socioculturales de Japón pueden ser relevantes para entender cómo la IA puede impactar a la sociedad mexicana y cómo esta puede responder regulatoriamente. Algunos ejemplos de estas características son la influencia de la cosmovisión japonesa en la percepción de la tecnología (Okazaki, 2020) y la forma en que las personas se relacionan entre sí y resuelven disputas jurídicas.
2. Japón y México tienen relaciones comerciales y sociales relevantes que se han fortalecido en los últimos 20 años (Valenzuela, 2006). Esto hace que sea relevante analizar cómo la IA puede afectar a estos vínculos y cómo pueden colaborar ambos países en el desarrollo y uso responsable de esta tecnología.
3. Japón enfrenta desafíos específicos que pueden ser útiles para comprender cómo la IA puede afectar a una sociedad y cómo esta puede responder a estos retos (Ohmae, 1986). Algunos de estos desafíos son una población envejecida, una política migratoria restrictiva y una economía estancada durante las últimas tres décadas. Al comparar cómo Japón aborda estos problemas con la IA, podemos obtener lecciones útiles para entender cómo México puede enfrentar sus propios problemas relacionados con la IA.

A nivel jurídico y legislativo, Japón ha tenido un papel destacado en la región asiática y ha servido como un espejo para occidente (Tsunoda, 1951). Esto se debe a que el país ha enfrentado la necesidad de modernizarse rápidamente para evitar ser colonizado por otros países durante el siglo XIX (Perry, 1853). Bajo la amenaza de abrir sus fronteras y perecer, un grupo de bushi o samurai decidió poner fin al Edo Bakufu, un sistema que había brindado paz y estabilidad al país durante dos siglos, pero que también había impedido el desarrollo de la sociedad japonesa (Tsunoda, 1951).

Como resultado, Japón comenzó un proceso de modernización que ha durado dos siglos y que ha llevado a un cambio profundo en el orden social y jurídico del país, aunque siempre manteniendo los valores y tradiciones de la cultura japonesa (Tsunoda, 1951). Y por ello, el periodo romántico del *bushi* o de los samurai se resquebrajó y comenzó una etapa a lo largo de dos siglos donde el orden social y jurídico cambió con un paso firme hacia la modernidad; aunque sobre las sandalias de la tradición. Lo cual se refleja en su orden jurídico y su relación con la tecnología.

A nivel tecnológico, Japón ha tenido un papel protagonista en la región asiática gracias a su proceso de industrialización acelerado y a la negociación de complejos y a veces injustos tratados comerciales con países como Estados Unidos (Fujimoto, 2018). Bajo la figura del emperador Meiji, el país comenzó también un proceso de centralización en torno a modelos políticos y jurídicos occidentales (Kornicki, 1995). Los intelectuales Meiji crearon neologismos como *jiyu* y *kenri* para describir conceptos políticos occidentales como libertad y derechos, inspirados en el modelo de desarrollo de Hegel (Kornicki, 1995).

En cuanto a la tradición jurídica, el gobierno decidió adoptar una mezcla de modelos francés y prusiano como base para su primera constitución (Harding, 2003). Tras el fracaso de una visión imperialista-militarista y colonialista que terminó en su derrota en la Segunda Guerra Mundial, Japón adoptó elementos del sistema jurídico de Estados Unidos (Harding, 2003). A pesar de los cambios históricos y políticos, Japón ha seguido abrazando la tecnología como elemento

fundamental de su desarrollo y como medio para superar los desafíos del pasado (Fujimoto, 2018)

Hay que resaltar el rol que la tecnología ha tenido desde la transición Meiji. Pero sin dejar de mencionar que este rol estuvo cargado de una gran ideología militarista. De acuerdo con Kitano (2019): Sobre la transición de Tokugawa a Meiji, hubo dos consignas políticas; bunmei-kaika (iluminación de la civilización, es decir, a Japón, occidentalización), y fukoku-kyohei (nación rica, ejército fuerte). En este sentido, a lo largo de este periodo de transición y de apertura, el país se convirtió rápidamente en una potencia económica y militar y tuvo que mirar y aprender del occidente para sobrevivir. Lo que significó un proceso de modernización y de occidentalización (más tarde de globalismo) no derivado de ser una colonia, sino para evitar convertirse en una colonia.

No obstante, los cambios tecnológicos y sociales de Japón están marcados por particulares sociales ligados a tradiciones y estilos de vida que todavía persisten. Y hoy en día, mientras el país enfrenta un estancamiento financiero, una sociedad envejecida, y una bandera tecnológica que buscan recuperar; su relación particular con el derecho, la sociedad y la tecnología sirve como espejo ante un país como México que necesita reconciliar rápidamente su relación con respecto a los impactos de una tecnología como la IA que cada día evoluciona mientras extiende sus raíces a cada aspecto de la vida de sus habitantes.

Esto deriva en otro elemento para justificar el comparativo entre México y Japón. Como plantean Jecker y Nakazawa (2022), y McStay (2021), el sistema de regulación y de valores para la correcta implementación y uso de la IA está sesgado con respecto a principios occidentales. Y si bien Japón no es una de las principales potencias de la IA al ocupar el octavo lugar en el nivel de preparación de esta tecnología (The Global AI Index - Tortoise, n.d.; AI Index Report 2023 – Artificial Intelligence Index, n.d.), es importante develar las especificidades en torno a sus valores (McStay, 2021). Esto es, porque tener una perspectiva de valores más holística y menos centrada en orientaciones éticas y morales que condujeron a una visión regulatoria desde el punto de vista occidental permitirán

plantear una visión más clara para México que reconozca la diversidad cultural y las necesidades específicas de su contexto social y económico.

Finalmente, México podría aprovechar las oportunidades que ofrece la IA para resolver sus problemas y desafíos, sin renunciar a sus valores y principios. Y además, México tiene la posibilidad de aprender de la experiencia y las buenas prácticas de Japón, que ha logrado integrar la IA en su sociedad, respetando su tradición y su identidad. Así pues, la comparación entre México y Japón es una comparación desde una perspectiva de espejo con dos sistemas jurídicos provenientes de tradiciones similares, pero en las que México se puede nutrir de perspectivas en torno a los planes no sólo de regulación, sino de la armonización de la IA en su sociedad como es Japón.

I. El sistema jurídico mexicano

La historia del derecho mexicano es el estudio de la evolución de las normas jurídicas que han regulado la vida social, política y económica de los habitantes de México desde la época prehispánica hasta la actualidad. El derecho mexicano se ha formado a partir de la influencia de diversas fuentes, como el derecho indígena, el derecho castellano, el derecho novohispano, el derecho liberal, el derecho revolucionario y el derecho constitucional. Cada una de estas fuentes ha respondido a las circunstancias históricas y a los valores culturales de cada época, y ha planteado desafíos y oportunidades para el desarrollo de los derechos humanos en México.

Hay que señalar que el sistema jurídico mexicano es el conjunto de normas, principios e instituciones que regulan la vida social, política y económica de los habitantes de México. Además, el sistema jurídico mexicano se ha formado a partir de la influencia de diversas fuentes históricas y culturales. El sistema jurídico mexicano se puede clasificar dentro de la familia neorromanista o civilista, que se caracteriza por la importancia de la ley escrita como fuente principal del derecho.

En primera instancia, el derecho indígena se refiere al conjunto de normas y principios que regían la organización social, política y religiosa de los pueblos originarios que habitaban el territorio mexicano antes de la llegada de los

españoles. Entre estos pueblos se destacan los olmecas, los mayas, los zapotecas, los mixtecas, los toltecas y los mexicas o aztecas. El derecho indígena se basaba en la tradición oral, en la autoridad de los ancianos y en la cosmovisión de cada cultura. El derecho indígena tenía como fines principales la armonía con la naturaleza, la cohesión comunitaria y la defensa del territorio (Martínez Rodríguez & Mirafuentes García, 2020).

Más tarde, el derecho castellano se relaciona al conjunto de normas y principios que regían la organización social, política y religiosa de los reinos cristianos de la península ibérica, especialmente el reino de Castilla. El derecho castellano se basaba en la tradición escrita, en la autoridad del rey y en la religión católica. El derecho castellano tenía como fines principales la unificación política, la expansión territorial y la evangelización (Adame Goddard, 1983).

Posteriormente, el derecho novohispano expulsó al conjunto de normas y principios que regían la organización social, política y religiosa de los territorios americanos conquistados por España, especialmente el virreinato de Nueva España. El derecho novohispano se basaba en la combinación del derecho castellano y del derecho indígena, así como en las leyes especiales dictadas por la Corona para regular las relaciones entre españoles e indígenas. El derecho novohispano tenía como fines principales la explotación económica, la administración colonial y la protección de los indígenas (Adame Goddard, 1983).

Acto seguido, el derecho liberal se refirió al conjunto de normas y principios que regían la organización social, política y religiosa de los territorios americanos que se independizaron de España, especialmente México. El derecho liberal se basaba en la influencia de las ideas ilustradas, en la soberanía popular y en la separación de poderes. El derecho liberal tenía como fines principales la construcción nacional, el desarrollo económico y la garantía de las libertades individuales (Martínez Rodríguez & Mirafuentes García, 2020).

Por otro lado, el derecho revolucionario se refiere al conjunto de normas y principios que regían la organización social, política y religiosa de México durante el periodo histórico conocido como Revolución Mexicana (1910-1920). El derecho

revolucionario se basaba en las demandas sociales, políticas y agrarias de los distintos grupos revolucionarios que lucharon contra el régimen porfirista. El derecho revolucionario tenía como fines principales la transformación social, la justicia distributiva y el reconocimiento de los derechos colectivos (Martínez Rodríguez & Mirafuentes García, 2020).

Además, el derecho constitucional, la parte más reciente del derecho mexicano, plantea un conjunto de normas y principios que rigen la organización social, política y religiosa de México desde 1917 hasta la actualidad. Finalmente, el derecho constitucional se basa en la Constitución Política de los Estados Unidos Mexicanos de 1917 y sus reformas posteriores. El derecho constitucional tiene como fines principales el orden jurídico, la democracia representativa y participativa y la protección de los derechos humanos (idem).

Asimismo, la constitución mexicana es el elemento más importante dentro de la jerarquía de las normas. Es decir, como plantea Cárdenas Gracia (2010, p: 164):

La norma constitucional establece por un lado el principio de supremacía constitucional y también la jerarquía de las normas en el ordenamiento mexicano, en donde la Constitución es la norma de mayor importancia y todas las demás normas deben estar conforme a ella para ser válidas, es decir, para ser reconocidas como parte del derecho mexicano.

De esta manera (y retomando muy someramente al Walter Benjamin), la constitución mexicana es el contador de historias del sistema jurídico mexicano. Cada sistema y ley que devenga de esta tiene que basarse en la realidad que esta construye; que, aunque en ocasiones escasas, difusa en su en su palabra, adquiere la capacidad de construir e influenciar la realidad. En términos de Latour, la constitución mexicana, y en su par japonesa, son los programas supremos de donde devienen las relaciones jurídicas y armónicas entre actores humanos y no humanos.

La importancia normativa de la constitución mexicana se puede ver en su artículo 133. Aquí se plantean que todos los tratados internacionales como las normas que emanen de ella ocupan un rango “inmediatamente inferior a esta” (Rico Álvarez, Garza Bandala y Cohen Chicurel, 2011). No obstante, para propósitos se tienen que recordar que las reformas a la constitución, particularmente el artículo primero

de esta, ya hacen vinculantes algunas normas internacionales, lo cual puede implicaciones importantes si México si adscribe a un hipotético, pero lejano escenario, de amortiguación jurídica y regulatoria supranacional en relación con la IA.

De esta manera, bajo el ordenamiento jurídico actual, todas las normas se desprenden de la constitución. Y al mismo tiempo, prácticamente todas las normas se desprenden de estas y cada ley debe de ser compatible con este documento. Así pues, el establecimiento de la constitución mexicana actual marcó el momento de madurez del sistema jurídico mexicano; y en este sentido, se asemeja a la trayectoria donde la constitución japonesa marcó su propio hito normativo que aun rige su sistema jurídico.

Estas son algunas de las características generales de la historia del derecho mexicano, que muestran su complejidad y diversidad, lo que permite comprender el origen y la evolución de las instituciones jurídicas que rigen la vida de los mexicanos, así como los problemas y desafíos que enfrentan en el presente y el futuro. Ahora bien, es importante resaltar cómo se relaciona el sistema jurídico mexicano la IA y los derechos humanos.

Hay que recordar que, como se mencionó antes, la reforma del artículo primero de la constitución mexicana hizo vinculante ésta con los tratados internacionales, y en especial con los elementos de la declaratoria de los derechos humanos. De esta manera se generan marcos generales que tentativamente protegen a la población contra los riesgos de la IA acotada.

Sin embargo, esta vinculación no es suficiente para garantizar el respeto y la protección de los derechos humanos frente a los avances de la IA. Es necesario que el Estado mexicano desarrolle una regulación específica y adecuada para la IA, que contemple los principios éticos en sus marcos normativos, que deben orientar su diseño, uso y evaluación. Asimismo, es indispensable que se promueva la participación ciudadana, la educación digital y la transparencia en el ámbito de la IA, para evitar posibles abusos, discriminaciones o violaciones de los

derechos humanos por parte de los actores involucrados en su desarrollo y aplicación (García-Herrero & López-Tarruella, 2020).

Es importante mencionar que el derecho mexicano y el derecho japonés tienen diferentes orígenes y fuentes, pero ambos se enfrentan al desafío de regular la IA de manera que respete los derechos humanos y promueva el desarrollo social. De acuerdo con Cárdenas Gracia (2010), el derecho mexicano se basa en el derecho romano, el derecho español, el derecho francés y el derecho anglosajón, y tiene como fuentes principales la Constitución, las leyes federales y locales, los tratados internacionales y la jurisprudencia. El derecho japonés, por su parte, se inspira en el derecho chino, el derecho alemán y el derecho anglosajón, y tiene como fuentes principales la Constitución, las leyes ordinarias, los decretos administrativos y las costumbres (Kawashima & Sakurada, 2007). Ambos países han elaborado propuestas regulatorias sobre la IA que buscan garantizar la protección de los datos personales, la transparencia de los algoritmos y la responsabilidad de los agentes involucrados (Gobierno de México, 2020; Gobierno de Japón, 2020), aunque las propuestas regulatorias se encuentran en una fase más avanzada en Japón y se centran en solucionar sus problemas sociales.

Y por ello, como lo plantea esta tesis, también es importante observar cómo se articulan los mecanismos, los objetos, del sistema jurídico mexicano para hacerle frente a los retos de la IA en torno a los casos empíricos de Alphabet-Google, Uber, y el C5. Además de examinar cómo se relacionan con un país que es jurídicamente el espejo de México como Japón. Y analizar la forma en la que ambos sistemas jurídicos responden a los problemas y retos de estos sistemas con IA.

II. El sistema jurídico japonés

El sistema jurídico japonés moderno se desarrolló durante la segunda mitad del siglo XIX. Y si bien está influida por la tradición francesa, alemana y más tarde con elementos de la americana; en práctica, contiene muchos elementos de las perspectivas sociales de sus habitantes. Esto representa un espejo de agua frente a la disciplina jurídica mexicana que también recoge muchos de sus elementos del

derecho romano; lo cual es relevante, dado que se puede aprender cómo Japón reconcilia su perspectiva jurídica y social con respecto a los alcances cada vez mayores de la IA. De acuerdo con Opereza (2019):

Extrañeza y suficiencia son dos palabras con las que (Octavio) Paz intenta atrapar al Japón del siglo XX, y que, a pesar del tiempo transcurrido, sus categorías siguen estando presentes en la esencia de una nación que ha sabido ser moderna sin perder en el camino la naturaleza de su milenarismo.

El mismo autor plantea que el derecho en una nación es producto de su cultura y de las necesidades e ideologías en torno a un momento, a un suspiro del reloj, en específico. Y en el caso de Japón menciona que sus diferencias respecto al occidente “reflejan históricamente caminos diversos que obedecieron a visiones distintas del mundo, de la vida y del orden social” (ídem). Es decir, a pesar de que el país adoptó ropas en torno a las tradiciones jurídicas de occidente, en especial la romana derivada de la francesa y la alemana, esta visión de derecho está tejida con hilos hechos con la esencia del pueblo japonés.

Y es justo esta otredad, esta extrañeza, esta similitud y particularidad, sirve como modelo para comprender cómo un país afronta los retos sociales y éticos de la IA. Lo cual, a modo de bisagra comparativa, arroja lecciones para que México pueda construir elementos que respondan a las diversas oportunidades y retos que esa tecnología arroja. Pero como Japón lo hace, desde su propia realidad social, y de momento, con base en casos empíricos concretos que se van a examinar a lo largo de este trabajo.

Hay que mencionar la raíz del sistema jurídico de Japón. Esta se puede resumir en tres etapas: la del periodo imperial, la de los bakufu y los estados beligerantes, y la de la era moderna a partir de la transición Meiji. La primera estaba muy relacionada con el sistema legal de la antigua China, la segunda se caracterizó por el periodo de violencia de los Bushi que terminó en el Shogunato de Tokugawa donde el país se dividió en tres clases (el bushi o la clase guerrera, los campesinos, y finalmente los comerciantes y artesanos), y la tercera etapa fue la de la Transición Meiji que tuvo como momento cumbre la constitución de 1955 que

se firmó entre el gobierno de Japón y las Fuerzas de ocupación de Estados Unidos.

Para fines de este trabajo, solamente se examinará la última etapa. Dado que, de otra forma este apartado se extenderá demasiado y con ello se entorpece el flujo de la lectura de este texto. Además, los elementos más importantes para comprender la relación entre IA, sociedad y respuestas regulatorias se recogen del periodo que comenzó desde la restauración Meiji hasta la fecha.

La constitución moderna de Japón se gestó a partir de la restauración Meiji, que fue una respuesta al miedo de una invasión, colonización o subordinación ante los poderes occidentales después de que Estados Unidos forzó la apertura comercial al demostrar su poderío militar y tecnológico encarnado en los acorazados que envió a las costas del país bajo el Comodoro Perry. Un sector de la sociedad japonesa decidió entonces derrocar al sistema del bakufu de Tokugawa e instaurar en su lugar la figura del emperador, que hasta ese momento solamente había tenido un papel simbólico dentro del aparato de poder del país.

La restauración Meiji no solamente abrió las fronteras del país al comercio exterior, sino que significó también el inicio de un proceso de modernización en torno a aparatos tecnológicos, políticos, filosóficos y jurídicos occidentales. En el aspecto tecnológico, el país comenzó un proceso de industrialización científica, tecnológica y militar. En el aspecto social, Japón adoptó e incorporó elementos filosóficos y, sobre todo, políticos y jurídicos para estar al nivel de las potencias occidentales. Así pues, de acuerdo con Mori (2019):

Precipitado por la llegada de la flota americana, el poder del shogunato *Tokugawa* disminuyó y en 1867 se derrumbó el shogunato *Edo*, que había durado por más de 250 años. El Emperador, que había perdido la autoridad real durante varios siglos, vio restaurado su poder y se instaló como el núcleo del gobierno. Sin embargo, en términos de las instituciones nacionales, en lugar de volver a la era *Ritsu-Ryo*, el objetivo fue crear algo nuevo y europeo.

El objetivo de este nuevo orden jurídico fue establecer una institución similar a las que tenían las potencias europeas, en particular Francia. Así pues, se crearon la corte suprema y la dieta del parlamento. Mori explica: “Primeramente, en 1875 se

instaló el Senado. Sin embargo, este no era un lugar donde los representantes electos se reunían, puesto que el poder legislativo simplemente descansaba en el *Dajyokan*². El emperador tuvo el poder de legislar directamente bajo el proceso de edictos.

Ahora bien, quizás uno de los elementos más importantes dentro del periodo de restauración Meiji fue la separación del poder legislativo y ejecutivo del judicial. De esta forma, el poder de ejecutar justicia ya no descansaba ni en el Shogun ni en el emperador, sino en un sistema de cortes independientes encabezadas por una suprema corte. Mori (ídem) elabora esto al citar el artículo 57 de la constitución de ese periodo: “La Judicatura será ejercida por los Tribunales de Justicia según la ley, en nombre del Emperador”.

Hay que resaltar que, dado que no había personas especializadas en materias jurídicas con corte occidental en Japón, el gobierno tuvo que buscar especialistas en el exterior. Mori detalla que “el gobierno *Meiji* invitó a abogados de ultramar para trabajar en la creación de un código de ley. Una figura central entre los invitados fue *Gustave Emile Boissonade de Fontarabie* (1825-1910), un francés que llegó a Japón en 1873”. Y en torno al borrador de Boissonade se escribió un 1907 código penal que más tarde sería enmendado en 1948 bajo la influencia de legislación de Estados Unidos (ídem).

No obstante, la relevancia del borrador de Boissonade, este estuvo sujeto a críticas. Estas críticas, de acuerdo con Mori (ídem), hicieron referencia al código civil influenciado por el Derecho alemán. El explican entonces:

“De este modo, en 1896 se anunció el Volumen 1 (Reglas Generales), el Volumen 2 (Ley de Propiedad) y el Volumen 3 (Demandas) del Código Civil así revisado, seguido por el Volumen 4 (Parientes Familiares) y el Volumen 5 (Herencia) en 1898 y, de su promulgación en 1898. Los Volúmenes 1 al 3 siguen en uso hasta el día de hoy” (ídem).

Otro elemento en torno movimiento jurídico y político de la restauración Meiji fue la incorporación de los derechos humanos. Esto es, pese a que la figura del

² Sistema de ministros que durante un tiempo sirvieron como base de gobierno antes del periodo de los estados beligerantes y del alza al poder del bushi o los samurai.

emperador representó aun de facto los poderes legislativo y ejecutivo. De acuerdo a Mori: “El progreso favorable de occidentalización del sistema legal por parte del gobierno *Meiji* fue respaldado por su enfoque de cultivar recursos humanos desde las etapas iniciales” (ídem).

El elemento de los derechos humanos es aun de suma importancia, dado los pocos cambios sustanciales que sufrió la constitución Meiji después del periodo de ocupación al finalizar la segunda guerra mundial. Cambios cuyos cambios sustanciales fueron la remoción del emperador japonés a una figura simbólica y sin poder real; así como el cambio a la ley familiar centrada en la familia extendida entrelazada con la figura del esposo, y donde la mujer tenía un rol inferior (ídem). El elemento de los derechos humanos es importante porque en torno a esto se aconseja enfrentar muchos problemas éticos de la IA; la discriminación, el acceso a la información, la privacidad y la responsabilidad o justicia.

En suma, como menciona Mori, durante los últimos 150 años Japón ha vivido una etapa de occidentalización influenciada por diversas tradiciones del derecho romano como la francesa y la alemana. Y más tarde, por la tradición del derecho estadounidense emanada de la ocupación después de la segunda guerra. No obstante, vale la pena tomar en cuenta, que estos elementos de transición hacían un modelo jurídico occidental donde se encuentran los derechos humanos se logró con cierto ápice de armonía porque fue por decisión propia, por una carrera por sobrevivir un mundo cada vez más globalizado a finales del siglo XIX, y que dejó como resultado un proceso jurídico y político occidentalizado, aunque con elementos japoneses.

El sistema jurídico japonés moderno está estrechamente relacionado con su sociedad. A pesar de que su formación se debió a la necesidad de modernizarse y occidentalizarse a finales del siglo XIX, muchos componentes de su sistema de derecho están atravesados por prácticas sociales cotidianas y por una cosmovisión particular del mundo centrada en la armonía. Estas similitudes y diferencias con modelos occidentales de derecho, como el de México, son relevantes al momento de comprender los contras programas jurídicos que la

nación japonesa utiliza para hacer frente a los problemas cada vez más presentes que provoca la IA.

Además del paralelismo entre occidente y oriente, examinar el Derecho japonés desde una perspectiva social también es necesario para México dada la estrecha relación comercial y cultural que ambas naciones han establecido. Por ejemplo:

“Con el arribo del presente siglo, México y Japón entraron en las negociaciones para concertar un convenio económico, lo cual fue motivado también por el cambio de la estrategia japonesa del comercio multilateral, y finalmente firmaron el Acuerdo de Asociación Económica México-Japón en 2004 que entró en vigor a partir de 2005 (Carrillo *et al.*, 2011 en Okabe 2019)”.

Además, como señalan Okabe y Carillo (2014), se han intensificado los estudios socioculturales entre ambos países, reforzando aún más la estrecha relación entre Japón y México al enfrentar desde perspectivas diferentes el impacto y las oportunidades de la IA. Por tanto, es necesario examinar brevemente la relación entre el sistema jurídico japonés y su sociedad. Sin embargo, debido a limitaciones de espacio en este trabajo, se explicará brevemente el sistema jurídico japonés premoderno y luego se examinará cómo la cosmovisión japonesa ha influido en el sistema de justicia moderno; al respecto, Okabe (2019) señala:

Las características de este sistema jurídico pre-moderno se basaron en las siguientes creencias: 1) En la sociedad humana originalmente no existían conflictos o controversias, y los pensamientos y credos de los humanos eran iguales; y 2) La aplicación práctica de la ley había sido flexible conforme a las circunstancias sociales a pesar de las letras de la ley (Aomi, 1987: pp. 58 y 59).

Es necesario tener en cuenta que el sistema jurídico japonés moderno se desarrolló durante el periodo de la Restauración Meiji, cuando el país buscó modernizarse y occidentalizarse a finales del siglo XIX. Aunque sus códigos fueron inspirados en la tradición jurídica francesa y alemana, fueron objeto de controversia debido a que se redactaron rápidamente sin tener en cuenta el entorno social y cultural japonés (Okabe, 2019). Como resultado, estos códigos se enfrentaron a resistencias, ya que no encajaban con la realidad de la sociedad japonesa. Es importante también mencionar que las características del sistema social premoderno no han tenido un impacto significativo en el sistema jurídico

japonés moderno (Okabe, 2019); por lo tanto, en este trabajo se centrará en el sistema de derecho japonés moderno y su relación con la cosmovisión japonesa.

Como medida para tratar de resolver estas diferencias, comenzaron a surgir escuelas dedicadas a la investigación del derecho japonés, así como elementos que conciliaba las diferentes tradiciones occidentales que el gobierno Meiji estaba adoptando. Okabe (2019) lo explica de manera sintética cuando menciona: "en 1893 se comenzó a investigar y, a modificar el código civil, en el que se decidió adaptar, en lugar del sistema de instituciones conforme al derecho francés, el sistema pandectista del que Friedrich Carl von Savigny inició la escuela Pandektistik de Alemania en el siglo XIX". Por lo tanto, el gobierno Meiji comenzó, con sus particularidades, una tradición jurídica basada en el derecho alemán y francés con una constitución inspirada en la de Prusia, donde la cabeza de esta era la del emperador o Ten'nō (天皇).

La restauración Meiji fue un periodo de modernización y occidentalización de Japón a finales del siglo XIX, durante el cual el país adoptó y tradujo muchas ideas y conceptos occidentales, incluyendo el derecho. Según Okaba (2019), la razón por la que no existía el concepto de derecho en Japón era porque simplemente no era necesario usarlo. Además, el autor señala que no había una distinción clara entre derecho objetivo y subjetivo, ya que ambos términos reflejaban "el mismo fenómeno social desde distintas esferas" (ídem). Esto se debe en parte a la concepción particular del derecho en Japón, que está influenciada por sus tradiciones y su entorno social.

Los japoneses tienen una concepción particular del derecho, influenciada por sus tradiciones y su entorno social. Esto puede provocar que la conciencia de derecho sea percibida como baja, ya que la forma en que se articulan las herramientas jurídicas no se ajusta a la percepción occidental de derecho objetivo y subjetivo. Por ejemplo, a los japoneses les provoca ansiedad los contratos largos, ya que estos suelen exponer condiciones demasiado detalladas y no se prestan a la flexibilidad necesaria. En cambio, se reconoce que ambas partes del contrato

tienen una obligación general de actuar de buena fe y considerar los sentimientos y beneficios mutuos, la cual no cambiará antes y después del contrato.

El elemento que impulsa a los japoneses a cumplir un contrato es la vergüenza social. Kitano (2009) explica esto al señalar que "existe una expectativa inexpresiva de que una persona actuará/deberá actuar de acuerdo con su posición social, y romper los límites posicionales conducirá a la condena social y se refleja en un sentimiento de vergüenza". Okaba (2019) también menciona este punto: "Si la contraparte no cumple con lo prometido, será expulsado de la comunidad, perderá la confianza y tendrá vergüenza, por lo tanto, procura cumplir con su obligación. Esto es la conciencia de los japoneses sobre el contrato que proviene de su propia moralidad, ética e incluso lógica".

Otro elemento importante que influye en el derecho de facto es la armonía. Para comprender lo anterior, el derecho de facto se entiende en este texto como el derecho en acción anclado a la resolución de conflictos y cómo estos se resuelven. Esto genera un ambiente donde se busca evitar los conflictos en sí mismos, lo cual también tiene que ver con la perspectiva ética que adaptaron, aunque no asimilaron completamente, de los occidentales durante la restauración Meiji. Según Kitano (2009), "Rinri se aplicó a la traducción de Ética y desde entonces se ha utilizado. En mi opinión, Rinri sostiene dos conceptos: el sentido japonés y la traducción de la Ética".

Este punto es afianzado por Okaba (2019), quien menciona que "los japoneses tienen una conciencia diferente del derecho que los occidentales". Esta conciencia del derecho se manifiesta en la constante búsqueda de la armonía y en un elemento constante de conformismo para no agravar ni exacerbar el conflicto. Incluso hay anécdotas de personas japonesas que piden disculpas después de haber incomodado a la sociedad debido a las protestas o inconformidades que han levantado.

Como señala Kitano, esto no significa que la sociedad japonesa sea uniforme, ya que sería como decir que a todos los mexicanos les gusta salir de fiesta y buscar el placer de forma inmediata. Sin embargo, las condicionantes sociales de la

mayoría sí influyen de manera constante en la percepción del derecho y su relación con la ética y propuestas regulatorias en torno a la IA. Y en este sentido, aunque no es una sociedad completamente uniforme, las características históricas y sociales del país, entre ellas la búsqueda constante de armonía, la han convertido en una sociedad con rasgos colectivistas con tintes occidentales. Okaba (2019) elabora lo anterior al explicar:

“Los japoneses suelen actuar colectivamente. Si bien, dicho colectivismo no impone la inmersión ni la esclavitud de los miembros. En este sentido, no representa el totalitarismo ni antiindividualismo, sino que los japoneses tienen la conciencia de que el actuar colectivamente con los demás miembros y socializar activamente con el grupo finalmente propiciará para el beneficio individual”.

En suma, si bien el derecho japonés moderno fue construido de manera independiente de la sociedad, contiene elementos que en la práctica son influenciados por ella. Esto es relevante tanto para comprender los marcos regulatorios en torno a la IA, la resolución de conflictos alrededor de esta tecnología, y las lecciones que la sociedad mexicana puede aprender de la adaptación de programas o mecanismos regulatorios basados en sus propias necesidades sociales. Como menciona Okaba (idem): “el sistema jurídico como parte de la disciplina social se forma y desarrolla de acuerdo con la estructura social del país”.

B. Un breve contraste sobre mirada tecnológica de Japón

México y Japón son excelentes casos para contrastar el impacto de la IA en su sociedad. Desde la perspectiva ética, su población tiene relación con los objetos que pueden impactar la creación de programas que respondan a los problemas sociales y de derechos humanos planteados por la IA. En el caso de México, su respuesta al impacto social de la IA parece estar rezagada en comparación con Japón, donde la respuesta a la IA ya forma parte de varias agendas legislativas y planes de desarrollo que buscan abordar la IA no solo mediante marcos regulatorios, sino también integrándose como parte activa de su entramado social a través de iniciativas como la sociedad 5.0.

El animismo es una creencia según la cual todos los objetos y seres, tanto naturales como artificiales, tienen alma o espíritu. Esta creencia está presente en algunas tradiciones religiosas y culturales y puede influir en la forma en que se percibe y se trata a la IA. El animismo y la ética japonesa conocida como *rinri* pueden tener una relación con la regulación de la IA y su impacto en la sociedad en el sentido de que pueden influir en la forma en que se percibe y se trata a la IA y en la forma en que se consideran los aspectos éticos y morales en su desarrollo y uso.

Las iniciativas y regulación de Japón en torno a la inteligencia tienen que ver con la creencia tradicional de que los objetos tienen cualidades humanas o espirituales. Esto se relaciona en particular con la noción de *kami* (神) o entidad divina practicada en el sintoísmo. Según Kitano (2017): "En Japón, existe una creencia tradicional de la existencia de vida espiritual en objetos o fenómenos naturales llamados *mi* (el dios) y *tama* /魂 (el espíritu). Desde la era prehistórica (o periodo paleolítico japonés), la creencia en la existencia del espíritu se ha asociado con las tradiciones mitológicas japonesas relacionadas con el sintoísmo" (Kitano, 2007).

Esta perspectiva japonesa contrasta con la tradición judeo-católica occidental mexicana. Por ejemplo, a pesar de su sincretismo con las religiones locales, la versión católica mexicana no contempla en sí misma objetos con un valor espiritual importante. E incluso, las tradiciones cuasi apócrifas encarnadas en los santos populares como San Valverde o la Santa Muerte tampoco conciben a actores no humanos como posibles divinidades que intervienen en lo sagrado y que merecen ser objeto de veneración.

El concepto de animismo se puede rastrear a Tylor. En el artículo "La religión como una dimensión de la cultura", Camarena y Tunal (2009) explican que:

"...la concepción y el estudio del animismo comenzado por Tylor se basa en la figura central del alma, la cual opera como una entidad explicativa de los fenómenos que los seres humanos no podemos explicar, aunque no es la adoración de los objetos una devoción por los mismos en su forma material, sino un culto a las características sobrenaturales que a éstos se les deposita o la

generación de un conjunto simbólico en torno a dichos objetos y de un escenario de la manifestación de las presencias no naturales o sobrenaturales, mismas que podemos llamar divinas”.

Otra definición de animismo se puede encontrar en el texto de "Techno-animism in Japan: Shinto Cosmograms, Actor-network Theory, and the Enabling Powers of Non-human Agencies" de Jensen y Block, en particular la definición que citan de Descola. Según los autores: “la cualidad distintiva del animismo, según Descola, es que proyecta características humanas sobre la naturaleza” (Jensen & Blok, 2013, p. 89). Esto significa que la cosmovisión animista otorga una cualidad de agencia a lo que forma parte del medio ambiente natural, como mencionó Kitano en relación con la tradición sintoísta japonesa.

Por lo tanto, el animismo es la práctica de atribuir cualidades humanas a objetos naturales hasta el punto de su veneración. Por ejemplo, las culturas que creen que los ríos tienen alma o que veneran a la madre naturaleza. Por consiguiente, las religiones animistas le dan un carácter de agencia a los objetos, a las cosas; como resultado, consideran que todo tiene alma, en especial, aquello que forma parte de la naturaleza

Es necesario tener en cuenta que no es correcto generalizar sobre la perspectiva religiosa occidental o sobre la religión en un país completo como México. Cada individuo y cada comunidad puede tener creencias y prácticas religiosas diferentes, y no es justo decir que toda la perspectiva religiosa occidental aboga por la separación de los reinos naturales y humanos. Además, aunque es cierto que algunas tradiciones judeocristianas pueden enfatizar esta separación, también hay muchas otras tradiciones y corrientes dentro de estas religiones que enfatizan la interconexión entre los reinos naturales y humanos y que promueven una relación respetuosa y responsable con el medio ambiente.

En cuanto a la regulación de la IA, es cierto que la perspectiva y las creencias religiosas de una sociedad pueden influir en cómo se percibe y se aborda la tecnología. Sin embargo, es importante recordar que la regulación de la IA debe ser un asunto más amplio que considere no solo aspectos éticos y morales, sino también cuestiones legales, técnicas y prácticas. Por lo tanto, es importante no

reducir la regulación de la IA a una cuestión de animismo o de cualquier otra creencia religiosa en particular.

Hay que tomar en cuenta que en México y Japón están implementando muchas herramientas de IA. En el caso de la Ciudad se están desplegando plataformas que hacen uso de la IA de manera pública y privada en la forma de aplicaciones y APIS que ofrecen servicio para mejorar la movilidad. Además, en la Ciudad de México ha desarrollado plataformas que usan IA para mejorar y agilizar los servicios de trámites digitales que ofrece.

Un ejemplo cercano de la alianza entre el sector privado y el gobierno de la ciudad de México fue su alianza con la empresa Cisco Systems en 2019. De acuerdo con el periódico Expansión:

“En presencia de Claudia Sheinbaum, jefa de gobierno de la ciudad y Chuck Robbins, director general de Cisco, la tecnológica firmó un acuerdo de colaboración con la Agencia Digital de Innovación Pública (ADIP) y el gobierno de CDMX para digitalizar la entidad bajo el programa Aceleración de la Digitalización de los Países (Country Digitization Acceleration Program, CDA) de Cisco”. (Expansión, 2019).

El objetivo de la alianza se centra en facilitar la transición a un gobierno inteligente o digital. Esto permitiría a la ciudadanía tener acceso a servicios digitales del gobierno a través del modelo de e-gobierno, así como proporcionar puntos de conexión a internet mediante wifi público. Sin embargo, hay que señalar que hay otras empresas, como Alphabet Inc, que ofrecen servicios infraestructurales a través de plataformas como Google Maps que también ayudan a la población.

Más tarde, el 11 de noviembre de 2021, la Ciudad de México obtuvo el récord Guinness World Records de "la ciudad más conectada del mundo al convertirse en la urbe con más puntos de wifi gratuitos, con un total de 21,500, superando las cifras de Moscú" (Staff de Forbes, 2021). Esto fue una noticia importante ya que el acceso al internet municipal facilita el uso de servicios de e-gobierno, lo que reduce la brecha digital. Además, al tener más población conectada, es más factible la adopción de plataformas con IA por parte del gobierno de la Ciudad de México.

En el caso de Japón, y en particular de Tokio, la IA está teniendo un papel más omnipresente en la sociedad. Esto es porque, además de construir infraestructuras y plataformas que utilizan la IA, se están poniendo a prueba parámetros éticos en torno a la ciudadanía de actores no humanos con inteligencia no humana. Un ejemplo notable es Mirai (未来/Futuro), un robot de chat que fue reconocido como residente de Tokio según una noticia publicada en la revista Forbes en 2017 (Caughill, 2017).

Además, a nivel más práctico, el gobierno de Tokio está implementando la IA en varios aspectos de la vida cotidiana. Por ejemplo, para los Juegos Olímpicos de 2020, el gobierno de Japón decidió desplegar un sistema de seguridad basado en la IA para agilizar la entrada en distintos puntos de acceso, según informó la Agencia de Seguridad Nacional de Japón (NSO) en 2019 (NSO, 2019):

“Como parte de su sistema de seguridad impulsado por IA, el software de reconocimiento facial de NEC se utilizará para autenticar la identidad de los atletas, jueces y personal durante la admisión a los eventos. El software de reconocimiento facial de NEC utiliza un motor de IA llamado NeoFace, que puede extraer rostros de videos y unirlos en tiempo real” (Vu, 2021).

También se usa la IA en el transporte público de Tokio. Un ejemplo es la línea de tren Yamanote, que es automatizada, pero no inteligente. Sin embargo, el gobierno de la ciudad planea implementar IA para mejorar la eficiencia de otras líneas de tren. Esto incluye, por ejemplo, tecnología de visión por computadora para inspeccionar rutinarias del equipo y la infraestructura del tren, y sistemas de asistencia al conductor para mejorar la seguridad y precisión del servicio. Además, el gobierno de Tokio también está explorando la posibilidad de utilizar la IA para mejorar la planificación del transporte y para proporcionar información en tiempo real a los pasajeros:

“...como líder digital en la industria ferroviaria, planeamos continuar participando activamente en esfuerzos similares. Queremos dejar el trabajo repetitivo y de gran volumen a las tecnologías digitales o la IA [IA], al tiempo que ponemos a los seres humanos a cargo de la toma de decisiones y del trabajo más creativo” (Yamamura³ en McKinsey & Company, 2021).

³ Akiyoshi Yamamura es el Director Ejecutivo del Metro de Tokio.

En ambas ciudades, la Ciudad de México y Tokio, se ha implementado la IA y la ciencia de datos para mejorar los servicios a la ciudadanía, ya sea a través de entidades privadas, públicas o una combinación de ambas. Sin embargo, se puede observar una diferencia en la forma en que estas ciudades se acercan a la implementación de la IA en los servicios y herramientas disponibles para la población. Mientras que la Ciudad de México lo ha hecho de manera más conservadora, principalmente mediante estrategias que faciliten la digitalización y el acceso a plataformas con IA y servicios de gobierno digital, Tokio ha adoptado una infraestructura de servicios más agresiva, que incluye desde sistemas de vigilancia hasta vehículos y sistemas de transporte sin tripulación.

Ahora bien, el uso de los sistemas de reconocimiento facial, dado que de estos se desprenden los casos empíricos para esta tesis. En la Ciudad de México, se ha implementado un sistema de reconocimiento facial llamado C5 (Centro de Control, Comando, Comunicación, Computo y Correlación) en un esfuerzo por mejorar la seguridad y el orden público. Este sistema utiliza cámaras de vigilancia y tecnología de reconocimiento facial para detectar y rastrear a individuos sospechosos y alertar a las autoridades. Sin embargo, ha generado controversia por preocupaciones sobre la privacidad y el uso discriminatorio de la tecnología.

En Tokio se han implementado diversos sistemas de reconocimiento facial en un esfuerzo por mejorar la eficiencia y la seguridad. Por ejemplo, se han instalado cámaras de reconocimiento facial en estaciones de metro y estaciones de tren para facilitar el acceso y mejorar la seguridad; y después de una visita de campo en abril de 2023, se constató que prácticamente todo el sistema de trenes y metro cuenta ya con sistemas de cámaras dentro de los vagones. También se ha implementado el uso de reconocimiento facial en algunos casinos y en algunos edificios gubernamentales para controlar el acceso. Al igual que en la Ciudad de México, el uso de la tecnología de reconocimiento facial en Tokio ha generado controversia debido a preocupaciones sobre la privacidad y el uso potencialmente discriminatorio de esta tecnología.

Sin embargo, hay que tener en cuenta que la implementación y el uso de sistemas de reconocimiento facial en Tokio pueden variar significativamente en comparación con los de la Ciudad de México. Por ejemplo, según un informe de la Agencia de Seguridad Nacional de Japón, el uso del reconocimiento facial en Tokio se ha centrado principalmente en el control del acceso y la verificación de la identidad, mientras que en la Ciudad de México se ha utilizado principalmente para fines de vigilancia y seguridad (Agencia de Seguridad Nacional de Japón, 2020). Además, en Tokio se han establecido reglamentos y leyes específicas para regular el uso del reconocimiento facial, mientras que en la Ciudad de México puede haber menos regulación y transparencia en cuanto al uso de esta tecnología (BBC, 2020).

En suma, es necesario destacar que Japón ha sido un líder en la implementación y uso de la IA en diversos aspectos de la vida cotidiana, desde el transporte público hasta el uso en edificios gubernamentales y casinos. Esto ha llevado a la creación de un marco ético para regular el uso de la IA en el país, con el objetivo de garantizar la protección de la privacidad y evitar el uso discriminatorio de esta tecnología. Por lo tanto, el análisis de cómo Japón ha abordado la implementación y regulación de la IA puede ser útil para México en su camino hacia una mayor integración de la tecnología en su sociedad.

C. Perspectivas sociales de México y Japón en torno a la IA

Es importante analizar las visiones sociales de México y Japón respecto a la IA. Esto permite entender mejor las estrategias que ambos países han adoptado para enfrentar los efectos de esta tecnología. Asimismo, esto contribuye a tener una visión más amplia de los derechos humanos que se incluyen en sus instrumentos regulatorios y de la postura de sus sociedades hacia esta tecnología.

Para ello, es útil comparar las características sociales de cada país a través de seis indicadores desarrollados por el profesor Hofstede, los cuales se pueden encontrar en el sitio web de su empresa. Estos indicadores son: Distancia del poder, Individualismo, Masculinidad, Aversión a la Incertidumbre, Orientación a largo plazo e Indulgencia. Estos rasgos reflejan un aspecto importante de la

sociedad de los dos países, y juntos permiten compararlas y comprender sus actitudes hacia la IA, desde el aspecto regulatorio hasta las decisiones individuales.

En el caso de México, según el sitio de Hofstede (2021), la distancia al poder es moderada. Esto quiere decir que la población espera y acepta cierto grado de desigualdad en la distribución del poder, pero también existe una cierta lucha por participar en el proceso de toma de decisiones. Por otro lado, en el caso de Japón, la distancia al poder es baja, lo que indica que la población espera y acepta una mayor desigualdad en la distribución del poder y hay menos lucha por participar en el proceso de toma de decisiones; lo cual, podrá tener un impacto en la forma en que ambos países abordan la implementación de la IA y en la forma en que sus sociedades reaccionan a esta tecnología.

En el caso de México, de un puntaje que va del uno al cien, Hofstede estima que la distancia al poder es de 81. De acuerdo con el sitio de la organización, esto significa que el país es una sociedad jerárquica. Es decir, que las personas en México tienden a aceptar un sistema jerárquico "...en el que todos tienen un lugar y que no necesita mayor justificación. Se considera que la jerarquía en una organización refleja desigualdades inherentes, la centralización es popular, los subordinados esperan que se les diga qué hacer y el jefe ideal es un autócrata benévolo" (ídem).

El párrafo anterior se basa en la teoría de las dimensiones culturales de Hofstede, que es un marco para la comunicación intercultural que muestra los efectos de la cultura de una sociedad sobre los valores de sus miembros y cómo estos valores se relacionan con el comportamiento (Hofstede, 1984). Una de las dimensiones que propone Hofstede es la distancia al poder, que se define como el grado en que los miembros menos poderosos de las instituciones y organizaciones dentro de un país esperan y aceptan que el poder se distribuya de manera desigual (Hofstede, 1985). Así pues, compara la distancia al poder en México y Japón, usando los datos del sitio web de Hofstede (2021), y sugiere que esta dimensión puede tener un impacto en la forma en que ambos países abordan la

implementación de la IA y en la forma en que sus sociedades reaccionan a esta tecnología.

En el caso de Japón, obtuvo una puntuación de 54. Esto significa que Japón es un país cuya sociedad es relativamente jerárquica. El sitio Hofstede Insights lo explica de la siguiente manera: "Sí, los japoneses siempre son conscientes de su posición jerárquica en cualquier entorno social y actúan en consecuencia. Sin embargo, no es tan jerárquico como la mayoría de las otras culturas asiáticas" (ídem).

Para comparar con otro país, Canadá obtuvo una puntuación de 9 en la distancia al poder. Esto significa que Canadá es una sociedad con una distancia al poder muy baja, lo que se refleja en una jerarquía relativamente plana y en la valoración de la igualdad y la participación democrática. Según Hofstede Insights, "en Canadá, la gente espera que se les trate con igualdad y respeto, independientemente de su posición jerárquica. Los líderes son esperados para ser participativos y consultivos, y el poder se distribuye más equitativamente" (ídem).

De acuerdo con el sitio de la organización de Hofstede, existe una percepción de que la sociedad japonesa es extremadamente jerárquica basada en la experiencia de las personas que hacen negocios allí debido a lo minucioso que es el proceso de toma de decisiones, desde el primer encuentro hasta en los centros comerciales como Tokio. Sin embargo, este lento proceso de toma de decisiones también muestra que en la sociedad japonesa no hay un solo líder que pueda tomar decisiones de manera autoritaria, como en sociedades más jerárquicas. Otro ejemplo de una distancia al poder no tan alta es que Japón siempre ha sido una sociedad meritocrática, donde el mérito y la competencia tienen un peso más grande, especialmente en su sistema educativo.

En este sentido, se puede apreciar que la sociedad mexicana es más jerárquica que la sociedad japonesa y que muchas de las decisiones que se toman en ella se dan desde una perspectiva centralizada. Por otro lado, la sociedad japonesa tiende a darle un peso mayor al sistema meritocrático y a la competencia (ídem). Y esto puede ayudar a comprender cómo se dan los procesos regulatorios en torno a la IA en un país y otro.

El siguiente rasgo que plantea Hofstede Insights es el Individualismo. De acuerdo con el sitio web de la organización, este rasgo plantea los grados de interdependencia entre los miembros de una sociedad. Dicho de otra manera: “En las sociedades individualistas, se supone que las personas deben cuidar de sí mismas y únicamente de su familia directa. En las sociedades colectivistas las personas pertenecen a 'grupos' que los cuidan a cambio de lealtad” (ídem).

En cuanto a la individualidad, México obtuvo 30 puntos de cien en la escala de Hofstede Insights. Esto significa que la sociedad mexicana no se puede considerar como individualista y que sus rasgos y decisiones se inclinan más hacia lo colectivo. Según Hofstede Insights, esto se manifiesta en un compromiso a largo plazo con el grupo de miembros, ya sea una familia, una familia o relaciones extendidas. La lealtad en una cultura colectivista es primordial y anula la mayoría de las demás normas y reglamentos sociales.

En lo que respecta a Japón, el país obtuvo 46 puntos de cien en torno a la categoría de individualidad. Esto rompe la percepción popular de que la sociedad japonesa es extremadamente colectivista, aunque no deja de tener rasgos importantes de ordenamiento colectivo como la persecución de la armonía relacionada con su visión particular de la ética y la moral o *rinri* (倫理). Esto significa que “los japoneses son experimentados como colectivistas según los estándares occidentales e individualistas según los estándares asiáticos. Son más privados y reservados que la mayoría de los demás asiáticos” (ídem).

En el caso de Canadá, obtuvo una puntuación de 80 en la escala de individualidad. Esto significa que Canadá es una sociedad más individualista, donde se valora la independencia y el espacio personal y se espera que las personas sean responsables de sus propias acciones y decisiones. Según Hofstede Insights, “en Canadá, la gente espera tener la libertad de tomar sus propias decisiones y de ser responsable de sus propias vidas. Los canadienses valoran el tiempo y el espacio personal y esperan ser tratados con respeto y consideración” (ídem).

Canadá tiene una puntuación alta en individualismo, lo que significa que valora la autonomía y la libertad personal, mientras que Japón tiene una puntuación baja, lo que indica que valora la armonía y la lealtad grupal. México tiene una puntuación intermedia, lo que sugiere que tiene una mezcla de valores individualistas y colectivistas. E incluir a Canadá como referente comparativo ayuda a mostrar el contraste entre las culturas individualistas y colectivistas, y cómo esto puede influir en la forma en que se percibe y se regula la IA, especialmente en relación con los derechos humanos de privacidad, acceso a la información y no discriminación.

En comparación con Canadá, la sociedad mexicana es más colectivista y menos individualista. Esto se manifiesta en un compromiso a largo plazo con el grupo de miembros y un valor primordial dado a la lealtad hacia el grupo. Canadá, por su parte, es una sociedad más individualista, donde se valora la independencia y el espacio personal y se espera que las personas sean responsables de sus propias acciones y decisiones.

Ahora bien, es cierto que Japón se considera colectivista según los estándares occidentales, pero no cuenta con un sistema de familia extendida como sí lo tienen otras sociedades asiáticas más colectivistas como China y Corea. Esto se debe a que Japón ha tenido una historia y tradiciones culturales diferentes a otras sociedades asiáticas y ha desarrollado un sistema de familia nuclear en lugar de la familia extendida tradicionalmente presente en otras sociedades del continente.

Además, Japón es conocido por ser una sociedad más reservada y privada que la mayoría de los demás países asiáticos. Esto se ve cómo los japoneses interactúan en la vida cotidiana, donde se valora la privacidad y se evita hacer preguntas personales o intrusivas. Esta reserva también se manifiesta en el sistema judicial japonés, donde se enfatiza el diálogo y el acuerdo entre las partes implicadas en lugar de la resolución de conflictos mediante el uso de la fuerza o la imposición de sanciones.

El tercer punto examinado por Hofstede Insights es la masculinidad. Un mayor puntaje en esta categoría no implica necesariamente que una sociedad tenga conductas machistas, sino que tiene que ver con la actitud de la sociedad hacia la

competencia. Una sociedad con un alto puntaje de masculinidad tolera y aprueba más un ambiente de competencia donde el éxito se define por el ganador o el mejor en el campo, un sistema de valores que comienza en la escuela y continúa a lo largo de la vida organizacional.

De acuerdo con Hofstede Insights, hasta 2021 México obtuvo 69 puntos de cien en la categoría de masculinidad. Esto indica que la sociedad mexicana tiende hacia la búsqueda de éxito ligada a la masculinidad y hacia la orientación hacia el trabajo. Además, esto también sugiere que la sociedad mexicana vive para trabajar y no trabaja para vivir (Hofstede Insights, 2021). Esto se refleja en las estadísticas que muestran que hasta el 2011, la población mexicana era una de las que más horas trabajaban en el mundo, aunque esto no se reflejó en su productividad (Giattino, 2020).

En cuanto a Japón, el país obtuvo 95 puntos de cien en la categoría de masculinidad según Hofstede Insights. Esto indica una alta competitividad en la sociedad japonesa; pero la masculinidad japonesa se manifiesta de manera diferente por su orientación colectivista. Por lo tanto, el sentido de competencia en Japón se da más entre grupos de corporaciones en lugar de entre individuos (Hofstede Insights, 2021). Esto también se puede observar en fenómenos como los equipos de béisbol que representan a corporaciones, el orgullo de representar a una empresa importante, y la cultura de regalos como las frutas de alto nivel como el mango del sol.

Según Hofstede Insights, México tiene una puntuación de 69 en la dimensión de masculinidad, lo que implica que la sociedad mexicana valora el éxito y la competencia. Esto podría explicar por qué las propuestas de regulación de la IA en México se basan en principios de responsabilidad más generales y menos específicos para cada organización. Por otro lado, Japón tiene una puntuación de 95 en la dimensión de masculinidad, lo que refleja que la sociedad japonesa es muy competitiva. Sin embargo, hay que considerar que en Japón la competencia se da principalmente entre grupos empresariales, y no tanto a nivel individual; lo que explica por qué las propuestas de regulación de la IA en Japón se enfocan

más en la responsabilidad de las empresas, pero no tanto en la responsabilidad individual.

El cuarto punto trata sobre la evasión de la incertidumbre, que se refiere a la forma en que las personas y sociedades se preparan para situaciones futuras que no pueden controlar. Según Hofstede Insights, este parámetro está relacionado con preguntas como: "¿debemos tratar de controlar el futuro o simplemente dejar que suceda?" Esta ambigüedad puede generar ansiedad y diferentes culturas han encontrado maneras de lidiar con ella de maneras distintas. Por lo tanto, se puede decir que sociedades con puntajes altos en esta categoría planifican más a futuro que aquellas con puntajes bajos, pero también son más propensas a seguir las reglas y a "mantener sistemas rígidos de creencias" (idem).

México tiene 82 puntos en esta categoría, lo que se considera un puntaje alto. Esto puede ayudar a explicar por qué la sociedad mexicana tiende a seguir ciertos códigos de creencias formales, religiosos e informales. En particular, la aversión al riesgo se manifiesta en la tendencia a querer ser parecidos a los demás y seguir las normas para evitar la incertidumbre o el rechazo. Según Hofstede Insights, "la medida en que los miembros de una cultura se sienten amenazados por situaciones ambiguas o desconocidas y han creado creencias e instituciones que intentan evitarlas se refleja en la puntuación de Evitación de la incertidumbre" (idem).

La evitación de la incertidumbre es cómo una cultura afronta lo desconocido y crea normas para evitarlo (Hofstede, 1984). Según Hofstede (2021), México evita mucho la incertidumbre y sigue códigos de creencias formales e informales. Canadá evita poco la incertidumbre y es más abierto y tolerante al cambio. Japón evita mucho la incertidumbre y es muy cuidadoso y previsor ante los riesgos. Comparar a Canadá con México y Japón muestra el contraste entre las culturas que evitan o aceptan la incertidumbre, y cómo esto afecta a la percepción y regulación de la IA y los derechos humanos relacionados.

De acuerdo con Hofstede Insights, Japón obtuvo 92 puntos de 100 en la categoría de evasión de la incertidumbre. Esto se atribuye a menudo a la frecuencia con la

que Japón es amenazado por desastres naturales como terremotos, tsunamis, tifones y erupciones volcánicas. Como resultado, los japoneses han aprendido a prepararse para cualquier situación incierta: perspectiva que es compartida por muchos estudiosos de la sociedad japonesa, como Shogo del canal de Youtube *Let's ask Shogo*, y durante la pandemia se ha reflejado en la lenta y cuidadosa apertura de su frontera a visitantes.

¿Cómo afecta esto la regulación de la IA en México y Japón? En el caso de México, aunque la sociedad tiene una alta aversión al riesgo, según Hofstede Insights (s.f.), las prácticas políticas y las tradiciones obstaculizan la presión por proponer regulaciones en torno a la IA. En cambio, en Japón, su aversión al riesgo se refleja en una serie de propuestas regulatorias en torno a la IA que buscan no solo abordar los problemas éticos que surgen de esta tecnología, sino también integrar de manera responsable para resolver problemas urgentes como la baja tasa de natalidad y el envejecimiento de la población (Kawakami & Tanaka, 2020).

En comparación con México y Japón, Canadá también ha implementado regulaciones en torno a la IA. Según el Ministerio de Innovación, Ciencia y Desarrollo Económico de Canadá (2021), el país ha establecido un marco de ética y responsabilidad para garantizar que la IA se utilice de manera ética y responsable. Además, Canadá ha creado el Consejo de IA Canadá, una organización sin fines de lucro encargada de promover el uso responsable de la IA en el país.

El quinto elemento de Hofstede Insights en su análisis y comparación entre países es la orientación a largo plazo. Según el sitio web de la organización, la orientación a largo plazo se refiere a "cómo cada sociedad mantiene algunos vínculos con su propio pasado mientras enfrenta los desafíos del presente y del futuro, y cómo las sociedades priorizan estos dos objetivos existenciales de manera diferente" (Hofstede Insights, s.f.). Además, las sociedades con un puntaje bajo en esta categoría tienden a ser más normativas y centradas en mantener tradiciones lejanas, mientras que las sociedades con un puntaje alto suelen ser más pragmáticas (Hofstede Insights, s.f.).

México tiene un puntaje de 24 de 100 en la categoría de orientación a largo plazo. Esto significa que la sociedad mexicana es más propensa a seguir normas y tradiciones incrustadas durante muchas décadas o años y a establecer verdades absolutas. Según Hofstede Insights (s.f.), las sociedades con un puntaje bajo en esta categoría "muestran un gran respeto por las tradiciones, una propensión relativamente pequeña a ahorrar para el futuro y un enfoque en lograr resultados rápidos".

En el caso de Japón, el país tiene un puntaje de 88 en la categoría de orientación a largo plazo. Esto significa que el país suele prepararse mucho más hacia el futuro, especialmente en términos de políticas corporativas y públicas, y a veces parece ir en contra de la corriente de otros países industrializados, como en su lenta apertura de fronteras durante la pandemia de COVID-19. Sin embargo, según Hofstede Insights, la orientación a largo plazo también se relaciona con una especie de fatalismo elegante, donde la vida no se da por sentado, y esto también está relacionado con la frecuencia de desastres naturales en Japón, ya que "un 20% de todos los terremotos mayores de 6.0 grados ocurren en Japón" (Let's ask Shogo, 2020).

En cuanto a la regulación de la IA, los resultados en la categoría de orientación a largo plazo pueden ayudar a explicar por qué la sociedad mexicana no siente la presión de examinar, definir y regular los riesgos y beneficios de esta tecnología. Por otro lado, el puntaje alto de Japón en esta categoría puede explicar su avance en propuestas regulatorias, investigación, desarrollo e implementación de la IA. Es interesante notar la diferencia en la visión a largo plazo de Japón en cuanto a la inversión en tecnología como medio de supervivencia en un mundo globalizado, incluso en comparación con México, que también ha sufrido desastres como guerras.

La última categoría para examinar consiste en comparar las diferencias culturales entre México y Japón es la de indulgencia. Según Hofstede Insights (s.f.), esta dimensión se refiere al grado en que las personas intentan controlar sus deseos e impulsos, dependiendo de la forma en que fueron criados. El control relativamente

débil se llama indulgencia y el control relativamente fuerte se llama restricción. La organización señala que el grado de indulgencia o restricción de una sociedad está vinculado y relacionado con los procesos de socialización que tienen las personas desde niños.

México tiene un puntaje de 70 de 100 en la categoría de indulgencia. Esto significa que la sociedad mexicana tiende a tener menos control sobre sus deseos e impulsos y a ser más indulgente en comparación con otras sociedades. Esto puede manifestarse en la forma en que la sociedad mexicana se relaciona con el placer, el consumismo y el disfrute inmediato. Por otro lado, Japón tiene un puntaje de 5 de 100 en la categoría de indulgencia, lo que significa que la sociedad japonesa tiende a tener más control sobre sus deseos e impulsos y a ser más restrictiva. Esto puede manifestarse en la forma en que la sociedad japonesa valora la disciplina, la responsabilidad y la moderación (siempre es importante tener en cuenta que estos puntajes son solo una aproximación y que cada individuo es único y puede tener sus propias formas de controlar sus deseos e impulsos).

Los resultados en la categoría de indulgencia pueden ser útiles para analizar las propuestas de regulación de la IA. Una sociedad indulgente podría estar más dispuesta a aceptar los riesgos y problemas éticos de una plataforma con IA a cambio de acceder rápidamente a sus servicios y obtener gratificación inmediata. Por otro lado, una sociedad como la japonesa, que tiene una mayor restricción, podría ser más vigilante y tener una mayor adversidad a los problemas éticos relacionados con plataformas con IA. Esto, junto con su orientación a largo plazo, podría llevar a mecanismos de regulación e implementación más robustos y adaptados a las necesidades de su población.

En síntesis, las sociedades mexicana y japonesa tienen similitudes y contrastes muy interesantes en algunos rasgos sociales examinados por la organización Hofstede Insights. Y esto puede ayudar a comprender las propuestas, actitudes y herramientas regulatorias en torno a la inteligencia, sobre todo con respecto a cómo se siente su población en torno a esta tecnología. Los puntos de mayor

cercanía son individualismo (ambas sociedades son colectivistas) y en la aversión a la incertidumbre que las hace seguir patrones normativos de forma rígida; y los rasgos donde divergen mucho son la Aversión a la Incertidumbre y la Indulgencia, lo que ayuda a comprender cómo y por qué plantean elementos en torno a la regulación de la IA y por qué hay ausencia de estos.

Hay que resaltar que estas seis dimensiones de análisis son generales. Y que no toda sociedad es en un país es la misma, ya que incluso se pueden notar diferentes actitudes sociales a lo largo de diferentes estados y prefecturas; además, hacen falta estudios empíricos, sobre todo en la dimensión de indulgencia, que ayuden a generar evidencia desde el nivel microsocial para apoyar el panorama general. Aun así, a nivel nacional, las seis dimensiones de análisis social de Hofstede Insights brindan herramientas para comprender cómo la IA transgrede los derechos de privacidad, acceso la información, no discriminación, y genera problemas en torno a la responsabilidad y explicabilidad; así como la forma y el por qué una sociedad responde ante eso.

D. Los aspectos ético-culturales de México y Japón con respecto a la IA

Más allá de los indicadores de Hofstede, hay que resaltar las diferencias culturales entre ambos países para comprender aun más la forma en la que articulan sus políticas, y la manera en que perciben los riesgos de la IA. Desde la perspectiva de Kitano (2007), la tecnología de robots japoneses se centra en perfeccionar la capacidad mecánica sin debatir mucho la ética del empleo de robots, mientras que, en Occidente, los expertos en robótica suelen plantear los desafíos sociales y éticos para integrar robots a las comunidades humanas. Asimismo, la ética japonesa, y con ello, la perspectiva de riesgos con respecto a las transgresiones de la IA, puede generar perspectivas, más no marcos, normativos diferentes a los de México.

Ahora bien, esto no significa que no haya influencias importantes en Japón, sobre todo a nivel regulación, con respecto a los problemas que ocasiona la IA. De acuerdo con McStay (2021), estas influencias abarcan organismos internacionales, como el Foro Económico Mundial (FEM) y la Organización para la

Cooperación y el Desarrollo Económicos (OCDE), todos fundamentados en el avance mediante el comercio económico liberal. No obstante, McStay retoma algunas nociones de autores como Morita (2012) que no considera que los derechos humanos internacionales sean incompatibles con los “valores asiáticos”, sino que sostiene que los derechos humanos universales deben fundamentarse en los propios términos y perspectivas de una región; y esto concuerda con las sugerencias más amplias sobre el pluralismo ético (Ess, 2020 en McStay, 2021).

Mcstay (2021) argumenta además que el debate ético occidental y global sobre las tecnologías emergentes se beneficiará de las ideas valiosas de las nociones filosóficas japonesas (o condimentos) de comunidad, integridad, honestidad y corazón que se encuentran en las perspectivas éticas japonesas. En este sentido, el punto es retomar una perspectiva de ética en torno a la noción japonesa de Rinri ya que Japón “Rinri es el estudio de la comunidad, o de la forma de lograr la armonía en las relaciones humanas, mientras que, en Occidente, hasta donde yo entiendo, la ética tiene una base más subjetiva e individualista” (Kitano, 2007). Ahora bien, más allá de la enorme influencia ética y jurídica occidental que Japón incluso comparte con México, el país oriental ha articulado políticas y planes con respecto a la IA como la Sociedad 5.0 que siguen reconociendo valores como la armonía (Society 5.0 – ASCA Tech Lab, s. f.).

¿Pero qué pasa con México, qué aspectos lo distinguen de la ética y sistemas de valores occidentales? México, por su parte, tiene una ética y unos sistemas de valores que se han forjado a partir de su historia, su diversidad y su mestizaje. En este sentido, México es un país que ha vivido procesos de colonización, independencia, revolución y modernización, que han dejado huellas en su cultura, su política y su sociedad. Asimismo, México es un país que alberga una gran variedad de pueblos indígenas, que han conservado sus lenguas, sus tradiciones y sus cosmovisiones; finalmente, el país ha experimentado una intensa mezcla de influencias europeas, africanas y asiáticas, que han enriquecido su identidad y su expresión.

En este sentido, un estudio realizado por Figueroa Rodríguez et al (2012) llamado *Análisis de los valores que construyen la identidad del mexicano*, refleja que hasta el 2012 los valores con los que se identifican los mexicanos son los siguientes: Al agrupar variables se observaron tres grupos:

Nuestros resultados nos permiten concluir que existen valores que son compartidos, como el hecho de ser “honrado(a)” o “trabajador(a)”, como símbolos que definen al otro y son parte de las autopercepciones de los mexicanos, mientras que el resto de los valores evaluados fueron más específicos de cada subcultura, definida por sus características sociodemográficas.

Ahora bien, hay que recordar que estos parámetros y valores son sólo generalidades. En este sentido, una persona del sur de México y de Japón se puede percibir a sí misma y a sus pares de manera diferente que una persona del norte; así como las diferencias entre zonas rurales y urbanas ¿Con base en esto, cómo se comparan México y Japón con respecto a los valores de las sociedades Western, Educated, Industrialized, Rich and Democratic (WEIRD) o Occidental, educada, industrializada, rica y democrática (RARO)?

El término de sociedad WEIRD es RARA fue acuñado por Henrich (2020) para describir los sistemas de valores, pero en especial, la psicología de básicamente países occidentales. En este sentido, un estudio realizado por Beyebach et al. (2021) y titulado *Bibliometric Differences Between WEIRD and Non-WEIRD Countries in the Outcome Research on Solution-Focused Brief Therapy* coloca a México como un país no RARO (aunque si se estuviera hablando de una definición convencional de raro el resultado sería muy diferente); de acuerdo con los autores, México no se considera un país occidental, sí un país medianamente educado, sí es medianamente avanzado, con una economía emergente, una clase media, y una democracia con problemas. En el caso de Japón, los autores califican a este país como RARO dado que a pesar de que no se le considera occidental, sí es país sumamente educado, sí está altamente industrializado, tiene una sociedad que un nivel de vida alto, y una democracia funcional.

Asimismo, hay que resaltar que el comportamiento de las personas cambia. En este sentido, por ejemplo, del cambio a la modernidad en las últimas tres décadas y media ha convertido a los mexicanos en personas menos cooperativas y

dispuestas a renunciar a algo por los demás, que ahora cuidan más de sus propios intereses y están más dispuestas a ignorar o aprovecharse de los otros, así como a atribuir sus fracasos a las situaciones, el entorno que las rodea o el azar en mayor medida que a su falta de decisión o voluntad (Abitia, 2020). Esto también se ve reflejado en Japón, particularmente con los enormes problemas de despoblación.

El último aspecto que vale la pena exponer es la relación de ambos países con la tecnología. Si bien México y Japón se consideran países no occidentales, el primero no comparte (salvo en grupos étnicos muy específicos) conductas o una cosmovisión animista. Mientras que de acuerdo con Kitano (2007): “Se considera que un objeto artificial posee una identidad con su propietario. En la medida en que el propietario trate al robot (o al espíritu del robot) de manera adecuada, el robot debe tener respeto por el propietario, actuar bajo la armonización y tener un comportamiento ético”.

En suma, existen diferencias importantes entre México y Japón con respecto a sus pares occidentales. Ambos países, a pesar de que uno comparte gran cantidad de rasgos y se considera como una nación WEIRD o RARA, tienen perspectivas de valores diferentes. Quizás la diferencia más importante entre ambos se da en las perspectivas de armonía que resaltan entre los japoneses y un enorme rasgo de comunidad que aun mantienen, al menos desde su perspectiva ética, los habitantes de este país (algo que en algunas partes de México parece perderse); no obstante, quizás el rasgo más distintivo la relación con los objetos de los japoneses que si bien no alteran sus perspectivas y mapas jurídicos que comparten en buena parte con los mexicanos, sí alteran la ruta para implementar a la IA en su sociedad.

E. La sociedad mexicana y japonesa en torno a la IA

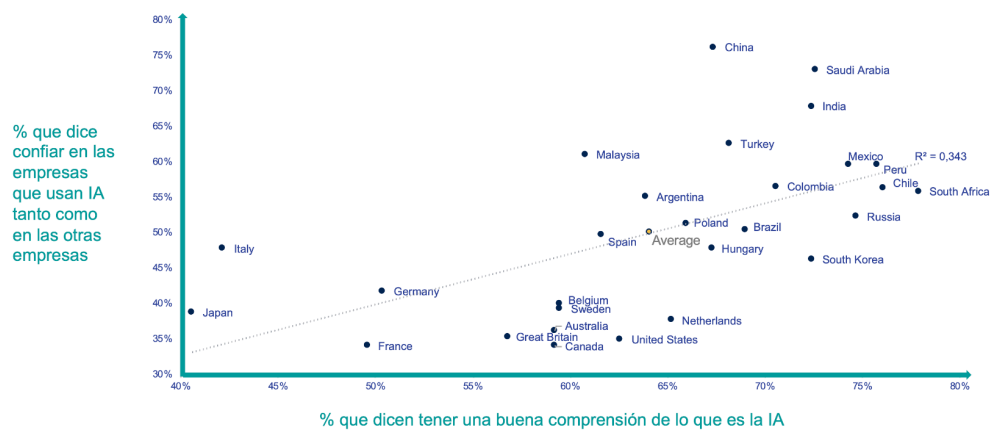
El siguiente apartado examinará la actitud de México y Japón hacia la IA. Se utilizará como base la encuesta "Opiniones globales y expectativas sobre la IA" realizada por la organización Global Advisor en 2022 para comparar la respuesta general de la población de ambos países en las siguientes categorías: nerviosismo

producido por la IA, áreas en las que se espera que la IA mejore, expectativas generales de mejora promedio de la IA, áreas que cambiarán debido a la IA, confianza en torno a la IA y comprensión o conocimiento sobre la IA. Estos aspectos tienen un impacto en la regulación y los mecanismos de respuestas con base en los marcos legales frente a los problemas sociales y de derechos humanos que pueda ocasionar esta tecnología.

En cuanto a la confianza y el conocimiento sobre la IA, el 75% de los mexicanos afirman tener una buena comprensión sobre la IA y casi un 60% dicen tener confianza en esta tecnología. En Japón, que ocupa el tercer lugar en cuanto a la IA, menos del 40% de la población afirma tener conocimiento y confianza en esta tecnología. Y al observar las gráficas de la encuesta, se puede apreciar una relación positiva entre el grado de confianza en la IA y el supuesto grado de conocimiento sobre ella; es decir, una población que dice conocer más sobre la IA mostrará más confianza en ella que una población que admite tener poco conocimiento sobre la IA.

Esto tentativamente se puede explicar a través de las diferencias culturales entre México y Japón (indulgencia y aversión al riesgo). La sociedad mexicana tiene una actitud menos fatalista ante la vida y el futuro, lo que contrasta con Japón. Lo que podría explicar por qué a pesar de tener grandes centros de investigación en IA, la población japonesa muestra más reservas en cuanto a su conocimiento sobre la IA y, en consecuencia, más desconfianza hacia esta tecnología.

LA CONFIANZA EN LA INTELIGENCIA ARTIFICIAL SE CORRELACIONA CON LA COMPRENSIÓN PERCIBIDA; AMBAS SON MAYORES EN LOS PAÍSES EMERGENTES



Base: 19,504 online adults aged 16-74 across 28 countries, Nov-Dec 2021
 Online samples in Brazil, Chile, mainland China, Colombia, India, Malaysia, Mexico, Peru, Russia, Saudi Arabia, South Africa, and Turkey tend to be more urban, educated, and/or affluent than the general population.
 The "Global Country Average" reflects the average result for all the countries and markets where the survey was conducted. It has not been adjusted to the population size of each country or market and is not intended to suggest a total result.

© Ipsos - Global Opinions and Expectations about AI - January 2022



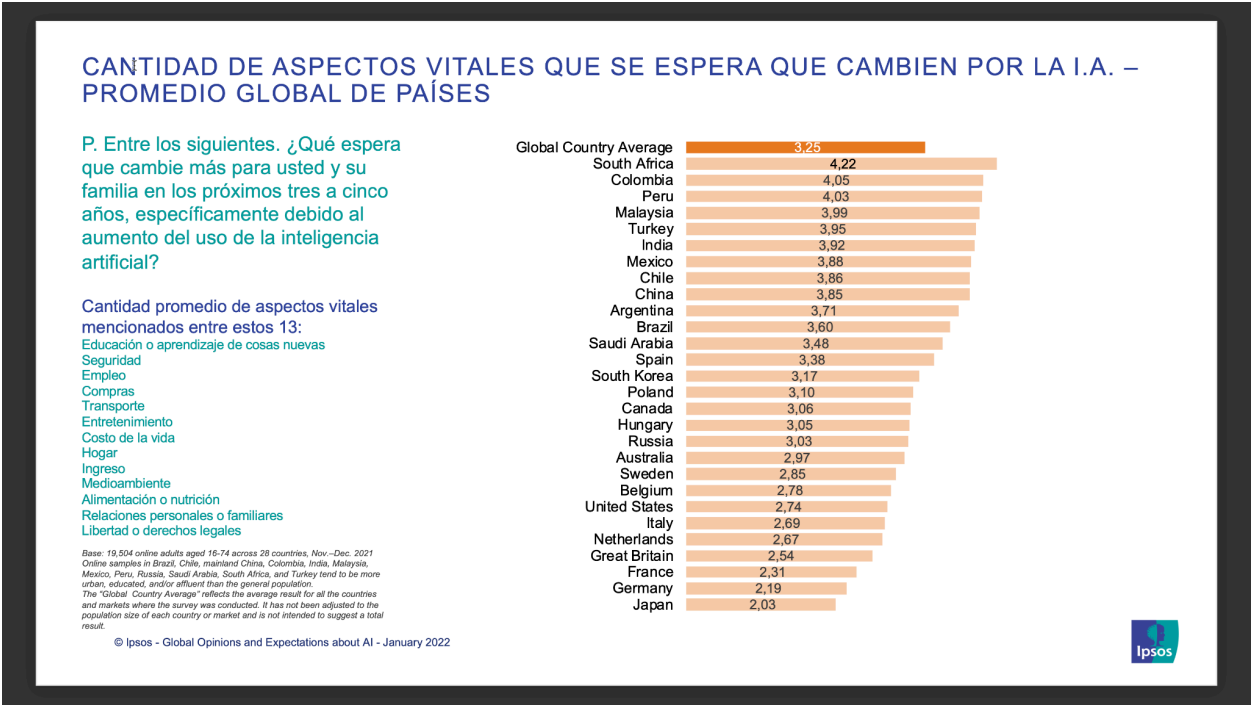
(Imagen 1, fuente: IPSOS, 2022)

En lo que respecta a las expectativas generales de mejora promedio de la IA y a las áreas que cambien debido a la IA, el 65% de los mexicanos piensan que los productos que utilizan la IA tienen más beneficios que desventajas, mientras que en Japón ese porcentaje es del 42%. Además, el 60% de las personas mexicanas confían en las empresas de IA tanto como en otras empresas, mientras que en Japón ese porcentaje disminuye a un 39%. A partir de estos datos, se puede apreciar una actitud más reservada de la sociedad japonesa hacia la IA, que parece estar más relacionada con sus características sociales que con sus perspectivas. En el caso de México, su alto nivel de indulgencia parece hacer que su sociedad sea más propensa a confiar en nuevas tecnologías y en los actores que las manejan (Imagen 1).

En cuanto a las áreas que cambiarán debido a la IA, las personas mexicanas respondieron que cambiarán 3.88 de los trece aspectos planteados en la encuesta (educación, seguridad, empleo, compras, transporte, entretenimiento, costo de vida, hogar, ingreso, medio ambiente, alimentación, relaciones personales y libertad o derechos legales). En Japón, las personas que respondieron la encuesta indicaron que en promedio solo el 2.03 de los trece aspectos cambiará debido a la

incorporación de la IA. Nuevamente se observa una actitud más cautelosa de la población japonesa en comparación con la mexicana, a pesar de que esta última tiene una cosmovisión en la que los objetos y la naturaleza forman parte del entorno social

En lo que respecta a las áreas que mejorarán debido a la IA, la mayoría de los mexicanos indicaron que estas serán la educación, el entretenimiento, el transporte y las compras. En el caso de Japón, la mayoría de las personas entrevistadas sugirieron que las áreas del hogar y las compras serán las que más se beneficien de la IA. Estos datos muestran en qué áreas hay un mayor entusiasmo por parte de la población mexicana y cuáles son los pocos sectores que entusiasman a la sociedad japonesa en torno a la IA (Imagen 2).



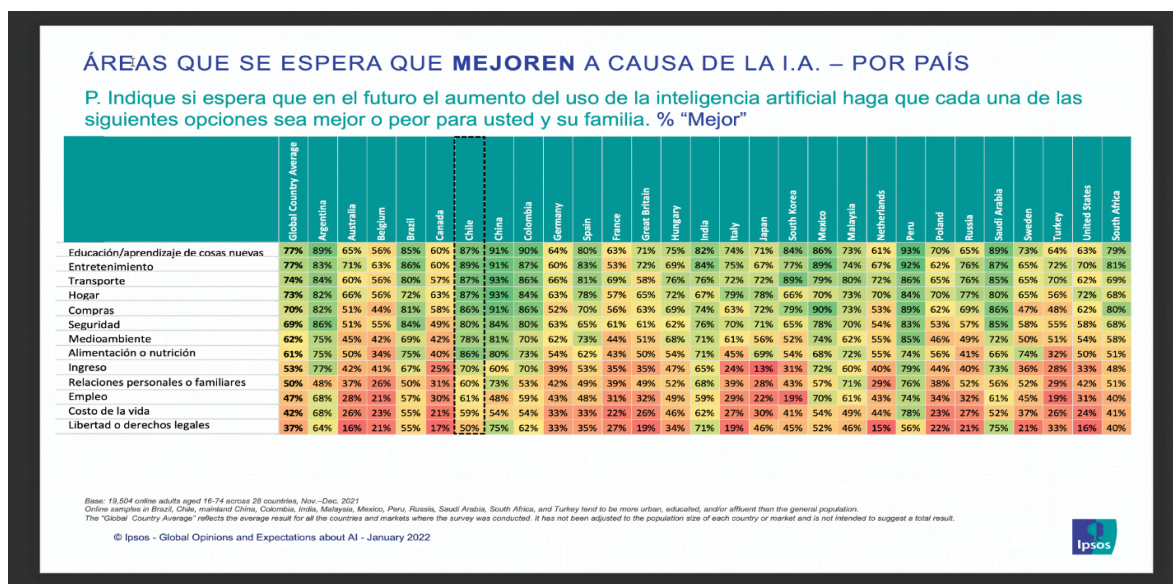
(Imagen 2, fuente: IPSOS, 2022)

El último punto para comparar es el nerviosismo que producen los productos que utilizan IA. El 38% de la población mexicana entrevistada comentó que estos productos los hacen sentir nerviosos, mientras que en Japón solo el 20% de las personas que respondieron la encuesta admitieron sentirse nerviosas con

respecto a los productos que utilizan IA, lo cual es el resultado más bajo dentro de la encuesta. Por lo tanto, se puede asumir que, a pesar de las diferencias culturales entre ambos países, la población japonesa muestra menos nerviosismo en torno a la IA, lo cual se puede explicar a través de su perspectiva ética particular (rinri), en la que los objetos forman parte del tejido armónico social.

En síntesis, se puede apreciar que la población mexicana tiene mayor confianza y conocimiento sobre la IA que Japón, a pesar de que este último sea uno de los países más desarrollados en esta área de investigación. En cuanto a los sentimientos que evoca la IA, la sociedad mexicana, según la encuesta, tiene más nerviosismo que la japonesa, que es la que menos muestra recelo hacia esta tecnología; y esto ayuda a explicar la presión de la sociedad japonesa para crear un entorno que regule la IA y la incorpore al tejido social desde una perspectiva humana. No obstante, se tiene que señalar que la regulación de la IA debe basarse en los derechos humanos y tener en cuenta los posibles problemas sociales y de derechos

humanos que pueda ocasionar esta tecnología (Imagen 3).



(Imagen 3, fuente: IPSOS, 2022)

5. Metodología

Es necesario encontrar la metodología adecuada para un trabajo con tantos componentes disciplinarios de análisis. En particular, se propone utilizar el Análisis de Constelación junto con la teoría de objeto de contorno, como propuso Feldhoff (2019), ya que este método permite interpretar la trayectoria de la tecnología desde una perspectiva multidisciplinaria. Como todo método, necesita fuentes de información; según se mencionó en la metodología de investigación, se utilizarán fuentes, documentales y se hará un trabajo de campo digital solicitando información al Instituto Nacional de Transparencia y Acceso a la Información Pública (INAI) con respecto a la plataforma C5. Todo sobre la creación de un mapa de constelaciones basado en datos empíricos que puedan analizarse de forma holística para obtener un resultado.

El método de análisis de constelaciones es importante porque está relacionado con la teoría. Al examinar la trayectoria de innovaciones de procesos tecnológicos y sociales, se observa que el método de análisis de constelaciones comparte rasgos con la teoría de la red de actores de la que se inspiró. Esto se refuerza por sus orígenes, ya que “...fue desarrollado como una herramienta para la investigación interdisciplinaria y transdisciplinaria sobre tecnología, sostenibilidad e innovación” (Ohlhorst y Schön, 2015).

El método de análisis de constelaciones se utilizará para responder a una pregunta sobre un fenómeno tan complejo como la IA. Junto con las fuentes y características metodológicas mencionadas en la introducción, ayudará a comprender el objeto de estudio a partir de sus elementos multidisciplinarios. Sin embargo, es importante explicar qué es el análisis de constelaciones para destacar su importancia en este trabajo, y todo esto será entrelazado y explicado con los planteamientos de Boltanski y la Teoría de actor-red, de donde se desprende en método de análisis de constelaciones, desde la mirada de Latour.

Entre las disciplinas enfrentadas y se confrontan en torno al objeto de estudio y la pregunta de investigación están la ética, la sociología, la antropología digital, el derecho, la economía y las ciencias de datos (especialmente sus conceptos

teóricos relacionados con la IA). Es importante abordar la investigación con herramientas teóricas y metodológicas que facilitan la traducción entre las ciencias sociales y humanidades, las ciencias de la información y entre quienes forman parte del comité tutorial, para proporcionar un producto sintético y ordenado.

A. Análisis de constelaciones

El método de análisis de constelaciones es un enfoque de investigación de la innovación que busca comprender trayectorias extremadamente complejas de innovación y difusión (Ohlhorst y Susanne, 2015). Se caracteriza por examinar factores heterogéneos que convergen y tiene una perspectiva de abajo hacia arriba similar a la teoría fundamentada. Es importante mencionar que esta aproximación metodológica, como lo mencionan Ohlhorst y Susanne (2015), se inspiró en la teoría del actor-red.

De acuerdo con Ohlhorst y Susanne (2015): "El análisis de constelaciones se basa en la conclusión de que los objetos sociales empíricos de investigación no pueden entenderse sin los artefactos técnicos y los sistemas de infraestructura interconectados, y viceversa". Esto es de suma importancia para la presente tesis de investigación, ya que examina la trayectoria de los componentes éticos ligados a la IA, un artefacto sumamente técnico. Además, el "análisis de constelaciones se basa en la conclusión de que los objetos sociales empíricos de investigación no pueden entenderse sin los artefactos técnicos y los sistemas de infraestructura interconectados, y viceversa" (Ohlhorst y Susanne, 2015).

Feldhoff et al. (2019) plantean que el análisis de constelaciones incluye cuatro tipos de elementos (actores, elementos técnicos, elementos simbólicos y elementos naturales) y "diferentes tipos de relaciones entre esos elementos (por ejemplo, relaciones simples, dirigidas, faltantes o conflictivas)". Estos elementos sirven como base de referencia para mapear la constelación y luego analizarla una vez dibujada la trayectoria o mapa.

El análisis de constelaciones tiene dos elementos importantes para abordar la respuesta a la pregunta de investigación. El primero es la parte visual, que se

basa en la creación de gráficas o constelaciones a partir de conceptos clave. El segundo es el análisis de esas constelaciones para comprender el objeto de estudio basándose en definiciones que pueden ser comprendidas por todas las disciplinas que lo componen.

Ohlhorst, Dörte; Schön, Susanne (2015) mencionan dos aproximaciones diferentes en el análisis de constelaciones. La primera es la perspectiva del análisis reconstructivo, que examina la trayectoria de innovación de un objeto o fenómeno de estudio en función de todos los factores que le rodean. Estos factores (técnicos, sociales, económicos, entre otros) son los que le permitieron sobrevivir a la innovación.

La segunda perspectiva del análisis de constelaciones es la perspectiva del análisis estratégico, que se centra en la innovación y examina los diferentes procesos estratégicos que permitieron esa innovación: uso prospectivo para sobrevivir y las innovaciones complementarias. De esta manera, se examina la trayectoria de un proceso de innovación a partir de la constelación de estrategias que lo llevaron al éxito. Por ejemplo, las tecnologías emergentes que convergieron, la forma en que se utilizaron esas tecnologías y la visión prospectiva para emplearlas en el proceso de innovación de un producto u objeto.

El fundamento de la tesis es el resultado interpretativo del análisis de constelaciones. Esta metodología combina dos visiones: retrospectiva y prospectiva. La retrospectiva hace una revisión bibliográfica de un proceso tecnológico o de un objeto, y la prospectiva crea un escenario de lo que puede ocurrir, un diagnóstico.

De esta manera, desde la perspectiva retrospectiva, se puede indagar el origen de la IA a partir de una constelación de actores que expliquen su relevancia. Además, en otro apartado se puede examinar el surgimiento de objetos regulatorios de la IA en Ciudad de México y Tokio. Finalmente, en un tercer apartado se buscará observar el origen de la implementación de los principios de la IA en dos espacios geográficos determinados.

La parte prospectiva implica hacer una predicción basada en el análisis de la trayectoria. Esto también sirve para hacer recomendaciones sobre los objetos centrales de esta investigación, articulándolos con la noción de objeto de frontera o de contorno. Así, al analizar las tres trayectorias y sus constelaciones sobre la IA (IA, principios éticos, y ética aplicada a la IA), se podrá responder si los objetos regulatorios éticos de la IA (regulaciones) son producto de la trayectoria de la IA y de los principios éticos, y se podrán hacer propuestas legislativas para el caso de la Ciudad de México.

El método de análisis de constelaciones ligado a la noción de objeto de contorno y frontera sirve para interpretar la trayectoria de los principios éticos de la IA a partir de variantes de un conjunto más amplio. Además, al ser un método que se basa en el análisis de abajo hacia arriba, ayuda a generar una interpretación robusta basada en conjuntos que pueden parecer dispersos pero que adquieren orden a partir de la traducción de estos. Por las diversas disciplinas que encierra la pregunta de investigación, el método de análisis de constelaciones es la herramienta ideal para examinar, interpretar y responder a la pregunta de investigación.

B. Análisis de constelaciones y el objeto de contorno

Dado que la pregunta de investigación tiene una naturaleza inter y transdisciplinaria, se requiere una herramienta metodológica adecuada. Esto llevó a la búsqueda de diferentes elementos teóricos y fundamentos metodológicos que inicialmente parecían inconexos. Sin embargo, al examinar con mayor profundidad el concepto de objeto de contorno planteado por Star y Griesemer (1989), se encontró un texto publicado por Feldhoff et al. (2019) se propone usar el concepto de objeto de contorno junto con el análisis de constelación, lo que ayuda a articular las diferentes perspectivas de la hipótesis planteada en este trabajo.

Hay que definir qué es un objeto de contorno antes de continuar como se usará el planteamiento de Feldhoff et al. De acuerdo con Star y Griesemer: “Los objetos de frontera son objetos que son lo suficientemente plásticos para adaptarse a las necesidades locales y las limitaciones de las distintas partes que los emplean,

pero lo suficientemente robustos para mantener una identidad común en todos los sitios". Y como plantean las autoras, crear y mantener un objeto de contorno es fundamental para mantener y sostener la cohesión social.

Ahora bien, de acuerdo con Feldhoff et al. (2019): "El análisis de constelaciones comparte un terreno común con la formulación de un objeto límite, como se muestra arriba. Tanto el concepto de objeto límite como el análisis de constelaciones se basan en una simetría metodológica entre factores materiales y biofísicos, por un lado, y factores sociales, por otro". Así, las autoras buscaron acercarse al problema de investigación sin privilegiar una disciplina sobre otra, lo cual es muy importante en esta tesis por la constelación de elementos que van desde las tecnologías de la información y la comunicación hasta entornos y contextos sociales.

Las autoras (Feldhoff et al., 2019) primero hicieron un mapa de diferentes constelaciones identificando varios elementos que caracterizan el problema de estudio. Luego, se definió y operacionalizó un objeto de contorno o frontera, lo que permitió tener un marco para mapear una constelación a partir de la definición de un objeto de contorno que sirve para anclar conceptos clave o nucleares. Por ejemplo, se definieron conceptos clave como redes neuronales, cuasi objeto y derechos, y se articularon alrededor de un objeto de contorno que será la IA.

La ventaja de usar en conjunto el concepto de frontera y el método de análisis de constelaciones, según lo planteado por las autoras, es que permite:

"Si bien el enfoque del objeto límite nos permite desarrollar un encuadre de problema compartido respaldado por un marco teórico, el análisis de constelaciones proporciona la base metódica para un proceso iterativo y autorreflexivo de integración de conocimientos de investigación y, por lo tanto, mapeando la constelación definida por el objeto límite" (Feldhoff et al, 2019).

En su trabajo, Feldhoff et al (2019) proponen tres pasos para aproximarse al análisis de constelaciones:

1. Mapear la constelación identificando los diferentes elementos de un problema.

2. Elaborar los principios fundamentales de la constelación: “Este paso aborda cuestiones relativas a actores especiales o dominantes, alianzas ocultas o relaciones de poder, lo que significa que el análisis es levantado a un nivel más conceptual, siguiendo las relaciones más descriptivas entre los elementos” (Feldhoff et al, 2019).
3. Hallar los procesos de transformación que alteran la constelación (Feldhoff et al, 2019).

En cuanto a los objetos de contorno, Feldhoff et al (2019), plantean dos usos para estos:

1. Los objetos como Flexibilidad interpretativa, o sea, que los de contorno están abiertos a diferentes interpretaciones. La estructura organizativa, que contiene un límite, un espacio de frontera que contiene un objeto en común entre las personas. Dinámica entre usos mal estructurados y más adaptados de los objetos: sus significados con más estrechos, más acotados, entre un mismo contexto social. Esto provoca que los objetos se vuelven más elásticos, más amplios, conforme entran a distintas constelaciones sociales.
2. Como herramienta de cooperación: básicamente lo que Star y Griesmar plantean como traducciones, ya que “un objeto límite facilita el diálogo entre diferentes actores y permite diferentes prácticas referidas al objeto límite” (Feldhoff et al, 2019).

Quizás lo que diferencia este trabajo del realizado por Feldhoff et al (2019) es que se trata de una tesis. Esto requiere una serie de constelaciones para analizar la trayectoria de la ética en la IA y su aplicación normativa en Ciudad de México y Tokio. Por lo tanto, además de los pasos planteados por las autoras, se realizará una constelación de destino, que busca integrar las distintas trayectorias dentro de un todo coherente (Ohlhorst y Schön, 2015).

Así pues, se propondrá a la IA como objeto de contorno o frontera que servirá para organizar los conceptos clave de una constelación. Luego se identificarán los procesos de innovación de esa constelación y se definirán los casos empíricos a

analizar. Finalmente, se creará una constelación de destino como conclusión, en la que se integrará el análisis de las distintas constelaciones (trayectoria de la IA, ética y la ética en la IA, derechos humanos) para responder la pregunta de investigación.

C. Ruta metodológica

La metodología propuesta consiste en analizar cómo se aplican los derechos humanos y un marco ético universal en la legislación y regulación de la IA en plataformas de movilidad en México y Japón a través de casos empíricos. Esto es, con la intención de evaluar el nivel de implementación de los derechos de privacidad, acceso a la información y no discriminación en relación con los problemas de la IA en estos países. Para ello, se elaborarán varios capítulos que aborden estos problemas en profundidad.

En el primer capítulo se abordará la teoría social necesaria para entender el contexto en el que se enmarca la investigación. Se discutirán conceptos claves de la obra de Latour y Boltanski, y algunos problemas valorativos con respecto a la IA; particularmente, el de la responsabilidad. Además, se examinarán algunas de las implicaciones sociales con respecto a los problemas de la IA.

El segundo capítulo profundizará la explicación de los términos y los conceptos de la IA desde las ciencias de datos. Además, se examinará cómo la IA se ha vuelto un término común en la discusión pública y cómo ha influenciado la forma en que se entiende y se habla sobre la tecnología. Pero principalmente, se hará énfasis en traducir los términos y conceptos de la IA a un lenguaje que se pueda aplicar al análisis de los marcos regulatorios con respecto a la teoría social de la teoría de actor-red y la perspectiva de resolución de conflictos desde Boltanski.

El tercer capítulo abordará la historia del aprendizaje de máquina y cómo ha evolucionado a lo largo del tiempo. Se examinarán algunos de los principales precursores del aprendizaje de máquina y cómo han contribuido al desarrollo de la IA. Además, se discutirán algunos de los avances y logros importantes en el

aprendizaje de máquina y cómo influyeron en la forma en que se entiende y utiliza la IA actualmente.

Para el capítulo cuarto se utilizará la trayectoria de los derechos humanos en torno a la IA: regulación y derechos humanos. Se examinará la trayectoria de los principios éticos de la IA y cómo han evolucionado a lo largo del tiempo. Se discutirá la importancia de la regulación y de la protección de los derechos humanos en el contexto de la IA y cómo se han abordado estos temas en el pasado. Además, se explorará la forma en que los principios éticos de la IA se relacionan con el marco ético universal y cómo se aplican y pueden ser aplicados a la regulación de la tecnología en las propuestas regulatorias de México y Japón en torno a la Unión Europea.

Finalmente, en el capítulo quinto se examinará el impacto de la IA en México y Japón, tanto desde una perspectiva positiva como negativa. Se discutirán algunos de los problemas y desafíos que surgen al utilizar la IA, incluyendo cuestiones de privacidad, acceso a la información y no discriminación. Además, se explorará quién es responsable de resolver estos problemas y cómo se está abordando la regulación de la IA en ambos países. Y se discutirá también la agencia de las personas y las comunidades en el contexto de la IA y cómo pueden influir en el uso y la regulación de la tecnología.

Además, en ese capítulo se realizará una investigación documental sobre los casos empíricos de Uber en México y las controversias legislativas en torno a los casos empíricos en Japón. Además, se llevará a cabo un trabajo digital solicitando documentos a través del Instituto Nacional de Acceso a la Información en México (INAI) para investigar cómo funciona la plataforma C5 en relación con la IA. A lo largo de toda la tesis, se utilizará el método de análisis de constelaciones junto con el objeto de contorno.

La investigación tendrá un enfoque acotado a los tiempos planteados para la presentación de la tesis (cuatro años) al tener en cuenta las condiciones de distanciamiento social y los plazos para entregar la tesis y los informes derivados. Dado que es importante recolectar, examinar e interpretar información para

comprender el fenómeno de los sociales que ocasiona la IA respetando los tiempos de trabajo que plantean los programas de doctorado en México. Como menciona Creswell (2014), se asume que recolectar información en mayor o menor medida nos permitirá hacer una investigación integral y tener una mayor comprensión del problema de estudio.

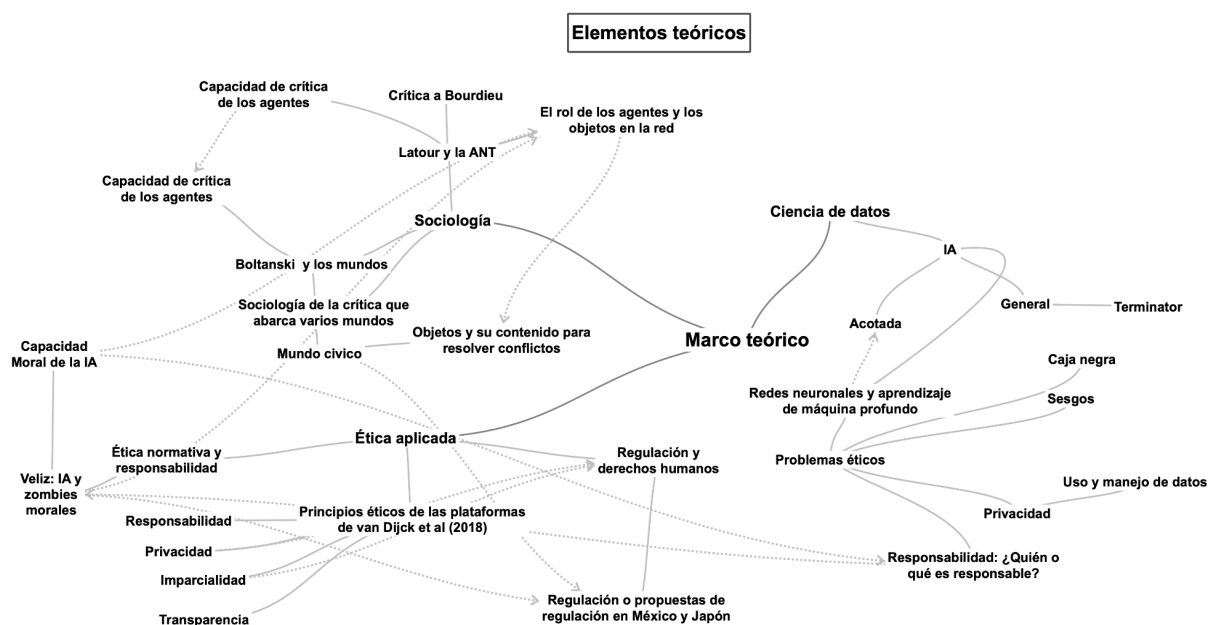
D. Los problemas sociales en México en Japón y la IA

Siguiendo la ruta planteada por el método análisis de constelaciones, el capítulo final, la conclusión de la tesis, buscará responder la pregunta de investigación a partir del análisis de los problemas sociales de la IA en casos empíricos comparativos en México y Japón. En Japón, se examinarán los problemas sociales en torno a la transgresión de los derechos humanos como la privacidad, acceso a la información y no discriminación, así como los mecanismos de respuesta con base en su regulación y propuestas regulatorias. En México, los casos empíricos se centrarán en el ecosistema de plataformas Uber y el sistema de vigilancia C5, donde se examinará si estos sistemas que utilizan IA transgreden los derechos humanos de las personas y causan problemas sociales.

Con el objetivo de enfocar el análisis de este capítulo en la pregunta de investigación, es necesario definir los campos del conocimiento que nos ayudarán a interpretar lo que se pretende responder en este apartado. Estos campos del conocimiento incluirán la política, la sociedad, el derecho y la regulación, la cultura y la ciencia de datos, manteniendo un formato coherente con el resto de la tesis. De esta manera, se examinarán las condiciones sociales que pueden influir en la trayectoria de la aplicación de la ética en la IA, pero de manera limitada.

Este capítulo examinará los problemas sociales ocasionados por las plataformas sociales en México y Japón. Además, analizará los mecanismos existentes o necesarios para abordar las controversias relacionadas con la IA en cuanto a derechos de acceso a la información, privacidad y no discriminación, así como problemas de responsabilidad y comprensibilidad. Para hacerlo, se examinarán casos empíricos en Japón, como Google Maps, y en México, como Uber y la plataforma C5.

De acuerdo con lo anterior, la tesis tratará de responder la pregunta de investigación a través de un análisis de elementos ordenados en constelaciones, dividido en niveles macro, meso y micro, y concluyendo con la interpretación de una constelación final. Cada capítulo proporcionará evidencia que se integrará en la parte final de la tesis para dar una respuesta holística a la propuesta planteada. Además, se realizará una investigación documental y empírica en torno al C5 para proporcionar un sustento empírico al analizar los casos con base en la teoría social que contribuirán a responder la pregunta de investigación en torno a los derechos humanos.



(Imagen 4, de autoría propia)

CAPÍTULO PRIMERO

MARCO TEÓRICO: UNA REFLEXIÓN SOCIAL EN TORNO A LA IA

Para abordar esta investigación es pertinente examinar dos categorías de conceptos teóricos. Primero, los conceptos éticos ligados a la IA y su relación con los problemas sociales que ocasiona esta tecnología. Segundo, las nociones sociológicas que servirán en torno los argumentos metodológicos, así como para describir los elementos con las se tratará de conocer al objeto de estudio.

En la parte sociológica, primero se partirá de un marco teórico holístico ligado a la Teoría de actores red desde la perspectiva de Latour. Asimismo, en torno a la teoría de actor-red, se abordará cómo los conceptos de programa y antiprograma pueden explicar los conflictos éticos entre las normas de los programas y la respuesta de los actores humanos y no humanos ante estas. Todo esto alrededor las herramientas del análisis interdisciplinario esbozadas en el Método de análisis de constelaciones que se centra en examinar asociaciones entre actores y los conjuntos en esas asociaciones.

Hay que resaltar, se examinarán los conceptos que ayuden a comprender la ética y la perspectiva japonesa a partir de las nociones de animismo, tecno-animismo y animismo postmoderno en los textos de Allison, Jensen y Blok, Kitano, Honda y Yoneyama.

A efecto de esclarecer de forma breve algunos de los conceptos que se van a emplear, primero es pertinente explicar qué es la Teoría de actores red. De acuerdo con Latour, esta teoría "...afirma que simplemente no deberíamos creer que la cuestión de las conexiones entre actores heterogéneos esté cerrada, que lo que normalmente se entiende por *social* probablemente tenga que ver con el reensamblaje de nuevos tipos de actores" (Latour, 2005). Esto es importante dado que desde la perspectiva de este trabajo se considera que los problemas sociales que las plataformas e instrumentos que emplea la IA forman parte de una enorme red de asociaciones humanas en la que los objetos forman parte.

La siguiente perspectiva sociológica que es importante explicar es la que está relacionada con el modelo de equivalencias. Esta teoría, de acuerdo con Boltanski y Thevenot, explora una sociología de la crítica donde las personas no son actores

pasivos frente a una injusticia. Es decir, que los seres humanos no son simples víctimas pasivas de la sociedad, el estado o las situaciones bajo conceptos como el *habitus*; por el contrario, tienen la capacidad de reclamar una injusticia en distintos regímenes o escenarios con la ayuda de sujetos o objetos de autoridad.

Hay que resaltar el concepto de *techno animismo*. Considerando la perspectiva de esta tesis, el concepto sirve tanto para explicar la relación de la gente japonesa con los objetos y para complementar algunos aspectos en la teoría de actores red. De acuerdo con Jensen & Block: “Nuestra sugerencia es que el *tecno-animismo* japonés nos permite identificar diferentes modos de convivencia humano-no humano, infundiendo así nueva energía al análisis de las no modernidades fuera de la órbita euroamericana” (Jensen & Blok, 2013).

Finalmente, hay que explicar algunas definiciones en torno a la IA. El doctor Marco Gillies plantea que la IA es “...el área de la informática que intenta hacer máquinas que puedan replicar la inteligencia humana para realizar tareas que los humanos pueden hacer, pero sí históricamente ha sido muy difícil para las máquinas” (Gillies, en Coursera, s.f). Asimismo, la IA se divide en varias áreas y subcampos como lo es el Aprendizaje de máquina donde se centrará este trabajo.

El aprendizaje de máquina, que funciona mediante el uso de grandes volúmenes de datos (*Big Data*), caracterizados por su gran volumen, diversidad y estructura compleja, enfrenta desafíos significativos en términos de almacenamiento, análisis y visualización (Yaseen & Obaid, 2020). Esta tecnología permite a los sistemas de inteligencia artificial aprender y mejorar mediante la interacción con funciones matemáticas complejas, como las redes neuronales artificiales, lo que ha dado lugar a aplicaciones de alto impacto social, como la asistencia en diagnósticos médicos y el desarrollo de ciudades inteligentes. Estas últimas se fundamentan en principios de sustentabilidad y utilizan la recolección e interpretación masiva de datos para optimizar su funcionamiento (Hoon Jo et al., 2019). Sin embargo, estas innovaciones también presentan riesgos importantes, como el sesgo algorítmico; por ejemplo, “la IA puede estar sesgada y discriminar injustamente a minorías u otros grupos vulnerables” (Ng, en Coursera, s.f.).

Y para atacar esos problemas, diferentes académicos e instituciones han desarrollado una serie de valores normativos en torno a la IA, los cuales están relacionados con algunos derechos humanos. Aun así, esos valores pueden estar supeditados a las prácticas éticas y nociones socioculturales de las naciones que los desarrollan; en particular, la sociedad china permite el uso de la IA para generar un puntaje social que sirve para penalizar el acceso a ciertos productos o servicios. Por consiguiente, vale la pena analizar la forma en la que surgen esos problemas desde una perspectiva sociológica centrada en distintas asociaciones, así como mecanismos de resolución de conflicto con base en la definición de responsabilidad de las transgresiones de la IA en torno a una perspectiva de derechos humanos.

1. Inteligencia Artificial

La Inteligencia Artificial, el aprendizaje automático profundo y los algoritmos asociados, así como las redes neuronales relacionadas con estas tecnologías, generan problemas importantes en torno a los derechos humanos. Estas problemáticas, como la privacidad, la transparencia y la comprensibilidad, se evidencian en una revisión de la historia y las aplicaciones de esta tecnología. Por lo tanto, no es sorprendente que investigadores como Van Dijck et al (2018) hayan planteado una serie de valores en torno a las plataformas: valores que tienen una estrecha relación con los derechos de a la privacidad, no discriminación, y acceso a la información.

Sin embargo, antes de abordar estos problemas de derechos en el análisis de la trayectoria ética en la IA, es importante desglosar algunos conceptos y términos relacionados con ella. Esto es, para hacer una especie de traducción para que este lenguaje pueda entenderse para cualquier lector de este trabajo, y para que estos conceptos se apliquen en el lenguaje de las ciencias sociales y humanas. Por lo tanto, primero hay que centrarse en el concepto central de esta tesis: la IA.

La IA es un término complejo definido de diversas maneras. Según Saeed y Al-Mushayt (2020), se refiere a "la capacidad de una computadora para imitar la inteligencia del comportamiento humano mientras mejora su propio desempeño". Por otro lado, Gillies (en Coursera, s.f.) lo describe como "el área de la informática que intenta hacer máquinas que puedan replicar la inteligencia humana para realizar tareas que los humanos pueden hacer, pero sí históricamente ha sido muy difícil para las máquinas". Ambas definiciones sugieren que la IA se basa en la imitación de la inteligencia humana.

Sin embargo, Saeed y Al-Mushayt (2020) señalan que también puede ser entendida como "un comportamiento inteligente de una máquina autónoma que describe el cerebro de la máquina y no su cuerpo; puede conducir un automóvil, jugar y realizar diversos trabajos sofisticados". Y es relevante mencionar que la idea de la IA, su concepto y desarrollo, no es nueva. La Real Sociedad (2017) explica que en 1952 el investigador Arthur Samuel creó una máquina de aprendizaje temprano, que logró aprender a jugar a las damas usando guías anotadas por expertos humanos y juegos que jugó contra sí mismo para distinguir los buenos movimientos de los malos.

El término de IA plantea la pregunta de qué es en realidad. A menudo se asocia con animales que poseen una gran capacidad de razonamiento, autodeterminación y ciertas habilidades como la creatividad o el lenguaje, que son características comunes de los seres humanos. Sin embargo, como señala Descartes, "no hay hombres tan aburridos y estúpidos, ni siquiera idiotas, que sean incapaces de unir palabras diferentes y construir así una declaración por la que hacer entender sus pensamientos" (Descartes, 2012). En este sentido, Ertel (2019) argumenta que "a veces se argumenta que los humanos nunca lograrán construir una máquina que sea más inteligente que su inventor, pero este argumento ignora el hecho de que las máquinas son capaces de aprender" (énfasis propio).

La reflexión de Ertel (2019) se enfoca en el estado actual de la IA, la cual ha logrado desarrollar tareas propias del ser humano como el aprendizaje, la creación

de música y el reconocimiento de imágenes, pero todavía se encuentra limitada a realizar correctamente solo una de estas tareas. Ertel (2019) destaca que "actualmente, la capacidad de aprendizaje todavía se limita a tareas específicas, sin embargo, no se puede descartar el desarrollo de algoritmos que lleven a máquinas capaces de aprender a un nivel meta y, por lo tanto, continuar evolucionando".

Para comprender el estado actual de la IA, hay que diferenciar entre dos tipos de inteligencia: la débil y la fuerte. La IA estrecha o débil es aquella que se especializa en realizar un solo tipo de tarea con excelencia. Según Andrew Ng, fundador de la plataforma educativa Coursera, esta es el área en la que se tienen los mayores avances en la actualidad. Ng (en Coursera, s.f) señala que la IA estrecha "se refiere a IA que realiza una sola cosa, como un altavoz inteligente, un automóvil autónomo, IA para búsquedas web o aplicaciones de IA en la agricultura o en una fábrica".

La IA estrecha, también conocida como IA débil, se especializa en realizar una sola tarea con gran precisión. Según Valverde (2019), "La IA estrecha aborda problemas específicos y avanza rápidamente en áreas como automóviles autónomos, juegos y robótica" (Valverde, 2019). Esta forma de IA ya tiene implicaciones significativas en la vida diaria, como la generación de rostros humanos, la creación de pinturas a partir de texto, la recomendación de contenido en plataformas de streaming y el apoyo en diagnósticos médicos y desarrollo de vacunas.

Por otro lado, la IA general, conocida como IA fuerte, puede realizar varias tareas a la vez. Según Searle (1980, citado por Müller, 2020), "las computadoras con los programas adecuados pueden comprender y tener otros estados cognitivos" (Searle, 1980, 417). La idea de la singularidad, según Müller (2020), es que una vez que se desarrolle la IA general, se desconoce lo que podría suceder. Müller (2020) define la singularidad como "el punto en el que los sistemas de IA superarían el nivel humano de inteligencia y serían 'súper inteligentes' ".

En síntesis, el desarrollo y la investigación en IA general es considerado un objetivo emocionante por investigadores como Andrew Ng, pero se necesitarán avances tecnológicos significativos antes de alcanzarlo. Hay dos tipos de IA: la acotada, que ha tenido avances trascendentes en diferentes subcampos y requiere de herramientas para funcionar, y la general, tan compleja como la inteligencia humana o incluso superarla.

2. Transgresión, sistemas y riesgos

La transgresión vincula los impactos de la IA en torno los derechos de Privacidad, acceso a la información y no discriminación; así como los enormes problemas de explicabilidad y responsabilidad derivada de esto. El concepto de transgresión también se aborda como un paraguas que contiene una tormenta en medio de la noche, principios elementales para generar confianza alrededor de los sistemas de IA. Esto es, en torno a elementos como la equidad, rendición de cuentas y transparencia. Y finalmente, la noción de la transgresión engloba debates alrededor del impacto social de esta tecnología, con respecto a determinar quién es responsable, y la forma en qué se pueden evaluar los riesgos y hacer propuestas regulatorias, auto-regulatorias, y de resistencia con respecto a la IA.

De acuerdo con la Real Academia Española de la Lengua (s.f), se puede definir Transgresión como la acción y el efecto de transgredir, es decir, de incumplir o violar una norma, una ley, un principio o un deber. Asimismo, la transgresión transgresión implica sobrepasar los límites materiales, sociales y simbólicos establecidos hasta el momento, ampliando el espacio de acción, quebrando los tabúes, y desafiando lo que el individuo es y lo que le pertenece (Kozielecki, 1997, p. 10; Mamzer, 2006); por ejemplo, una infracción, una vulneración, un quebrantamiento, una desobediencia, un atropello, un atentado o un delito, según la gravedad o la intención de la conducta; lo contrario de la transgresión es el acatamiento, el cumplimiento o la observancia de lo establecido.

Surge una pregunta: ¿cuáles son los niveles de transgresión de la IA? En este trabajo se proponen cuatro niveles: a nivel de derechos humanos, a nivel de asociaciones humanas en torno a la teoría de actor-red, a nivel técnico, y a nivel sociotécnico:

1. A nivel de derechos humanos, se refiere a las transgresiones que afectan a los derechos fundamentales de las personas, como el derecho a la vida, a la libertad, a la igualdad, a la privacidad, a la educación, a la salud, a la participación, a la cultura, entre otros aspectos. Estas transgresiones pueden ser cometidas por la IA, por las entidades que la desarrollan o la usan, o por las personas que interactúan con ella. Por ejemplo, se puede mencionar el caso de los sistemas de reconocimiento facial que pueden violar el derecho a la privacidad o a la no discriminación de las personas, o el caso de los sistemas de IA que pueden afectar el derecho al trabajo o a la educación de las personas.
 - a. Transgresión a la privacidad: Un caso reciente en Palma, España, involucró a un menor de 15 años que utilizó una aplicación de IA para generar imágenes manipuladas de compañeras de clase, haciéndolas parecer desnudas, y luego las difundió entre sus amigos. Este incidente destaca cómo la IA puede ser empleada para violar la privacidad y dignidad de las personas, especialmente de menores (García, 2025). Y este tipo de casos también han ocurrido en países como México (Velázquez, 2023).
 - b. Transgresión a la no discriminación: En Suecia, una investigación reveló que la Agencia del Seguro Social utilizaba sistemas de IA que seleccionaban injustamente a grupos marginados para inspecciones sobre fraude en prestaciones. Estos sistemas discriminaban por género, origen extranjero, nivel de ingresos y nivel educativo, lo que llevó a Amnistía Internacional a exigir la interrupción inmediata de su uso (Amnistía Internacional, 2024).

- c. Transgresión al derecho de acceso a la información: La plataforma DeepSeek, propiedad de China Mobile, ha sido señalada por rastrear y almacenar información detallada de sus usuarios sin su conocimiento explícito. Este tipo de prácticas opacas impide que los usuarios comprendan cómo se recopilan y utilizan sus datos, afectando su derecho al acceso a la información y a la privacidad (El País, 2025). Asimismo, muchas empresas mantienen cerrado el acceso al funcionamiento de sus modelos, como es el caso de OpenAI y ChatGPT.
- 2. A nivel de asociaciones humanas en torno a la teoría de actor-red, se refiere a las transgresiones que afectan a las relaciones o a las interacciones entre los humanos y la IA, o entre los humanos a través de la IA. Estas transgresiones pueden ser cometidas por la IA, por las entidades que la desarrollan o la usan, o por las personas que interactúan con ella. Por ejemplo, se puede mencionar el caso de los sistemas de IA que pueden manipular, engañar, influir o persuadir a las personas, o el caso de los sistemas de IA que pueden generar conflictos, desconfianza, desigualdad o exclusión entre las personas.
 - a. Transgresión a la explicabilidad de los sistemas de IA: De la
 - i. Dentro de estas transgresiones también resaltan algunas que se pueden considerar como de *algorithm cruelty* o crueldad de los algoritmos:
 - 1. El uso de IA para el control social, la vigilancia masiva, la censura, la represión o la violación de los derechos humanos.
 - 2. El uso indiscriminado de la IA para fomentar sistemas de explotación de trabajo mediante algoritmos, como se ha documentado en la economía de plataformas o *gig economy* (Morton, 2020).

3. El uso de IA para la guerra, el terrorismo, el espionaje, el sabotaje o el ciberataque.
 4. El uso de IA para la explotación, el abuso, el acoso, el engaño o la extorsión.
 5. El uso de IA para la generación de contenido falso, malicioso, ofensivo o dañino, como las noticias falsas, los deepfakes, el discurso de odio o la pornografía infantil.
 - ii. El uso de IA para la creación de armas autónomas, letales o incontrolables, que puedan actuar sin supervisión humana o ética.
 - iii. El posible de IA para la experimentación, la tortura, el sacrificio o la eliminación de animales o de otras formas de vida.
3. A nivel técnico, se refiere a las transgresiones que afectan a la calidad, la seguridad, la eficacia o la ética de los sistemas de IA. Estas transgresiones pueden ser cometidas por la IA, por las entidades que la desarrollan o la usan, o por las personas que interactúan con ella. Por ejemplo, se puede mencionar el caso de los sistemas de IA que pueden tener errores, fallas, vulnerabilidades o ataques, o el caso de los sistemas de IA que pueden tener sesgos, opacidad, incomprensibilidad o incoherencia.
4. A nivel sociotécnico, se refiere a las transgresiones que afectan a la gobernabilidad, la legalidad, la regulación o la responsabilidad de los sistemas de IA. Estas transgresiones pueden ser cometidas por la IA, por las entidades que la desarrollan o la usan, o por las personas que interactúan con ella. Por ejemplo, se puede mencionar el caso de los sistemas de IA que pueden violar las leyes, las normas, los principios o los valores vigentes, o el caso de los sistemas de IA que pueden generar incertidumbre, ambigüedad, impunidad o irresponsabilidad.

Algunos casos de transgresión a nivel técnico y sociotécnico derivados de la implementación de sistemas de inteligencia artificial son el uso de LLMs para generar noticias falsas, desinformación y extorsión (Barman et al., 2024; Gobierno del Estado de Sonora, 2025), así como el empleo de algoritmos de reconocimiento facial en sistemas de vigilancia masiva (Bhumarkar et al., 2024) y la automatización de decisiones en plataformas de trabajo, lo que puede reforzar sesgos y afectar las oportunidades laborales de algunas minorías (Morton, 2023).

A esto se suman los problemas técnicos relacionados con el uso y tratamiento de los datos de los usuarios por medio de plataformas que emplean sistemas de IA; por ejemplo, la aplicación DeepSeek ha sido señalada por transmitir información sin protección avanzada y utilizar claves de cifrado obsoletas, lo que facilita la vulneración de datos sensibles y el robo de credenciales (García, 2025). Esto genera un ambiente de total incertidumbre, de falta de transparencia, una nube constante que amenaza con una tormenta con respecto al impacto de la IA en las asociaciones de los usuarios con estos sistemas, y al tratamiento de cosas fundamentales como son sus datos.

En suma, la transgresión es un concepto que permite analizar los diferentes tipos y niveles de violaciones o de daños que puede causar la IA a los humanos, a las organizaciones, a la sociedad o al medio ambiente. La transgresión también se relaciona con los principios elementales para generar confianza alrededor de los sistemas de IA, como la equidad, la rendición de cuentas y la transparencia. Y la transgresión también engloba los debates alrededor del impacto social de esta tecnología, con respecto a determinar quién es responsable, y la forma en qué se pueden evaluar los riesgos y hacer propuestas regulatorias, autorregulatorias, y de resistencia con respecto a la IA.

3. La teoría de actor-red: Latour

La teoría en este trabajo deriva en buena parte de las ideas de Bruno Latour. En particular, de los armamentos teóricos y metodológicos en la Teoría de actor-red; además, estas aportaciones del autor se usarán en conjunto con los planteamientos hechos por Boltanski y Thévenot en aras de explicar tanto el rol de

la IA en la sociedad, así como la respuesta de esta en torno a la transgresión de los principios éticos mencionados a lo largo de este trabajo. Por ello, es importante explicar las ideas principales alrededor de la teoría de actor-red (el punto de frontera de la metodología de este trabajo) desde la perspectiva de Latour, para después analizar cómo estas coinciden y conviven con las aportaciones de Boltanski particularmente en la capacidad de los actores humanos de mostrar resistencia.

La Teoría de actor-red propone que los actores no humanos, como los objetos, también influyen en el entorno social al formar parte de redes de asociaciones con actores humanos. Esto implica que los objetos no son meros instrumentos o recursos, sino que tienen agencia y capacidad de modificar las redes y los actores en estos. Por eso, esta teoría rompe con la visión antropocéntrica que solo considera a las personas como actores sociales y se enfoca en las asociaciones como redes de actores humanos y no humanos.

La IA es una tecnología que no solo crea herramientas y algoritmos pasivos, sino que también genera impactos sociales a través de las interacciones que establece con otras personas y objetos (tecnologías). Esto se evidencia en las plataformas sociales (TikTok, Twitter, Youtube, entre otras) donde los algoritmos de la IA diseñan sistemas que captan la atención; y según Trejo et. al (2020) "...el sistema se enfoca en la moderación de contenido. El método más común para este tipo de sistemas es el de maximizar el número de clics. Esto implica que el sistema privilegia el contenido que ha sido comentado o compartido muchas veces".

Ahora bien, con base en que se descarta el rol de los objetos, es importante responder qué es y qué no es la teoría de actor-red. En principio, la teoría de actor-red destaca el rol de todos los actores dentro de las asociaciones entre actores, y entre ellos están los objetos; se centra en el papel de las asociaciones que luego empujan acciones entre sí. No obstante, esto no significa que un objeto tenga el mismo papel ni la misma relevancia que un actor humano, sino que su papel forma funciona como elemento dentro de las asociaciones de las que

forman parte los humanos, pero que no se pueden comprender puntualmente sin los objetos mismos; en palabras de Latour (2005):

“La teoría de actor-red no es la afirmación vacía de que los objetos hacen cosas 'en lugar' de los actores humanos: simplemente dice que ninguna ciencia de lo social puede siquiera comenzar si la cuestión de quién y qué participa en la acción no se explora en primer lugar a fondo, a pesar de que podría significar dejar elementos en los que, a falta de un término mejor, llamaríamos no humanos” (en *Reassembling the Social: An Introduction to Actor-Network-Theory*).

Desde el punto de vista de Latour, la novedad en la teoría de actor-red (ANT por sus siglas inglés) no es solo el papel de los objetos, sino también “lo que explica el paisaje contrastado con el que comenzamos, los poderes generales de la sociedad, las enormes asimetrías, el ejercicio aplastante del poder” (Latour, 2005). De nueva cuenta, esto no quiere decir que los objetos tengan el mismo protagonismo en las asociaciones, sino que estos abren la posibilidad de que aquello llamado social (ídem) tenga en su ensamblaje nuevos actores. Y esto se puede ver de forma en la vida cotidiana; por ejemplo, un bolardo protege a los ciclistas en su carril confinado, una pantalla indica a las personas el estado de su vuelo, y una ley establece nuevas condiciones de convivencia, entre otros lo que muestra que estos objetos forman asociaciones con los seres humanos y generan un movimiento y acción en los seres humanos.

Así pues, la ANT genera herramientas teóricas y metodológicas para comprender por qué las condiciones sociales rara vez consisten únicamente de asociaciones humano a humano (ídem). Es decir, casi siempre las asociaciones contendrán actores no humanos como *mediadores*. Y esto implica una tarea compleja para la persona que decide usar las herramientas de la teoría de actor-red; como bien lo explica Latour (ídem), quien aplique esta teoría no debe de limitarse a los lazos sociales ni convertirse en un experto tecnológico (lo que la hace una herramienta inter y transdisciplinaria por excelencia).

Ahora bien, se pueden distinguir dos tipos de mediadores con respecto a la IA y usando los elementos de la teoría de actor-red. Mediadores a favor de la IA, que son aquellos que promueven y facilitan su integración y desarrollo en la sociedad como lo son innovadores tecnológicos, las empresas que desarrollan soluciones

de IA, y los defensores de la tecnología que ven en la IA una herramienta para el progreso y la resolución de problemas complejos. Y los mediadores en contra que plantean cuestiones críticas sobre la IA, como los éticos, los activistas de derechos humanos y los académicos, quienes cuestionan las implicaciones de la IA en la privacidad, la autonomía y la equidad social.

Limitar es la otra advertencia que hace Latour en cuanto a la teoría de actor-red. De acuerdo con el autor, es poco práctico y realista contemplar tanto a los pequeños agentes individuales como a los grandes actores. Por ello aconseja centrarse en un solo elemento, en un solo actor, y para fines más prácticos, en una sola pregunta de dónde se desprenda el elemento central de una investigación.

En suma, no hay un límite de cuántos actores hay ni cuáles se deben de escoger. Esto depende de la pregunta y la naturaleza de la investigación. No obstante, sí es importante tener en cuenta que las asociaciones sociales no tienen un perfil meramente antropocéntrico; y con respecto al número de actores, Latour explica lo siguiente:

“...para la ANT, no hay un tipo preferible de agregados sociales, existe un sin número de mediadores, y cuando estos se transforman en intermediarios fieles no es la regla, sino una rara excepción que tiene que ser explicada por algún trabajo extra, generalmente por la movilización de aún más mediadores” (Latour, 2005).

Es necesario explicar los componentes técnicos de la teoría del actor-red. Es decir, más allá de plantear qué es y no es la teoría, hay que retomar los conceptos fundamentales en torno a esta para poder aplicarlos a un trabajo y una investigación *per se*. Además, dado que este trabajo retoma elementos de Boltanski, y en especial de la sociología de capacidades, la comprensión de estos elementos ayudará a resaltar las diferencias y coincidencias entre ambos modelos explicativos.

El primer elemento que se debe aclarar es el de actante. De acuerdo con Latour, actante es un término retomado de la literatura donde se ilustra la capacidad de un actor para imprimir movimiento sobre el otro. Y Latour explica cómo decidió emplear la palabra actante de la siguiente forma:

“La clave consiste en definir al actor por lo que hace -por sus actuaciones*- en las pruebas de laboratorio*. Más adelante, su competencia se deduce y se añade como parte de una institución*. Dado que, en inglés, el uso de la palabra «actor» se suele restringir a los humanos, la palabra «actante», tomada en préstamo de la semiótica, se utiliza a veces como el elemento lingüístico que permite incluir a los no humanos en la definición” (Latour, 2001).

El siguiente elemento que se debe de definir es el cuasi objeto. Este término básicamente se refiere a un actor en movimiento, a un actor que está haciendo algo; por ejemplo, cuando tu teléfono inteligente te manda una notificación, o cuando cambia la señal de tráfico. Pero en términos de Latour: “Como regla, un cuasiobjeto debe pensarse como un actor en movimiento, porque transforman el objeto en movimiento” (Latour, 2017).

De esta manera, Latour llama a cuasi objetos a los híbridos que surgen entre las asociaciones de objetos y personas, pero más concretamente entre el mundo natural y el mundo social. De acuerdo con el autor, los cuasi objetos explican aquellos fenómenos que se dan más allá de la división moderna entre lo natural y lo social. Es decir, un cuasi objeto surge para complementar el vacío explicativo de la división entre las ciencias sociales y las naturales, así como la contradicción de la segunda de dar demasiado o poco peso a los actores no humanos.

De esta manera, como afirma Latour, los objetos no son simples receptáculos de categorías sociales duras o suaves. Esto quiere decir que los objetos tienen un rol importante en las asociaciones sociales y este rol no está ligado a las implicaciones de su trascendencia. Es decir, un objeto no queda relegado a los aspectos explicativos de las ciencias sociales; además, tampoco se convierte en la base de la realidad social, por ejemplo, o desde la perspectiva del autor:

“De los cuasiobjetos, cuasisujetos, diremos simplemente que trazan redes. Son reales, bastante reales, y los humanos no los hemos creado. Pero son colectivos porque nos unen unos a otros, porque circulan en nuestras manos y definen nuestro vínculo social por su propia circulación” (Latour, 1993, p. 89).

La teoría de actor-red se propone como una alternativa a lo que Latour denomina figuración. Este concepto se refiere a “...la tendencia a explicar lo social por medio de formas antropomórficas... Es necesario comprender que hay muchas otras

formas posibles que no son antropomórficas... Solo se trata de hacerlas ideográficas en lugar de antropomórficas” (Latour, 2005, pp. 53). Es decir, la teoría de actor-red incorpora elementos, actores, que enriquecen una visión del mundo que asume que lo social era social por sí mismo.

Otro elemento importante en la teoría de actor-red son los mediadores de los objetos o de los actores no humanos. De acuerdo con Latour (1993), se convierten en mediadores, es decir, actores dotados de la capacidad de traducir lo que transportan, de redefinirlo, reubicarlo y también de traicionarlo; en otras palabras, redefinen lo que se dice con otras palabras, manteniendo el mismo significado y tono. En este sentido, los mediadores se convierten en una especie de voceros entre los actores humanos y no humanos, entre este grupo de actantes.

En el caso de la IA, los mediadores a favor de estos sistemas son aquellos actores que promueven, difunden, implementan y defienden el uso de la IA para diversos fines, como la innovación, la eficiencia, la competitividad, la seguridad, el entretenimiento, etc. En este sentido, los mediadores pueden ser empresas, gobiernos, organizaciones, investigadores, medios de comunicación, influencers, entre otros que hablan en nombre de la IA y le atribuyen capacidades, beneficios y valores. Sin embargo, estos mediadores también pueden ocultar, ignorar, minimizar o justificar los riesgos, los problemas, los errores o los daños que la IA puede causar a las personas, a la sociedad o al medio ambiente. Por eso, es necesario someter a estos mediadores a un análisis crítico y a una evaluación de su veracidad y su responsabilidad.

De acuerdo con Latour (2008), los actores de una red pueden ser tanto humanos como no humanos, como objetos, artefactos, tecnologías, etc. Estos actores tienen la capacidad de influir en el comportamiento de otros actores mediante lo que él denomina programas y antiprogramas. Un programa es una acción o una intención que un actor quiere realizar o imponer a otros (Latour, 1992). Y un antiprograma es una resistencia o una oposición a ese programa por parte de otro actor (Latour, 1992); por ejemplo, un semáforo tiene el programa de regular el tráfico mediante sus luces roja, amarilla y verde: un conductor puede tener el

antiprograma de pasar el semáforo en rojo si tiene prisa o no hay nadie más en la calle. Así pues, los programas y los antiprogramas son parte de la dinámica de la red, que se construye y se transforma a través de las interacciones entre los actores (Latour, 2005).

Los programas, anti-programas y anti-anti-programas explican las relaciones de poder y la estabilidad de la traducción dentro de una red de asociaciones. La traducción, según Star y Griesemer (1989), es el proceso mediante el cual actores con intereses diversos logran coordinarse a través de la creación de objetos frontera, los cuales permiten interpretaciones flexibles sin perder coherencia dentro de la red. En el contexto de la IA, este proceso de traducción se refleja en cómo desarrolladores, reguladores y usuarios negocian significados y prácticas en torno a los sistemas algorítmicos, definiendo así el impacto de estas tecnologías en los derechos humanos.

Latour, en *Technology is Society Made Durable* (1990), plantea que un anti-programa es la resistencia de un actante a un programa. Su ejemplo del llavero ilustra cómo la necesidad de que los huéspedes devuelvan la llave (programa) se enfrenta a la resistencia de quienes la olvidan o se la llevan como recuerdo (anti-programa). Para contrarrestar esto, se introduce un anti-anti-programa: un llavero enorme que dificulta llevárselo, asegurando el retorno de la llave.

Cuando un anti-programa genera una tensión en la red de asociaciones y compromete las intenciones originales del programa, ocurre lo que Callon (1984) denomina una traición a la traducción. Supongamos que una empresa extranjera instala cámaras con reconocimiento facial en una universidad como la UAM-Cuajimalpa o Northeastern University en Boston, prometiendo mayor seguridad. Sin embargo, si ocurre una filtración de datos o el sistema identifica erróneamente a personas racializadas como sospechosas, se genera una tensión en la red. La resistencia estudiantil contra estas cámaras emerge como un anti-anti-programa, reconfigurando la red de asociaciones en respuesta a los riesgos percibidos.

Este mismo proceso puede observarse en plataformas digitales de trabajo como Uber. Inicialmente, los sistemas de IA se presentan como herramientas para optimizar el trabajo de los conductores, pero terminan funcionando como supervisores algorítmicos que imponen restricciones, sanciones y control sobre ellos, convirtiéndose en un anti-programa. En respuesta, los trabajadores implementan estrategias para manipular los sistemas a su favor (Morton, 2023) o desarrollan anti-anti-programas al incorporar nuevos actantes que estabilizan la red, como cooperativas digitales o herramientas alternativas de optimización (Toxtli & Savage, 2023).

De esta manera, la teoría del actor-red, complementada con conceptos y metodologías satélite como los objetos frontera y el análisis de constelaciones, emerge como un modelo explicativo, organizacional y metodológico que permite mapear las asociaciones entre distintos actores en el desarrollo y aplicación de la inteligencia artificial. A su vez, facilita la identificación de puntos de tensión y traición en estos sistemas, revelando en qué momentos surgen anti-programas que vulneran a los usuarios. Con ello, se pueden rastrear, analizar y proponer soluciones a problemas específicos como la transgresión a la privacidad, la discriminación algorítmica y la opacidad en el acceso a la información.

4. El modelo de equivalencia: Boltanski y Thévenot

Boltanski ofrece una herramienta clave para analizar la reacción de los actores humanos frente a los desafíos éticos y sociales de la IA. A partir de los conceptos de programa y antiprograma de Latour, y de traducción de Callon, se propone un mapa que muestra cómo la IA impone un programa basado en códigos que vulneran los derechos de acceso a la información, no discriminación y privacidad. Asimismo, se recurre a Boltanski para entender la resistencia social y civil que se articula en torno a actores no humanos vinculados con instancias legislativas.

Hay que resaltar que a lo largo de este trabajo de investigación se observó que se necesitaba una herramienta sociológica que complementará a la teoría de actor-red. Y esta herramienta es el modelo de equivalencias propuesto por Boltanski y Thévenot (1999). Esto es importante, porque, según los autores,

existen muchos modelos de justificación que a veces son incompatibles, estos modelos requieren de herramientas para solucionar un conflicto y muchas veces esto se dan a partir de equivalencias ligadas a otras personas u objetos.

En este trabajo de investigación se observó que la teoría de actor-red (TAR) requería de una herramienta sociológica complementaria. Esta herramienta es el modelo de equivalencias de Boltanski y Thévenot (1999), que propone diferentes formas de justificar las acciones y los valores de los actores en situaciones de conflicto o controversia. Estas justificaciones pueden ser incompatibles entre sí y necesitar de ciertos recursos para establecer equivalencias con otros actores o entidades, tanto humanos como no humanos.

Boltanski y Thévenot (1999) explican que existen diversos modelos de justificación que los actores humanos utilizan para resolver los conflictos que surgen en torno a la IA. Estos modelos se basan en diferentes principios y valores que pueden entrar en contradicción. Para lograr acuerdos o consensos, los actores recurren a equivalencias que establecen con otras personas u objetos. Estas equivalencias son formas de resistencia ante el programa impuesto por la IA, que se puede analizar desde la perspectiva de Latour.

El modelo de equivalencias también aporta de manera significativa a los análisis comparativos y a la diversidad de estudios planteados en este trabajo. Esto se debe, en primer lugar, a que se articula con las nociones teóricas y metodológicas expuestas a lo largo de este trabajo. En segundo lugar, permite hacer un análisis más profundo de las situaciones locales y las respuestas ante los impactos de la IA desde estas.

El modelo de equivalencias también fortalece y complementa las herramientas de análisis de esta tesis debido a lo heterogéneo de las plataformas y de los mundos o escenarios donde se sitúan los principios éticos que la IA transgrede. Además de las complejas diferencias entre las sociedades, lo que hace indispensable la aplicación de un instrumento común para poder tener un resultado interpretativo y holístico del impacto de la IA en la sociedad con respecto a los derechos a la privacidad, acceso a la información, y a la no discriminación.

Así pues, considerando que los sistemas de la IA generan un programa, y este programa está compuesto por códigos que a su vez transgreden los derechos humanos, es necesario analizar la respuesta de la sociedad de las personas mediante regulaciones. Esto es, porque mediante una regulación se responde a la transgresión de los derechos humanos. Y más importante aún, se requiere examinar cómo se da este paso, y ahí es donde entra en juego el esquema inspirado en las observaciones Boltanski y Thévenot (1999).

5. La sociología de la capacidad crítica y modelo de equivalencias

El artículo La sociología de la capacidad crítica de Boltanski y Thévenot (1999) examina cómo las situaciones sociales requieren, en muchos casos, la justificación de una acción o la necesidad de una acción justificada. Para ello, los autores proponen modelos de fijación de valor que explican cómo los individuos pueden formular críticas dentro de un marco de justificación. Es importante señalar que, según su planteamiento, no todas las situaciones sociales pueden analizarse desde una perspectiva de justicia basada en emociones como el amor o en relaciones familiares.

El concepto de capacidad crítica hace referencia a la facultad de los individuos para evaluar y cuestionar situaciones percibidas como injustas dentro de un determinado contexto social. En este sentido, cuando una persona expresa su descontento o inconformidad, se genera lo que los autores denominan una escena. Esta puede manifestarse de distintas maneras, desde la confrontación abierta hasta estrategias más sutiles de evasión o negociación. Lo relevante es que, en todos los casos, los actores sociales ejercen resistencia ante aquello que consideran injusto.

Sin embargo, no todas las interacciones desembocan en una crisis. En situaciones donde el conflicto emerge, la negociación se convierte en una herramienta clave para alcanzar acuerdos. Para ello, es necesario que los involucrados compartan ciertos parámetros comunes que permitan construir un marco de entendimiento. De lo contrario, la falta de un terreno común puede llevar a respuestas impulsivas

y desproporcionadas, como reaccionar con violencia ante una situación de descontento.

Boltanski y Thévenot (1991) explican que los desacuerdos surgen cuando los actores operan desde distintos mundos cívicos (como el inspirado, el doméstico, el de la opinión, el cívico, el industrial y el mercantil) o cuando compiten por imponer su propia lógica como dominante. En estos casos, es necesario negociar o encontrar equivalencias entre los diferentes marcos de referencia para alcanzar un compromiso. La equivalencia, en este contexto, puede adoptar distintas formas: el reconocimiento mutuo de los mundos en disputa, el uso de recursos compartidos o incluso la construcción de un nuevo marco híbrido que integre elementos de ambas partes.

En cuanto a la manifestación de la escena, los autores señalan que esta puede adquirir diversas formas, dependiendo del contexto en el que se produzca (Boltanski y Thévenot, 1999). Por ejemplo, en un entorno doméstico, los conflictos pueden resolverse mediante mecanismos informales, mientras que en escenarios institucionales la mediación suele estar regulada por normas jurídicas y procedimientos formales. En ambos casos, la persona inconforme debe justificar su posición frente al otro actor, ya sea un individuo, un objeto o una institución.

Finalmente, las justificaciones empleadas deben seguir un criterio de aceptabilidad (Boltanski y Thévenot, 1999). Es decir, no basta con expresar una inconformidad basada en preferencias personales o prejuicios sin fundamentos sólidos. Por ejemplo, no sería válido argumentar una queja simplemente porque no se aprueba la vestimenta o los gustos musicales de alguien. En este sentido, el análisis de Boltanski y Thévenot subraya la importancia de establecer marcos de referencia comunes para la resolución de disputas, destacando el papel central del principio de equivalencia, el cual permite fundamentar los argumentos en elementos con legitimidad reconocida dentro del orden social.

6. El principio de equivalencia

El principio de equivalencia es un elemento innovador en el planteamiento de Boltanski. Según el autor, este principio no debe confundirse con la idea de un intercambio de valores iguales, como en la alquimia o el mercado. En realidad, el establecimiento de una equivalencia es un momento clave cuando una persona entra en una disputa. Para justificar, criticar o explicar una situación, es necesario conectar diferentes personas, objetos e ideas y establecer referencias comunes que permitan la comparación y el juicio.

Este proceso de equivalencia es esencial para la resolución de desacuerdos, ya que permite a los individuos salir de sus posiciones particulares y generar un espacio de discusión basado en elementos compartidos. Un ejemplo de este principio es el uso de leyes o normas que funcionan como referencia común en una disputa. Así, una ley civil puede servir como un objeto de equivalencia en un divorcio, del mismo modo que las reglas de tránsito estandarizan la interacción entre automóviles y carreteras.

Boltanski y Thévenot explican que, para evitar la violencia, las personas deben reducir los motivos de descontento privado y construir una definición compartida de los objetos relevantes en una disputa. Sin embargo, no todas las equivalencias son aceptadas de manera uniforme. Algunas personas pueden rechazarlas o descartarlas, lo que complica la posibilidad de alcanzar acuerdos.

En este contexto, los autores identifican diferentes regímenes de acción. En el régimen emocional, las personas dejan de lado las equivalencias y se guían por sentimientos subjetivos. En el régimen de acción, en cambio, se intenta construir equivalencias para justificar posiciones y resolver conflictos.

Un caso particular es el régimen de ágape, en el que los individuos renuncian a las equivalencias y justificaciones propias de otros regímenes. Aquí, las acciones se motivan por el amor, la compasión o la generosidad, sin buscar comparaciones o reciprocidad. Este régimen se asocia con experiencias religiosas, artísticas o

revolucionarias y rompe con las estructuras de equivalencia establecidas en la sociedad.

El principio de equivalencia, por tanto, permite entender cómo los actores justifican sus reclamos y resuelven disputas mediante la creación de referencias comunes. Boltanski y Thévenot subrayan que su objetivo no es minimizar el papel del poder o la dominación, sino comprender cómo las personas experimentan y solucionan injusticias a partir de su vida cotidiana. Lo cual también puede dar pautas para comprender y solucionar los problemas que ocasionan los sistemas de la IA.

¿Cómo llegaron los autores a plantear un modelo donde la solución de una situación de informalidad se da a partir de equivalencias encarnadas muchas veces en especialistas, objetos e instituciones con un poder superior a lo simbólico capaz de actuar sobre los actores humanos? La respuesta a ello se encuentra en tres niveles:

1. Los datos empíricos que sustentan los argumentos.
2. Los textos clásicos de filosofía moral con base en autores de autoridad que resaltan el bien común como herramienta de equivalencia. Sin embargo, estos elementos del bien común son heterogéneos, y una persona debe saber transitar de un principio de equivalencia a otro según la situación.
3. El rol de los objetos como guías en las que las personas se refugian y buscan resolver un conflicto con base en diferentes situaciones. A partir de estos puntos, los autores plantean una serie de mundos. Estos mundos son escenarios con elementos para la resolución de conflictos mediante datos empíricos anclados a textos filosóficos específicos para cada situación. Y los objetos que encarnan estas equivalencias, como una obra particular, una constitución, una IA que detecte noticias falsas o incluso instituciones, sirven como mecanismo de autoridad al momento de reclamar una injusticia.

Con base en los puntos anteriores, los autores plantean una serie de mundos. Estos mundos son escenarios con elementos para la resolución de conflictos mediante datos empíricos anclados a textos filosóficos específicos para cada

situación. Y los objetos que encarnan estas equivalencias, como una obra particular, una constitución, una IA que busque noticias falsas, e incluso instituciones sirven como mecanismo de autoridad al momento de reclamar una injusticia.

Asimismo, los autores mencionan que los procesos para expresión de desacuerdo de los mundos se componen de cuatro elementos. El primero es el mundo de evaluación o el valor (gracia, reputación, interés colectivo, renombre y precio), el segundo es el formato de información relevante (emocional, oral, formal, semiótico y monetario), el tercero es la relación elemental (pasión, confianza, solidaridad, reconocimiento e intercambio), y el cuarto es la calificación humana (creatividad, autoridad, igualdad, celebridad, deseo de compra). Estos cuatro elementos forman parte de seis mundos ideales (porque no se representan exactamente de esta manera en la realidad) y los datos empíricos o textos filosóficos que los respaldan; y estos mundos son:

1. **El mundo de la imaginación:** La fuente es San Agustín que decía que había un elemento externo de donde todo el valor provenía. Su información relevante es la gracia porque se basa en obtener valor sin importar lo que piensen otros, el formato de información relevante es el emocional, su relación elemental es la pasión, y su sistema de calificación humana es la creatividad. Y es el mundo de los niños y de los locos.
2. **El mundo doméstico:** depende de la jerarquía de una serie de interdependencias personales. Su información relevante es la reputación, el formato de información relevante es el oral, su relación elemental es la confianza, y su sistema de calificación humana es la autoridad. Y para hacer equivalencias están las guías de buenos modales.
3. **El mundo cívico:** El contrato social de Rousseau es el texto empírico que proponen los autores. Y sus equivalencias son objetos ligados a códigos inmateriales y leyes como por ejemplo los derechos humanos y las regulaciones a las que están anclados. Su información relevante es el interés colectivo, el formato de información relevante es el formal u oficial, su relación elemental es la solidaridad, y su sistema de calificación humana es

la igualdad; y justamente este el mundo donde se pueden resolver muchas de las injusticias en torno a la IA y la transgresión a los derechos humanos, dado que está lleno de equivalencias y argumentos que se pueden convertir en equivalencias, para lograrlo.

4. **El mundo del renombre u opinión:** Basado en el Levitan de Hobbes y su peso al honor. Su objeto son las guías de principios y técnicas en relaciones públicas. Su información relevante es la reputación, el formato de información relevante es el semiótico, su relación elemental es reconocimiento, y su sistema de calificación humana es la celebridad.
5. **El mundo del mercado:** La riqueza de las naciones de Adam Smith es el texto empírico que proponen los autores. Los objetos son libros pedagógicos sobre el mercado. Su información relevante es el precio, el formato de información relevante es el monetario, su relación elemental es el intercambio, y su sistema de calificación humana es el deseo o poder adquisitivo.
6. **El mundo industrial:** Los textos empíricos que proponen los autores están en los escritos de San Simón. Los objetos de equivalencia se pueden encontrar en manuales de procedimiento. Su información relevante es productividad, el formato de información lo mensurable, su relación elemental es la funcionalidad, y su sistema de calificación humana son las capacidades profesionales.

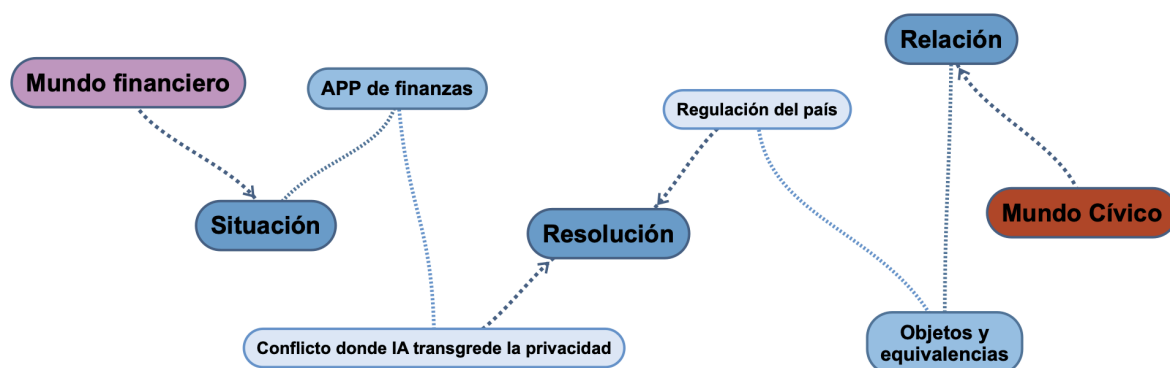
A continuación, los autores explican que un conflicto puede darse en torno a dos mundos. Esto de nueva cuenta se relaciona con el régimen de ágape, pero más que nada se centra en las equivalencias que dos mundos distintos traen al momento en que un actor manifiesta una injusticia. En el caso de la IA, es muy común que el conflicto traiga equivalencias dos mundos; por ejemplo, el mundo financiero que norma la IA busca extraer la mayor cantidad de datos para maximizar su ganancia a pesar de que esto afecte la privacidad de los usuarios; mientras que el actor que se ve afectado levanta una queja con base en las equivalencias del mundo cívico ligadas a la resolución de un país.

De nueva cuenta esto se puede explicar mediante la noción de programa y antiprograma. Por un lado, los objetos, en este caso la IA de una compañía, generan programas y estos programas están compuestos por códigos que se mueven bajo la equivalencia de aumentar la ganancia sustentada justamente en los mismos códigos de la IA. Por otro lado, los actores humanos crean un objeto que contiene un programa compuesto por códigos sustentados en las equivalencias civiles para contrarrestar el programa de la IA; así pues, el programa es el actor que comete una falta, mientras que el antiprograma es el objeto de la persona que percibe una injusticia y esto se puede ver en el esquema de abajo (imagen 5, de autoría propia).

Los compromisos son formas de coordinar la acción entre actores que pertenecen a diferentes mundos o ciudades, es decir, que se rigen por diferentes principios de justificación. Y estos implican una negociación entre las partes para encontrar un punto de acuerdo o una equivalencia común que permita resolver un conflicto o una controversia. Asimismo, los compromisos pueden ser más o menos estables y más o menos explícitos, dependiendo del grado de compatibilidad entre los mundos involucrados y de la capacidad de los actores para argumentar y reconocer las justificaciones de los otros. Los compromisos también pueden estar mediados por objetos que funcionan como referentes comunes o como fuentes de autoridad para los actores.

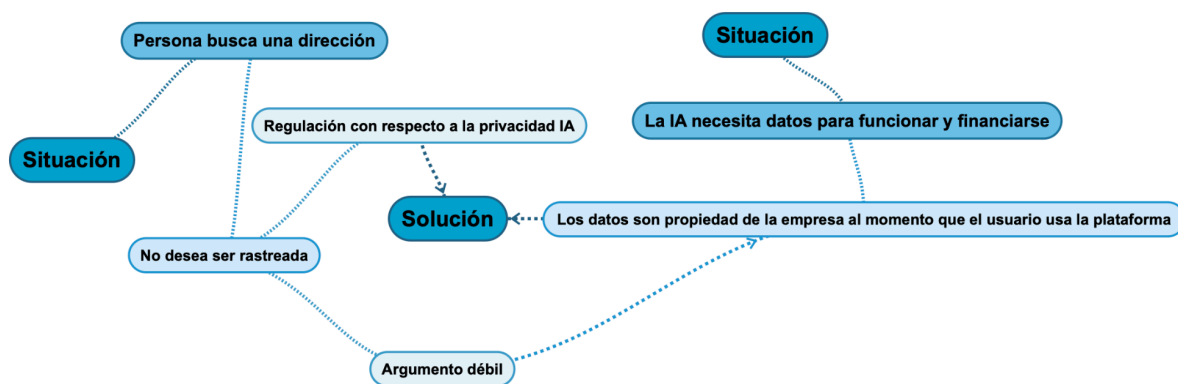
Estos objetos pueden ser desde obras de arte o documentos legales hasta sistemas de IA o instituciones públicas. Un ejemplo de compromiso entre mundos en el ámbito de la IA y los derechos humanos es el caso de la Declaración de Toronto sobre la protección del derecho a la igualdad y la no discriminación en los sistemas de aprendizaje automático. Esta Declaración es un documento que busca establecer un marco ético y normativo para el desarrollo y uso de la IA, basado en los principios de derechos humanos. La Declaración es el resultado de un proceso de diálogo y consulta entre diversos actores, como organizaciones de la sociedad civil, empresas tecnológicas, académicos y expertos, que pertenecen a diferentes mundos o ciudades: el mundo cívico, el mundo industrial, el mundo inspirado, entre otros; y busca crear un compromiso entre estos mundos,

reconociendo sus intereses y valores, pero también estableciendo límites y responsabilidades para prevenir y remediar los posibles impactos negativos de la IA en los derechos humanos (Amnistía internacional, 2018).



(imagen 5, de autoría propia)

Los autores, más allá de los mundos en torno a la resolución del conflicto, también mencionan la noción de *compromisos*. De acuerdo Boltanski y Thévenot (1999), los compromisos se dan cuando dos o más mundos conviven y las equivalencias para resolver estos se tienen que negociar con base en la situación particular. Aquí puede haber objetos que convergen entre dos mundos, o situaciones donde las equivalencias están en conflicto; no obstante, un argumento débil se puede fortalecer bajo la construcción de un objeto o institución como se ve en la siguiente figura (imagen 6, de autoría propia).



(Imagen 6, de autoría propia)

El esquema muestra una situación en la que una persona usa una aplicación de movilidad para buscar una dirección, pero esta aplicación la rastrea violando su privacidad. La empresa puede justificar que necesita rastrear a los usuarios para que su IA funcione, y que esos datos también pueden ser usados para financiarse al venderlos a un data broker, además de que argumenta que los datos del usuario ya son de su propiedad. El usuario puede expresar su malestar por esa situación, pero sin apoyarse en ninguna equivalencia común que dé fuerza a su argumento, hasta que su inconformidad y su argumento se conviertan en un objeto encarnado en una regulación para proteger su privacidad de la IA.

Ahora bien, no todo conflicto requiere de equivalencias. Dado que, a veces una disputa se termina sin todos esos pasos puesto que simplemente las personas desisten de esta. Pero hay que examinar qué hay detrás de esas resoluciones o no limitarlo a todo a relaciones de poder o alineación ya que en ocasiones puede desistir de denunciar un conflicto porque no existen las equivalencias para argumentar su cosa, porque consideran que el costo beneficio de una denuncia no les favorece, o porque el único régimen existente es el emocional ligado a la violencia.

La inteligencia artificial opera dentro de múltiples “mundos” en el marco del modelo de equivalencias de Boltanski. En el mundo del mercado, su desarrollo e implementación están impulsados por la búsqueda de ganancias y ventaja competitiva. En el mundo industrial, se justifica como una herramienta para optimizar la productividad bajo el pretexto de ampliar las capacidades humanas. En ciertas ocasiones, la IA también habita el mundo de la imaginación, donde algunos de sus avances más significativos han surgido de la pura curiosidad intelectual, como en el caso de Geoffrey Hinton y su exploración sobre el funcionamiento de la mente humana.

La justificación para desarrollar e implementar IA varía según el mundo en el que se inscriba. En el mundo del mercado, el principio rector es avanzar rápidamente y romper paradigmas, sin considerar necesariamente las consecuencias, con el fin de consolidar una posición dominante. En el mundo industrial, la IA se introduce

para automatizar tareas y liberar a los humanos de trabajos tediosos o poco placenteros. En el mundo de la imaginación, la motivación es la exploración misma: diseñar un modelo de IA que pueda preparar el mejor sushi de Japón simplemente porque es un reto fascinante.

Las equivalencias emergen cuando surgen tensiones entre estos mundos o entre las constelaciones que los conforman. Si el mundo del mercado genera monopolios o vulnera la privacidad mediante el uso de sistemas de IA, el mundo cívico buscará una equivalencia jurídica, como leyes o impuestos, para contrarrestar estos efectos. Si el mundo industrial desarrolla modelos que desplazan el trabajo creativo, el mundo cívico responderá con regulaciones para mitigar el impacto. Y si el mundo de la imaginación produce modelos que reproducen patrones de discriminación, eventualmente surgirá un mecanismo para corregir sus efectos no deseados.

Con base en esto, el modelo de Bolstanski no permite rastrear cómo se usan las equivalencias para generar resistencias, algo similar al anti-anti-programa en la teoría de actor-red. Pero más allá de eso, y de lado de la teoría de actor-red, permite conocer como diferentes actantes se puede agrupar en estos mundos para cumplir un objetivo en torno a las intenciones y desarrollo de la inteligencia artificial. Por ejemplo, en el mundo del mercado, empresarios se asocian con sistemas de IA para maximizar ganancias y expandir su influencia, mientras que en el mundo industrial, ingenieros usan sistemas de inteligencia artificial, la infraestructura y sus actantes para que esta funcione, y los dispositivos para desplegarla para optimizar procesos; y en el mundo cívico, activistas y legisladores se apoyan en otros actantes como marcos regulatorios a modo de equivalencias para exigir transparencia, lo que muestra cómo los actantes—tanto humanos como no humanos—se entrelazan en redes sociotécnicas que configuran el desarrollo y la resistencia frente a la inteligencia artificial.

En suma, el modelo de equivalencias refuerza los conceptos de objeto de contorno, traducción y el esquema objeto-programa-códigos/objeto-antiprograma-códigos expuestos a lo largo de

esta tesis. Esto resulta crucial para entender cómo los países responden a la implementación de plataformas y sistemas de IA que transgreden principios fundamentales como la privacidad, la transparencia y la imparcialidad, así como para analizar las respuestas sociales ante las transgresiones de estas plataformas en el ámbito de la movilidad.

7. La discreta armonía entre Boltanski y Latour

Tanto Boltanski como Bruno Latour ofrecen herramientas clave para analizar las dinámicas sociotécnicas del impacto social de la inteligencia artificial (IA). Aunque sus enfoques difieren, Guggenheim y Potthast (2012) los describen como “gemelos simétricos”, señalando tres ejes de convergencia: el cambio social, el papel de los objetos y la crítica. Estas coincidencias permiten articular una perspectiva combinada para examinar cómo la IA reconfigura las relaciones entre humanos, tecnologías y marcos normativos.

En cuanto al cambio social, Latour lo concibe como un proceso continuo dentro de redes dinámicas de actores humanos y no humanos, mientras que Boltanski lo vincula con la emergencia de nuevas formas de justificación dentro de órdenes normativos. Esta diferencia influye en el análisis de la IA: la teoría del actor-red (ANT) permite rastrear las trayectorias y ensamblajes de los sistemas inteligentes, mientras que la sociología de la capacidad crítica ayuda a comprender los regímenes de justificación que legitiman la IA y las disputas que genera.

El segundo eje de convergencia, el papel de los objetos marca otra diferencia clave. Para Latour, los objetos tienen agencia dentro de redes de controversia; para Boltanski, su relevancia radica en cómo intervienen en los regímenes de justificación, especialmente cuando su materialidad se convierte en un factor de poder. Aplicado a la IA, esto implica que no basta con examinar cómo los algoritmos operan en redes sociotécnicas, sino que también es crucial analizar cómo se utilizan para resolver conflictos o perpetuar desigualdades. Además de los algoritmos, esto involucra instrumentos como leyes, instituciones y autoridades, que actúan como equivalentes funcionales en la estructuración del orden social.

Guggenheim y Potthast (2012) sostienen que la convergencia entre ambos autores parte de su distanciamiento de Pierre Bourdieu. Ambos critican la centralidad del concepto de campo y la marginación de los actores en la teoría de Bourdieu, a la que consideran determinista. Para Latour, el problema radica en la idea de que las razones de la acción no son accesibles a los actores; para Boltanski, en la rigidez de los campos sociales como estructuras predefinidas. Mientras Latour adopta una perspectiva microsociológica, observando laboratorios y redes de controversia, Boltanski se centra en la capacidad crítica de los actores y en la pluralidad de los órdenes de justificación.

Esta diferencia se refleja en la estructura de sus obras principales: *Nunca hemos sido modernos* (Latour, 1993) y *Sobre la justificación* (Boltanski y Thévenot, 2006). Latour reconstruye tanto el mundo natural como el social a través de redes de actantes humanos y no humanos, mientras que Boltanski analiza las controversias mediante seis órdenes de valor: inspiración, doméstico, fama, cívico, de mercado e industrial. Mientras Latour enfatiza la disolución de la frontera entre naturaleza y sociedad, Boltanski se centra en los principios normativos que los actores movilizan para justificar sus acciones.

Ambos autores comparten una visión crítica que se aleja del modelo de Bourdieu sin asumir una postura de superioridad intelectual. En lugar de ello, buscan una crítica empírica y detallada. Esta diferencia metodológica influye en sus aproximaciones al estudio de la inteligencia artificial, ya sea como una red de relaciones sociotécnicas o como un campo de disputas normativas.

Desde esta perspectiva, la IA puede analizarse en dos niveles complementarios. La teoría del actor-red permite rastrear la multiplicidad de actores involucrados en los sistemas inteligentes, desde desarrolladores hasta usuarios y reguladores. Por su parte, la sociología de la capacidad crítica ayuda a identificar los regímenes de justificación en los que se enmarcan las disputas sobre la IA, así como los objetos y mecanismos que pueden resolverlas. Esto permite observar los escenarios de conflicto a través de distintos “mundos” donde los programas, anti-programas y

anti-anti-programas entran en tensión al traicionar expectativas compartidas sobre la implementación de un sistema de IA en plataformas como Uber.

Un ejemplo de esto es el sesgo algorítmico en plataformas de reclutamiento, donde la IA ha aprendido a favorecer currículos con lenguaje típicamente masculino, discriminando a mujeres (Trejo et al., 2020). Desde la perspectiva de Boltanski, este problema puede situarse en un régimen cívico, donde la regulación actúa como un objeto de autoridad para corregir la injusticia. Sin embargo, la naturaleza cambiante de la IA exige una reevaluación constante de estos regímenes y sus mecanismos de resolución.

En conclusión, las teorías de Latour y Boltanski, aunque divergentes en su enfoque, se complementan al analizar la IA. Latour enfatiza la agencia de los objetos y la reconstrucción de controversias, mientras que Boltanski destaca la estructuración de los conflictos y los mecanismos de justificación. Esta combinación no solo permite rastrear el impacto de la IA en la sociedad, sino también comprender cómo los actores intentan resolver sus disputas. No obstante, ambas teorías enfrentan un desafío común: abordar la cuestión de la responsabilidad en los sistemas de IA, un problema central en la ética de estas tecnologías.

8. Velíz, los zombis morales, y la responsabilidad en la IA

Uno de los mayores problemas en torno a los problemas sociales de la IA es la responsabilidad. Sí, la IA causa problemas relacionados con la privacidad; sí, la IA aprende sesgos discriminatorios que impactan en la vida real de muchas personas; y sí, es complicado comprender por qué surgen de raíz estos problemas debido a que la IA es poco transparente bajo pretexto de propiedad intelectual o por problemas para saber simplemente cómo toma decisiones. En torno a estos tres problemas clave, es difícil determinar quién o qué es el responsable dada la capacidad de agencia que la IA y sus plataformas tienen, y dada la gran cantidad de actores que se elaboran, distribuyen y comercializan.

Por ello, es importante dedicar una sección al problema de la responsabilidad. Ahora bien, como se planteó en la parte dedicada a la teoría sociológica, se requiere de una herramienta conceptual que ayude a explicar este problema. Y de preferencia, esta herramienta debe de estar acompañada de un diagrama que complemente su explicación y que esté acorde con el método de análisis de constelaciones y la teoría del actor-red.

Así pues, se decidió elaborar un mapa y delimitar un concepto clave para explicar quién o qué es el responsable en las transgresiones de la IA. El concepto clave que se eligió fue el de zombi moral desde la perspectiva y análisis de la doctora Véliz Velarde. Y con base en el método de análisis, se creó una herramienta visual a forma de diagrama de flujo en torno al concepto clave y a la constelación de actores que forman parte de su órbita.

Véliz Velarde, a partir de la articulación del concepto de zombis morales aporta directamente a esta tesis porque desde un principio a ayudar a responder la pregunta de quiénes son los responsables de los problemas que generan los sistemas de IA ¿Los propios modelos, o los desarrolladores o participantes que los implementan? Tener esta noción ayuda a evitar puntos grises como el que no se puede juzgar a Uber por un resultado no esperado de sus sistemas de IA ya que estos tienen cierta agencia a similar a la humana; y esto, permite liberar en análisis para enfocarse en desarrollar planteamientos o anti-anti-programas que ayuden a solventar los problemas ocasionados por la IA.

El diagrama de flujo parte de los derechos humanos. Pero principalmente se centra en el problema de la responsabilidad en torno a la IA. Complementa los argumentos explicativos sobre los derechos de privacidad, acceso a la información y no discriminación; que servirá como elemento de análisis que se examinará con mayor profundidad a lo largo de la tesis.

Al detenerse en el problema de responsabilidad, el diagrama hace la pregunta de si el objeto encarnado en herramientas de la IA es un actor moral. Para responder esto, es necesario explicar el concepto de actor moral con base en la forma que lo introduce Véliz. En este sentido, Véliz (2021) sostiene que “la sensibilidad es

necesaria para la agencia moral, porque la experiencia consciente del tipo que permite sentimientos como el placer, el dolor y la empatía parece necesaria para la agencia moral”.

Así pues, la IA acotada (la que existe en todas las plataformas con IA) no tiene agencia moral. Dado que, a pesar de su capacidad de imprimir movimiento sobre otros actores a modo de actante, no tiene la capacidad reflexiva ni emocional para comprender el impacto de estas acciones. Es decir, no es más que una versión más compleja y con mayor capacidad de incidencia que un instrumento del laboratorio que examina Latour: relevante, más no moral ni reflexivo.

Para elaborar lo anterior, Véliz argumenta: “Los agentes morales toman decisiones y actuaron en consecuencia, al menos en parte, porque responden a razones. Si una entidad no es el tipo de criatura que puede entender las razones, entonces no se gobernará a sí misma de una manera que sea relevante para la moralidad”. Esto es de suma importancia, porque si la IA no es un agente moral, entonces esta no es responsable de sus acciones pese a que tenga capacidades de aprender y dar predicciones y resultados. La responsabilidad de los problemas de la IA será de los actores humanos que la desarrollan, implementan o comercializan: lo que es relevante cuando se presenta y se resuelve una inconformidad.

Esto se encuentra en la parte de abajo del diagrama de flujo. Ahí se explica de manera simple que la IA puede ejecutar tareas y tomar decisiones, además de que incide en el funcionamiento de las plataformas de las que forma parte. Aun así, estas capacidades no la hacen responsable de sus propias acciones al carecer los elementos necesarios para ser un agente moral, y de elementos de reflexión en la toma de decisiones que le permitan pasar de actante a agente: por ello, como lo plantea claramente Véliz con respecto a los algoritmos que forma la matriz de la IA:

1. No necesitamos pedirle permiso a un algoritmo para modificarlo o incluso terminarlo porque los algoritmos no tienen valores propios; no les importa su propia existencia.

2. No puede persuadir a un algoritmo para que haga algo dándole buenas razones; sólo puede programarlo de una forma u otra (idem).

Otro elemento importante en torno a la relevancia de por qué la IA no es un agente moral, es su capacidad reflexiva. Si bien la IA puede aprender sesgos discriminatorios por medio de datos como un niño lo hace por medio de su entorno o de sus padres, esta, a diferencia de un niño que se vuelve adulto, no tendrá la capacidad autorreflexiva de cambiar de opinión o de reflexionar acerca del daño que ha hecho. Por ello, es prácticamente imposible hacer responsable a la IA de sus acciones.

De esta manera, la responsabilidad de las acciones que toma la IA recae en los actores humanos que la desarrollan, elaboran y comercializan. Sin embargo, determinar la responsabilidad en este sentido resulta sumamente complejo porque cada las decisiones de cada persona pueden incidir de maneras distintas en el funcionamiento de la IA, ya sea intencionalmente o no. Por ejemplo: “La comunidad de visión computacional, encargada de hacer sistemas inteligentes con aplicaciones como la identificación de personas, el reconocimiento de estas y su monitoreo, ha estado plagada de errores” (Trejo et al, 2020).

En este sentido la persona responsable por los errores de los sistemas reconocimiento facial puede ser la persona que diseñó el algoritmo, la persona que lo entrenó (en caso de que sea aprendizaje de máquina supervisado), la persona que eligió la base de datos, o hasta la intención al momento de comercializar. Ahora bien, supongamos que alguien entrena a su modelo con una gran base de datos de internet con la intención de evitar sesgos por género o por etnicidad. Si esta base de datos es extraída de internet, es muy posible que el modelo contenga sesgos de género, raciales, de edad, entre otras, dado que los primeros usuarios de internet fueron varones blancos de países de primer mundo.

¿Quién es el responsable de que la IA sea responsable de esta falta? Como se aprecia, la responsable no puede ser la IA porque no es un actor moral, y probablemente tampoco lo sea quien eligió la base datos. Los responsables deben ser los directivos de las empresas, instituciones o corporativos que no generaron

los mecanismos de supervisión y de pruebas lo suficientemente robustos en torno a una herramienta que ya usa el público y que afecta a la sociedad.

En particular los directivos de estas redes de investigadores, desarrolladores e implementadores de la IA con fines comerciales o de gobierno tienen la tarea de prevenir que los productos u objetos con esta tecnología sean implementados en el campo sin que hayan pasado por procesos que prevengan sesgos o invasión de la privacidad. Además, estos actores, requieren promover que los productos con IA sean lo suficientemente transparentes como para ser auditados. Y más tarde, para hacer correcciones con respecto a las fallas de estos; y para que los actores que los usen estén conscientes de su impacto.

Saber qué es un actor moral sirve para identificar quién es el responsable en torno a los problemas sociales que ocasiona la IA. Esto es de suma importancia; en especial, por la capacidad que tiene esta tecnología de tomar decisiones y generar resultados. Sin embargo, simplemente se hace presente la necesidad de mecanismos de regulación y de resolución de controversias que muchas veces se harán valer de otros objetos que contengan autoridad o que sirvan para auditar las causas; y con ello, generar las reglas para sancionar los responsables, lo que se verá en el tercer capítulo.

La responsabilidad de los problemas sociales de la IA recae principalmente en los actores humanos que *participan en su creación y uso*. Como argumenta Carissa Véliz (2021), la IA no tiene conciencia ni voluntad propia, y por lo tanto no puede ser considerada como un agente moral; sin embargo, esto no significa que la IA sea inocua o neutral, ya que puede causar daños y beneficios a las personas y a la sociedad en función de los programas y antiprogramas que genera. Estos programas y antiprogramas son el resultado de las redes de asociaciones entre actores humanos y no humanos, que pueden estar basadas en intereses, valores, normas o sesgos; por lo tanto, es necesario que los actores humanos sean conscientes de las implicaciones éticas y políticas de la IA, y que ejerzan su capacidad de resistencia y cambio frente a los antiprogramas que atentan contra la dignidad, la justicia y el bienestar común.

CAPÍTULO SEGUNDO

LA TRAYECTORIA DE LA IA

El objetivo de este capítulo analizar la trayectoria de la IA con el fin de tener una herramienta de análisis que ayude a responder en qué medida las plataformas digitales que utilizan IA vulneran los derechos humanos relacionados con la privacidad, transparencia, e imparcialidad. Asimismo, se analizará cómo estos problemas afectan la responsabilidad y explicabilidad de las decisiones tomadas por la IA, y cómo se están abordando estos problemas en México y Japón. Para ello, se utilizarán casos empíricos relacionados con Lovelace, Adam Turing, Hinton, Rosenblatt, McCarthy, Hinton, así como las máquinas Alpha Go y Watson, que representan un avance significativo en el aprendizaje de máquina. También se estudiarán las plataformas de Didi, TikTok, Facebook, ya que estos actores han tenido un papel fundamental en el desarrollo de la IA.

La historia de estos actores y sus luchas ayudará a ilustrar cómo la IA pasó de ser una idea etérea constreñida por las limitaciones técnicas de distintas épocas, para convertirse en un campo concreto del conocimiento. Cabe señalar que se definirán ideas como la de traducciones y objetos de contorno que luego se emplearán en la tesis bajo este mapa con base en el método de análisis de constelaciones. Aun así, es necesario precisar más la definición de los objetos de contorno y las traducciones.

Con base a la metodología de análisis de constelaciones, este capítulo buscará examinar la trayectoria social de la IA en torno a su situación política, cultural, económica y tecnológica. Para ello se centrará en la formación de conceptos clave, además de una asignación de un objeto de frontera. Dentro de los conceptos clave sobresalen la IA, las redes neuronales, y el aprendizaje de máquina.

Hay que resaltar que la IA también será asignada como un objeto de contorno, pero es relevante explicar cómo surgió esta tecnología clave a partir de su trayectoria. Y dado que las redes neuronales, y el aprendizaje de máquina son

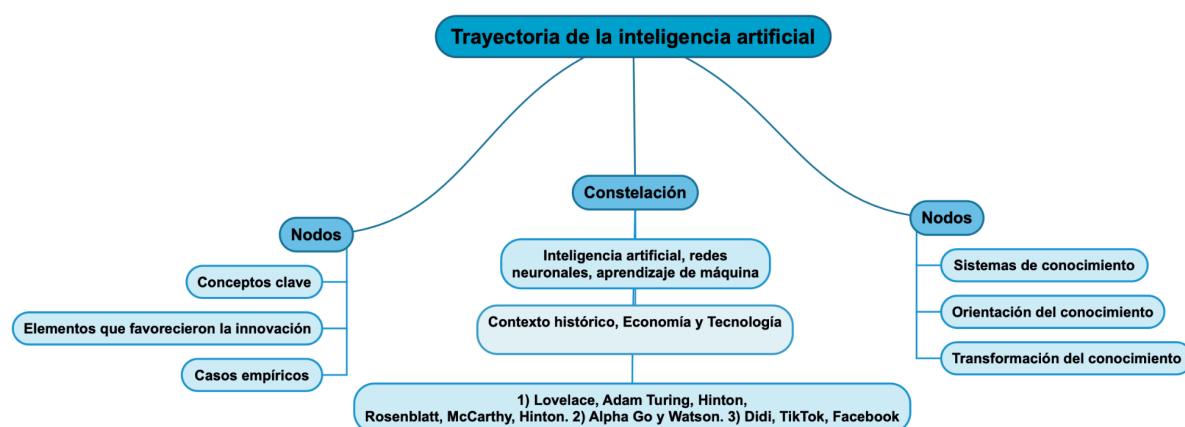
subcampos de la IA, serán examinados como parte de una constelación que formará una pieza de los sistemas de conocimiento. Mientras que su orientación de conocimiento se centrará en el análisis de una trayectoria simada en la innovación basada en distintas áreas del conocimiento como la economía, el estudio de la innovación tecnológica, las condiciones sociales, políticas, y la situación socio histórica de los investigadores que desarrollaron las herramientas en torno a la IA.

Los casos de estudio, los casos empíricos que forman parte de la constelación de la transformación del conocimiento, se centrarán en dos trayectorias: la de los actores humanos, y la de los no humanos. Dentro de la trayectoria de los actores humanos, se verá el caso de Ada Lovelace, Adam Turing, McCarthy, Rosenblatt y Geoffrey Hinton. Esto es, porque se consideró que, entre la gran variedad de investigadores y teóricos de la IA, este grupo de actores, esta constelación, es la que representa los retos, fracasos y triunfos de la IA con base en distintas situaciones que favorecieron o entorpecieron su avance.

Dentro de la constelación de los objetos, este capítulo decidió centrarse en la trayectoria de DeepBlue, AlphaGo. Estos objetos representan avances significativos en torno a la IA; y en especial, alrededor del aprendizaje de máquina. El primero, DeepBlue, logró implementar redes de lenguaje ante un escenario donde las opciones de conocimiento no habían sido previamente complicadas; el segundo, AlphaGo, combinó el método de Montecarlo con herramientas del aprendizaje de máquina para derrotar al campeón mundial del juego de Go (un juego con mucha mayor cantidad de combinaciones que el ajedrez).

Finalmente, como tercer análisis de los casos empíricos, se optó por centrarse en tres plataformas de uso cotidiano. En especial, esas plataformas emplean herramientas que se examinaron en la trayectoria de la inteligencia. Las tres plataformas que examinarán son DiDi, una plataforma de la economía gig; TikTok, una plataforma de entretenimiento y creación de contenido por parte del consumidor; y Facebook, una plataforma social.

Para concluir, se inquirirá el resultado del análisis de la trayectoria para comprender los retos y desafíos a los que se enfrenta la IA. De esta manera se comprenderá de mejor manera la trayectoria de innovación en torno a esta tecnología. Y finalmente se comprobará si los elementos socio culturales, políticos, sociales y tecnológicos, influenciaron en su desarrollo, ya sea desde la trayectoria de sus investigadores o desde el trayecto en torno a objetos específicos (imagen 8, de autoría propia).



(imagen 8, de autoría propia)

1. Tres canciones de una máquina pensante: Lovelace, Babbage y Turing

La IA comprende diferentes actores, cada uno anclado a un momento y circunstancia histórica que es importante examinar para comprender cómo se construyó esa rama de la academia, esa herramienta tecnológica y, más específicamente, lo que ese objeto representa. Por lo tanto, para analizar el impacto de la IA, este apartado examinará de manera transversal la historia de la IA y sus actores. Se creará un mapa de constelaciones, similar a una biografía económica y culturalmente informada, en torno a la trayectoria de la IA como actor y objeto de contorno en formación. Es importante destacar que una biografía económica y culturalmente informada es "una entidad construida culturalmente, dotada de significados culturalmente específicos, y clasificada y reclasificada en categorías construidas culturalmente" (Kopytoff, 1988 en Bell et al., 2018).

De esta manera, el análisis de la trayectoria de la IA que se realizará en este apartado se centrará en la interrelación de los actores humanos que la transformaron y fueron transformados por ella. Para ello, se utilizarán una serie de documentos centrados en las biografías de Ada Lovelace y Charles Babbage, así como en la de Alan Turing. En el caso de Turing, el trabajo también examinará su famoso ensayo "Computing machinery and intelligence" del cual se desprenden muchos de sus postulados sobre las máquinas inteligentes, así como una interesante discusión atemporal con Lovelace.

Precisamente, la idea de seres con IA es un campo de estudio que se puede rastrear hasta la antigua Grecia (Mijwel, 2015) en donde se concebían seres mecánicos creados por algunas deidades. Sin embargo, la noción de un aparato mecánico capaz de hacer cálculos inteligentes o con un resultado a partir de información previa se remonta concretamente al siglo XVIII, y específicamente a el año 1763, de la mano de un "teorema matemático para la probabilidad, que vino ser conocido como Teorema de Bayes" (Royal Society, 2017). Pero el término académico nació de la mano de John McCarthy y otras series de investigadores que se congregaron en una conferencia del Darmouth College "junto con Marvin Minsky, Nathaniel Rochester, y Claude Shannon 1959" (Mateo, 2020).

Dentro de los teóricos que concibieron a las máquinas como capaces de generar un resultado a partir de información previa, es importante mencionar a Ada Lovelace. Como señala Valverde (2019): "A mediados del siglo XIX, Ada Lovelace se dio cuenta de que las máquinas analíticas tenían el poder de resolver problemas de naturaleza compleja (Menabrea y Lovelace, 1843)". Y esta máquina analítica también la imaginaba y teorizó Charles Babbage, de quien Lovelace era asistente.

En particular, la máquina o motor analíticos fue un objeto interesante en dos sentidos. El primero, tiene que ver con su antecesora, la máquina diferencial; el segundo, la máquina diferencial terminó siendo un prototipo y la máquina analítica quedó delegada al reino de los esquemas y las teorías. La máquina diferencial, en estricto sentido, se puede considerar como una calculadora capaz de realizar

operaciones matemáticas precisas; por otro lado, la máquina analítica —en palabras de Lovelace—, podía representar sobre todo objetos que no son números (Chas, s.f) como las melodías o los símbolos.

La biografía de Lovelace resulta sumamente interesante. Desde el punto de vista de Solaeche (1994), Ava Lovelace se puede considerar como la primera programadora de la historia. Aunque su biografía se ha centrado en las personas de su vida, como su padre (el poeta Lord Byron, que la consideraba la hechicera de los números), su mayor aportación fue en torno a la idea de las máquinas inteligentes.

Ahora bien, es importante detenerse a examinar las contribuciones de Babbage. En cuanto al aspecto matemático, sus logros van más allá de la máquina diferencial y la propuesta de la máquina analítica. Según Stein (1985), Babbage hizo importantes aportaciones al campo del álgebra moderna y entendió que, tanto en el trabajo intelectual como en la manufactura, las herramientas utilizadas pueden hacer avanzar o retrasar los procesos. Cuando Ada trabajaba junto a Babbage, ella propuso que "podemos decir más acertadamente, que el motor analítico teje patrones algebraicos al igual que el telar Jacquard teje flores y hojas" (Carlucci Aiello, 2016). El telar de Jacquard era otro instrumento interesante que permitía tejer patrones sumamente complejos con la ayuda de tarjetas perforadas. Stein (1985), quien escribió una de las biografías de Ada, recoge un testimonio interesante sobre la vida de la matemática al ilustrar la forma en la que Sophia De Morgan narra la reacción de Lovelace ante la máquina diferencial.

«Recuerdo muy bien que la acompañé a ver el maravilloso motor analítico del Sr. Babbage. Mientras que otros visitantes miraban el funcionamiento de este hermoso instrumento con el tipo de expresión, y me atrevo a decir el tipo de sentimiento, que se dice que algunos salvajes mostraron al ver por primera vez un espejo o al oír una pistola, si, de hecho, tenían una idea muy clara de su maravilla: la señorita Byron, joven como era, comprendió su funcionamiento y vio la gran belleza de la invención. Había leído Cálculo diferencial hasta cierto punto y, después de casarse, prosiguió con el estudio y tradujo una pequeña obra del matemático italiano Menabrea en el que se explican los principios matemáticos de su construcción» (Stein, 1985).

Lamentablemente, Lovelace nunca pudo poner en práctica sus ideas ya que la máquina de Babbage nunca se completó debido a la falta de fondos. De alguna manera, el proyecto era ambicioso para su época y ni siquiera la máquina diferencial, menos demandante en recursos técnicos y monetarios, pudo terminarse. Según Stein, "al final, el Gobierno contribuyó con un total de £ 17.000 para la construcción (de la máquina diferencial), lo cual resultó ser una tarea mucho más larga y complicada de lo que se había estimado en un principio" (Stein, 1985); por consiguiente, la máquina diferencial no pasó de ser un bello prototipo y la máquina analítica una maravillosa idea que sembró las semillas que hicieron florecer una segunda y mejorada visión de la máquina diferencial (O'Regan, 2012).

La relación entre Lovelace y Babbage se puede definir como una de complementariedad. El primero era el visionario que intentaba construir complejas máquinas de cálculo y análisis, mientras que el segundo era la persona que vislumbraba todo su potencial y aplicaciones. Charman-Anderson describe esta relación al mencionar que: "Se dice que Babbage tenía una llamada Silver Lady, que podía inclinarse y levantar sus gafas a intervalos, como si se tratara de conocidos que pasaban" (Charman-Anderson, 2020).

La autora, además explica que quizás la mayor importancia de Ada Lovelace en torno a la IA yace en que ella vio lo que otros no podían en torno a las máquinas de Babbage, en especial la máquina analítica:

Uno de los mayores avances que puede dar la mente humana es la extrapolación de las capacidades actuales a las posibilidades futuras. El *arte de lo posible*, como se le ha llamado, es una habilidad esencial para los innovadores y emprendedores, pero imaginar una clase de máquina completamente nueva es algo para lo que pocas personas tienen la capacidad. El diseño de Babbage para el motor analítico fue asombroso, pero ninguno de sus compañeros pareció comprender realmente su significado. Ninguno excepto Lovelace (Charman-Anderson, 2020).

Hay que detenerse a examinar cómo se llegó a la idea de IA. Posiblemente la figura más famosa dentro del campo de la IA, especialmente en la computación, fue el británico Alan Turing. Durante la Segunda Guerra Mundial, el científico fue

convocado por el gobierno del Reino Unido para formar parte de un equipo encargado de descifrar el código Enigma. Es importante señalar que Turing conocía muy bien a Ada Lovelace (Carlucci, 2016).

El código Enigma se utilizaba en la Segunda Guerra Mundial mediante un aparato desarrollado y utilizado por los alemanes que generaba un lenguaje de comunicación cifrado altamente complejo. Por ello, a pesar de diversos e infructuosos intentos, los países aliados se enfrentaban a una barrera técnica que dificultaba generar estrategias para enfrentar los planes del aparato bélico del eje, basados en un código mutable generado con la ayuda de un dispositivo mecánico sumamente avanzado. En 1940, Alan Turing y su equipo decidieron construir un aparato electromecánico para romper el misterio detrás del Enigma.

Sin embargo, las aportaciones de Turing no se limitaron a la construcción de un aparato que cambió el rumbo de la guerra. La segunda aportación más importante del científico británico fue un artículo académico en el que propuso una prueba para detectar cómo una computadora podía hacerse pasar por un ser humano. La prueba de la imitación, en teoría, consistía en averiguar si un ser humano podía distinguir un diálogo entre un humano y un ente concebido artificialmente.

Hay que detenerse un momento para hablar del artículo "Computing machinery and intelligence" y en especial del Juego de la imitación de Turing. Este artículo académico ilustra la curiosidad e intuición que Turing tenía en torno al campo que más tarde sería conocido como IA. La primera línea del Juego de la imitación es la pregunta que Turing plantea: "¿Pueden pensar las máquinas? Esto debería comenzar con definiciones del significado de los términos máquina y pensar" (Turing, 1950).

El juego de la imitación se concibió para averiguar si las máquinas son capaces de imitar y hacerse pasar por humanos. De acuerdo con Turing, esta prueba se basa en un juego que solía jugar con sus conocidos: "Se juega con tres personas, un hombre (A), una mujer (B) y un interrogador (C) que pueden ser de ambos sexos. El interrogador permanece en una habitación separada de los otros dos. El objeto del juego es probar si el interrogador es capaz de determinar cuál de los otros dos

es el hombre y cuál es la mujer". En esta versión, se busca averiguar si una máquina puede engañar completamente a un interrogador humano.

Es importante detenerse a hablar del artículo "Computing machinery and intelligence" y en especial del Juego de la imitación de Turing. Este artículo académico ilustra la curiosidad e intuición que Turing tenía en torno al campo que más tarde se conocería como IA. La primera línea del Juego de la imitación es la pregunta que Turing plantea: "¿Pueden pensar las máquinas?" (Turing, 1950).

Uno de los planteamientos importantes de Turing es que "las computadoras digitales pueden explicarse diciendo que estas máquinas están destinadas a realizar cualquier operación que pueda realizar una computadora humana" (Turing, 1950). A pesar de que las computadoras digitales solo existían en el mundo teórico y de la ciencia ficción en ese momento, el matemático inglés logró plantear muchos de los elementos que ahora forman parte del mundo de la computación. Un ejemplo de esto son los tres componentes de una computadora: Almacenamiento (el lugar donde se realizan los cálculos), unidad ejecutiva (la herramienta para realizar los cálculos) y control (la parte que asegura que las reglas dadas a la unidad ejecutiva se sigan adecuadamente): Turing también fue más allá y planteó un elemento aleatorio que podría existir en algunas de estas computadoras, lo que ayudó a distinguirlas de la percepción teórica de que las computadoras solo sirven para representar cosas, pero no para crearlas.

Como ya lo mencionaron otros autores, Turing estaba consciente de la existencia de los estudios de Babbage y Lovelace. Eso se refleja cuando menciona: "La idea de una computadora digital es antigua. Charles Babbage, profesor lucasiano de matemáticas en Cambridge de 1828 a 1839, planeó una máquina de este tipo, llamada Máquina Analítica, pero nunca se completó" (Turing, 1950). Más aun, la discusión de las ideas de Lovelace se hace palpables cuando debate con ella a partir de su ensayo.

"Una variante de la objeción de Lady Lovelace afirma que una máquina *nunca puede hacer nada realmente nuevo*. Esto se puede detener por un momento con la sierra, No hay nada nuevo bajo el sol. ¿Quién puede estar seguro de que el trabajo original que ha realizado no fue simplemente el crecimiento de la semilla

plantada en él mediante la enseñanza, o el efecto de seguir principios generales bien conocidos?" (Turing, 1950).

Turing planteó que la postura de Lovelace se debía a la falta de capacidad de almacenamiento. El autor explica: "La crítica de que una máquina no puede tener mucha diversidad de comportamiento es solo una forma de decir que no puede tener mucha capacidad de almacenamiento" (Turing, 1950). Esto se relaciona con las capacidades técnicas de la época, las cuales influenciaron las posturas de Lovelace: "Se notará que no afirma que las máquinas en cuestión no tuvieran la propiedad, sino que la evidencia disponible para Lady Lovelace no la animó a creer que la tenían".

En torno a las reflexiones sobre Lovelace, Turing hace críticas sobre muchos de los dogmas que circulaban en torno a las computadoras. Un apartado interesante es cuando critica a sus colegas al mencionar que "Debe investigarse una gran parte del espacio-tiempo para obtener resultados fiables. De lo contrario, podríamos (como la mayoría de los niños ingleses) decidir que todo el mundo habla inglés y que es una tontería aprender francés" (Turing, 1950). En torno a estas reflexiones, Turing comienza a vislumbrar otro aspecto que le permitiría ser inteligente a una máquina.

Así pues, Turing introduce el error como elemento de la inteligencia: "Al observar los resultados de su propio comportamiento, puede modificar sus propios programas para lograr algún propósito de manera más efectiva. Estas son posibilidades del futuro cercano, más que sueños utópicos" (Turing, 1950). Así, comienza a vislumbrar un proceso de inteligencia que se desarrolla de abajo a arriba (bottom-up), donde la máquina adquiere inteligencia a partir de la experiencia y no mediante comandos e instrucciones que imitan la inteligencia.

La observación anterior es muy importante. Las postulaciones de la inteligencia mediante el aprendizaje son los fundamentos del campo del aprendizaje de máquina y aprendizaje profundo que sacaron a la IA de la oscuridad. Para explicar esto, Turing desarrolla la siguiente analogía: "En lugar de intentar producir un programa para simular la mente adulta, ¿por qué no intentar producir uno que

simule la del niño? Si esto fuera sometido a un curso de educación apropiado, se obtendría el cerebro adulto” (Turing, 1950).

Otro punto interesante que se puede rescatar de la disertación sobre IA en el ensayo de Turing es el concepto de aprendizaje supervisado. El matemático planteó que se debe experimentar con la enseñanza de las máquinas para encontrar con qué herramientas y técnicas aprenden mejor, y así afinar los procesos para obtener mejores resultados.

Hay que pensar un momento más sobre las implicaciones de la biografía científica de estos tres matemáticos. A pesar de que existieron en localizaciones distintas, salvo Lovelace y Babbage –quienes colaboraron juntos–, sus aportaciones y discusiones parecen ser elegantemente atemporales, pero eficientes: tres personas discutiendo bajo el cobijo de la curiosidad y con el peso intempestivo de las palabras que desafían el tiempo. Aunque, eternamente quizás, Latour lo explica mejor:

“...tenemos que pasar de una temporalidad a la otra, ya que una temporalidad, en sí misma, no tiene nada de temporal. Es un medio de conectar entidades y archivarlas. Si cambiamos el principio de clasificación, obtenemos una temporalidad diferente sobre la base de los mismos eventos” (Latour, 1993, p. 75).

La idea central de Latour es que en ningún momento nació lo moderno. Su afirmación es relevante para explicar la trayectoria de un objeto como la IA a partir de los agentes. Es decir, porque el actor tiene como rol la “doble construcción, ciencia con sociedad y sociedad con ciencia, surge de un conjunto de prácticas que la noción de deconstrucción capta de la mejor manera posible” (Latour, 1993, p. 6); es decir, esta sociedad y ciencia que la contiene surge de las asociaciones de actores al tiempo que construye a estos actores a partir de las redes en las que estos están inscritos.

Así pues, se puede ver que el entorno tanto en la práctica como en la teoría de la computación es atemporal y atravesado por constantes traducciones por parte de sus agentes humanos y no humanos. Por un lado, al examinar sus aportaciones, ya sea por fuentes primarias o secundarias, se puede apreciar que Lovelace y

Babbage concibieron un aparato que estaba más allá de las posibilidades técnicas y materiales de su época. Por otro lado, a pesar de que Lovelace y Babbage no eran contemporáneos de Turing, este último retomó muchas de sus ideas: "En su búsqueda secreta de la máquina para descifrar mensajes, Turing se encontró estudiando los escritos de una mujer que había vivido un siglo antes: Ada Lovelace" (Chas, s.f).

Hasta ese momento, la IA no era un objeto estable cuyas características permanecieran intactas a pesar de las distintas interpretaciones a las que estaba sujeto. En primer lugar, no tenía una forma constante cuyos significados y aplicaciones pudieran ser reconocidos; en segundo lugar, era una idea y objeto que aún estaba siendo moldeado por las concepciones teóricas y desarrollos tecnológicos, lo cual hacía que fuera difícil de interpretar. Por lo tanto, la IA, en sus primeras etapas, era lo que Latour llama un "cuasi objeto" que puede "pensarse como un actor en movimiento, porque transforma el objeto en movimiento" (Latour, 2017).

Para concluir, la historia temprana de la IA ayuda a comprenderla como un "cuasi objeto". Bruno Latour explica que "Los cuasiobjetos son mucho más sociales, mucho más fabricados, mucho más colectivos que las partes 'duras' de la naturaleza, pero de ninguna manera son los receptáculos arbitrarios de una sociedad en toda regla" (Latour, 1993, p. 55). Sin embargo, a pesar de que la IA no fue influenciada directamente por la sociedad, sí fue afectada por las historias de los primeros actores que la concibieron con distintos nombres, aplicaciones y pruebas.

2. McCarthy, el padrino de la IA

La siguiente figura importante en torno a la IA es John McCarthy. Según Hayes y Morgenstern (2007), las contribuciones de McCarthy a la IA son nada menos que extraordinarias, por ejemplo: "Revolucionó el uso de computadoras con sus innovaciones en el tiempo compartido; inventó Lisp, uno de los lenguajes informáticos más duraderos en uso; hizo contribuciones sustanciales a los primeros trabajos en la teoría matemática de la computación; fue uno de los

fundadores del campo de la IA" (Hayes y Morgenstern, 2007). Pero quizás su mayor contribución fue reunir a distintos científicos en el campo de la IA para que esta disciplina floreciera adecuadamente.

En cuanto a su contribución específica a la disciplina de la IA, Hayes y Morgenstern (2007) explican que "Lo que distinguió el plan de McCarthy fue su énfasis en el uso de la lógica matemática como lenguaje para representar el conocimiento que debería tener una máquina inteligente y como medio para razonar con ese conocimiento". Esto está relacionado con cómo funciona la IA, ya sea mediante un razonamiento probabilístico o mediante funciones programadas.

La vida de McCarthy tuvo su dosis de tragedia. Hijo de migrantes, él vivió la pérdida de su hogar a temprana edad. Sin embargo, desde muy joven, McCarthy demostró una cualidad que le ayudaría a superar muchas de sus desventajas materiales. En el texto principal para este apartado, "On John McCarthy's 80th Birthday, in Honor of His Contributions", Hayes y Morgenstern (2007) explican que "Como muchos niños prodigio, John McCarthy fue en parte autodidacta. Debido a una enfermedad de la infancia, comenzó la escuela con un año de retraso, pero rápidamente recuperó el tiempo por su cuenta, saltó varios grados y terminó graduándose de la escuela secundaria dos años antes".

Los códigos Nazi eran creados por una máquina conocida como Enigma. Desde 1920, este tipo de códigos fueron muy perjudiciales para la seguridad de Polonia. Como se mencionó anteriormente, el matemático inglés Alan Turing desarrolló una máquina electrónica conocida como Bombas que logró descifrar ese código continuando el trabajo del criptógrafo polaco Marian Rejewski (Inman, 2020); por lo tanto, se ve una relación entre agentes humanos y no humanos que germinaron en la curiosidad de McCarthy.

Es importante destacar que los objetos siempre hablan. Las maquinaciones de Turing despertaron el interés de McCarthy en las computadoras y por lo que estas podían hacer. Hayes y Morgenstern (2007) mencionan: "Para McCarthy, el fermento intelectual en ese momento era palpable. La emoción generada por el

uso de las primeras computadoras para decodificar los mensajes enemigos y, por lo tanto, ayudar a ganar la Segunda Guerra Mundial, todavía estaba en auge".

La próxima etapa de la vida de McCarthy estuvo marcada por una conexión más cercana con uno de los investigadores mencionados anteriormente. En 1948, asistió al Simposio Hixon sobre Mecanismos Cerebrales en la Conducta (Jeffress, 1951), una conferencia que reunió a líderes en diferentes áreas relacionadas con las ciencias cognitivas, incluidos los matemáticos Alan Turing y Claude Shannon y el psicólogo Karl Lashley (Hayes y Morgenstern, 2007). Las discusiones durante el simposio tuvieron un gran impacto en McCarthy, despertando su interés en hacer que las máquinas pensaran como personas (Rajaraman, 2014). Así pues, el interés de McCarthy en la IA no fue el resultado de un momento de inspiración, sino de una serie de influencias recibidas de distintos investigadores, entre ellos Alan Turing.

La carrera académica de McCarthy progresó sin obstáculos. En solo dos años, completó su doctorado en la Universidad de Princeton y luego se desempeñó como instructor en la misma institución durante dos años más. Según Hayes y Morgenstern (2007), fue durante su estancia en Princeton donde conoció a Marvin Minsky y rápidamente establecieron un vínculo en torno a las máquinas inteligentes, aunque en ese momento aún no se llamaba IA.

Es importante mencionar que Minsky tuvo varios logros relevantes, como la construcción de la primera máquina de aprendizaje de redes neuronales cableada aleatoriamente (Durand-Lose y Margenstern, 2007) y, junto con Seymour Papert, la publicación del libro *Perceptrones*, un texto controvertido en el que se hace una crítica fuerte a un trabajo relacionado con las redes neuronales y que se considera en parte responsable del llamado "invierno de la IA" (Olazaran, 1996 y Wikipedia, s.f.).

En cuanto a McCarthy, después de un intento fallido de publicar una serie de ensayos bajo el término genérico de IA, formalizó el término al proponerlo como título para una conferencia. La conferencia de IA en Dortmund, realizada en 1956, contó con la participación de Minsky, Shannon y Nathaniel Rochester, un

investigador de IBM (Hayes y Morgenstern, 2007 y Rajaraman, 2014). Aunque la conferencia contaba con investigadores interesados en distintos campos, tuvo como objetivo principal "conseguir que cuatro investigadores que estaban trabajando en el campo: McCarthy, Minsky, Newell y Simon se reunieran, hablaran y planificaran futuros proyectos de investigación en IA" (Hayes y Morgenstern, 2007).

La conferencia de Dortmund refleja el papel del entorno social en la formación de objetos. McCarthy tuvo que negociar no solo para obtener recursos para el congreso, sino también para consolidar el término de IA tanto con sus colegas como con instituciones, enfrentando el rechazo de otros investigadores como Claude Shannon en el proceso. Según Star y Griesemer (1989), esto se debe a que "cada mundo social tiene jurisdicción parcial sobre los recursos representados por ese objeto, y los desajustes causados por la superposición se convierten en problemas de negociación".

McCarthy pasó los siguientes dos años en Dortmund hasta que la universidad le dio la oportunidad de dar clases durante un año en cualquier universidad de su elección. Decidió asistir al MIT en 1956 (Hayes y Morgenstern, 2007). La elección del MIT no solo se debió a la reputación de esa universidad, sino también a los recursos materiales a los que tendría acceso, especialmente a una de las pocas computadoras IBM disponibles en esa universidad. Además, al estar en la zona de Boston, McCarthy podría volver a tener contacto con Minsky, quien en ese entonces era profesor en Harvard.

De nuevo se puede apreciar cómo McCarthy se movía en torno a un objeto y los actores no humanos relacionados con él. Sus decisiones no solo se basaban en intereses económicos, sino también en las posibilidades materiales y colectivas que le permitieran acercarse a la materialidad de su objeto de estudio y a colegas cercanos que compartían sus intereses. Star y Griesemer (1989) ayudan a explicar las acciones de McCarthy al proponer que "el problema para todos los actores de una red, incluido el emprendedor científico, es reducir (temporalmente) su incertidumbre local sin correr el riesgo de perder la cooperación de los aliados".

Después de su periodo como profesor invitado en MIT, esa misma universidad le ofreció una plaza como profesor titular. Según Hayes y Morgenstern (2007), los seis años que McCarthy pasó en MIT se convirtieron en uno de los periodos más productivos de su vida. Por un lado, se dedicó a desarrollar teoremas matemáticos de su interés y, sobre todo, a trabajar en el razonamiento de las máquinas. Por otro lado, logró colaborar directamente con Minsky, quien se unió al MIT en 1957.

Es importante señalar que McCarthy y Minsky no siempre tuvieron una relación tersa. Minsky fue muy crítico de la perspectiva lógica que McCarthy intentaba desarrollar en torno a la IA. En cambio, la perspectiva de Minsky se centró en “Representaciones conceptuales de alto nivel de la lógica y el pensamiento” y en una perspectiva donde interactúan diferentes agentes (Knight, 2020). Esta disputa entre McCarthy y Minsky permite notar que la IA es un agente de contorno de límites coincidentes cuyos límites son iguales, pero poseen contenidos internos distintos.

Además, el trabajo de McCarthy durante su estancia en el MIT tampoco era bien visto por el resto de sus colegas:

“Su trabajo era radicalmente diferente al de sus colegas de IA en al menos dos formas. En primer lugar, el trabajo anterior en IA se había centrado en conseguir que una computadora reprodujera las actividades que son un desafío para los humanos, como jugar al ajedrez y demostrar teoremas de las matemáticas. En contraste, McCarthy estaba preocupado por tareas mundanas y aparentemente triviales, como construir un plan para llegar al aeropuerto” (Hayes y Morgenstern, 2007).

En 1962, John McCarthy aceptó una oferta de la universidad de Stanford, pero solo después de haber convencido a la universidad para que le diera una plaza de profesor de tiempo completo. Según sus propias palabras, Stanford no tuvo otra opción más que ceder a su petición debido a que reconocían el enorme potencial de sus áreas de estudio, especialmente en IA (Hayes y Morgenstern, 2007).

En su estancia en Stanford, y hasta su retiro en 2001, McCarthy siguió trabajando en proyectos relacionados con la computación, las matemáticas y el razonamiento. Su legado más importante fue cómo cristalizó y luego tradujo la noción de IA como un objeto constante cuya definición está en disputa. Además, al igual que él fue

influenciado por Turing, quien también tomó nota de las teorías de Babbage y Lovelace, sus ideas servirían para la formación de importantes institutos de IA en las universidades de Texas en Austin, Linköping en Suecia, el Imperial College de Londres y la Universidad de Toronto (Hayes y Morgenstern, 2007).

En cuanto a sus ideas más íntimas en torno a la IA, McCarthy dejó una serie de notas y blogs para que el público los consultara. A continuación, se presentarán algunas preguntas que la gente le hacía, y por supuesto, las respuestas que él daba:

- “P. Sí, pero ¿qué es la inteligencia?
- R. La inteligencia es la parte computacional de la capacidad para lograr metas en el mundo. En las personas se producen diversos tipos y grados de inteligencia, así como en muchos animales y algunas máquinas.
- P. ¿La IA no se trata de simular la inteligencia humana?
- R. A veces, pero no siempre o incluso habitualmente.
- P. Por un lado, podemos aprender algo sobre cómo hacer que las máquinas resuelvan problemas observando a otras personas o simplemente observando nuestros propios métodos. Por otro lado, la mayoría de trabajo en IA implica estudiar los problemas que el mundo presenta a la inteligencia en lugar de estudiar personas o animales. Los investigadores de IA son libres de utilizar métodos que no se observan en las personas o que implican mucha más información de lo que la gente puede hacer.
- P. ¿Qué es la prueba de Turing?
- A. El artículo de Alan Turing de 1950 *Computing Machinery and Intelligence* discutieron las condiciones para considerar que una máquina es inteligente. Argumentó que si la máquina pudiera pretender ser humana con éxito para un observador experto, entonces ciertamente debería considerarla inteligente. Esta prueba satisfaría a la mayoría de la gente, pero no a todos los filósofos. El observador podría interactuar con la máquina y un humano por teletipo (para evitar requerir que la máquina imita la apariencia o la voz

de la persona), y el humano trataría de persuadir al observador de que era humano y que la máquina trate de engañar al observador” (McCarthy, 2007).

En suma, si se observa detenidamente las preguntas y respuestas contenidas en el texto *What is artificial intelligence* (2007), se puede apreciar la forma en que McCarthy intenta traducir los contenidos generales, pero constantemente cambiantes de la IA a un público más allá de las aulas y pasillos de la academia. De nuevo, esta habilidad para comunicar sus ideas fue lo que le permitió avanzar en el campo cuyo término él acuñó. Además, esto ayuda a entender que la historia de los actores en torno a la IA no es aislada, sino que está influenciada por objetos sin límites definidos o en proceso de definición, constreñida por situaciones políticas y económicas y contrastada por distintos actores humanos y no humanos.

3. De los inviernos de la IA a la primavera de los objetos

A. El perceptrón de Rosenblatt

La siguiente parte de la historia de la IA es protagonizada por los objetos en los que se materializa. Sin embargo, sigue siendo una historia donde los actores humanos siguen siendo importantes, ya que la historia de la IA después del entusiasmo inicial que despertó este campo del conocimiento estuvo marcada por decepciones, culpas y señalamientos. Aun así, también fue un momento en el que se dieron importantes avances ligados a máquinas y algoritmos impulsados, diseñados e implementados por actores que no dejaron de ver el potencial en este campo.

Según la Real Sociedad: "El perceptrón de Frank Rosenblatt fue un primer intento de crear una red neuronal con resistencia giratoria (potenciómetro) accionada por un motor eléctrico". Esta máquina podría tomar una entrada, digamos, los píxeles de una imagen, y crear una salida, como una etiqueta". Un perceptrón, en el campo del aprendizaje máquina, es un algoritmo que se utiliza para aprendizaje supervisado. Es decir, se etiquetan una serie de funciones, una serie de datos de entrenamiento, y se espera que la máquina genere un resultado (Freund y

Schapiro, 1999); por ejemplo, y de forma simplificada, un algoritmo de aprendizaje se articula mediante diferentes funciones binarias o nodos.

A diferencia de Lovelace, Turing y McCarthy, la trayectoria de Frank Rosenblatt no tuvo sus raíces en las matemáticas. Frank Rosenblatt se graduó en psicología en la Universidad de Cornell. Y quizás lo anterior, la formación de Rosenblatt ayuda a explicar cómo el investigador “Se inspiró en el trabajo de la neurociencia en la década de 1940, lo que lo llevó a crear una réplica burda de las neuronas del cerebro” (Schuchmann, 2019).

Junto a sus antecedentes en la psicología y posteriormente en la neurociencia, Frank Rosenblatt demostró un enorme entusiasmo, no solo alrededor de su perceptrón, sino de la idea de las máquinas inteligentes:

“—Las historias sobre la creación de máquinas que tienen cualidades humanas han sido durante mucho tiempo una provincia fascinante en el ámbito de la ciencia ficción—, escribió Rosenblatt en 1958.—Sin embargo, estamos a punto de presenciar el nacimiento de una máquina así, una máquina capaz de percibir, reconocer e identificar su entorno sin ningún entrenamiento o control humano—” (Lefkowitz, 2015).

Este entusiasmo fue replicado por la prensa. Lefkowitz (2015) recogió una cita del periódico *New York Times* que decía: “El psicólogo (Rosenblatt) muestra el embrión de una computadora diseñada para leer y crecer más sabiamente...De hecho, nos parece el primer rival serio del cerebro humano jamás concebido”. En esa misma publicación, el diario de *The New York Times* describió que se espera que el perceptrón, el *embrión de un cerebro humano*, “camine, hable, vea, escriba, se reproduzca a sí mismo, y sea consciente de su propia existencia” (The New York Times, 1958).

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo
of Computer Designed to
Read and Grow Wiser

WASHINGTON, July 7 (UPI)—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's \$2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of \$100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human be-

ings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

1958 New York Times...

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

Rosenblatt generó lo que Schuchmann (2019) considera un ciclo de entusiasmo basado en las expectativas de la IA. Dentro de este ciclo, también se puede explicar la llegada de lo que se conoce como el primer invierno de la IA. Había enormes expectativas en torno a lo que la IA podría ser y, por tanto, había recursos económicos destinados a universidades o centros de investigación, además del apoyo de la iniciativa privada, pero sobre todo del gobierno.

Según Schuchmann (2019), el ciclo de entusiasmo, junto con los elementos sociales, históricos y tecnológicos planteados en este análisis, permite comprender de forma holística la decepción en torno a la temprana IA. El ciclo de entusiasmo consta de distintas fases: "Las fases clave del ciclo son el pico, donde el interés y la emoción están en su punto más alto, y el punto más bajo de la desilusión, donde la expectativa inicial no se puede cumplir. Aquí, el interés en el campo está en su nivel más bajo. Luego, el campo se recupera lentamente y

alcanza la meseta de productividad”. Esto se pudo apreciar claramente en lo que cuando el interés, la emoción, se manifestaron tanto en medios de comunicación, así como en escenarios académicos.

Desde la perspectiva de la teoría del actor-red, el ciclo de entusiasmo puede ser visto como una dinámica de asociación y disociación entre actores humanos y no humanos que participan en la construcción de la IA. Los actores humanos, como los científicos, los medios de comunicación, los inversores y el público en general, se relacionan con los actores no humanos, como las máquinas, los algoritmos, los datos y los conceptos, para crear expectativas, promesas, narrativas y proyectos sobre la IA. Sin embargo, estas asociaciones pueden ser frágiles o inestables, y pueden romperse cuando las expectativas no se cumplen, cuando surgen problemas técnicos o éticos, o cuando cambian las condiciones sociales o históricas; entonces, los actores se disocian y pierden interés o confianza en la IA. Pero esto no significa el fin de la red, sino que abre la posibilidad de nuevas asociaciones que pueden reactivar el ciclo de entusiasmo y generar innovaciones o transformaciones en el campo.

Así pues: “...las limitaciones tecnológicas, como el perceptrón experimentado en la década de 1960, inhibieron el progreso. El perceptrón, del que se esperaba que pronto fuera consciente de su existencia, no pudo resolver el problema XOR en ese momento” (Schuchmann, 2019). Además de eso, comenzaron a surgir documentos que cimbraron dudas en torno a la IA, en especial el aprendizaje máquina que estaba relacionado con el perceptrón. Y no hubo documento más lapidario en contra de la IA que el Informe de Lighthill.

Finalmente, el perceptrón, como modelo temprano de red neuronal, estableció las bases para diversas técnicas de inteligencia artificial, incluido el aprendizaje automático (Cruz, 2023). No obstante, desde su concepción, también reveló desafíos en torno a la explicabilidad, es decir, la capacidad de comprender cómo toma decisiones, una problemática que persiste en modelos más avanzados de redes neuronales (Alvarez-Melis & Jaakkola, 2018). Estos desafíos sentaron las bases para problemas más complejos, como los sesgos algorítmicos derivados de

los datos utilizados en el entrenamiento de redes neuronales profundas y modelos de lenguaje de gran escala (Páez, 2021).

B. El reporte Lighthill

“En la década de 1970, estaba claro que el progreso en el campo no fue tan rápido como se esperaba” (Royal Society, 2017).

De acuerdo con la Real Sociedad, el informe conocido como El informe Lighthill generó un ambiente de apatía en torno a la investigación de la IA y sus diferentes ramas como el aprendizaje de máquina. Así pues, el informe, junto con las limitaciones tecnológicas de la época—como la falta de grandes volúmenes de datos y la escasez de poder de cómputo, esenciales hoy para modelos como ChatGPT—frenaron el desarrollo de la IA. Ejemplo de ello fue el perceptrón, experimentado en la década de 1960, cuyo desempeño limitado contribuyó a la desilusión en el campo y dio lugar a lo que se conoce como el primer invierno de la IA (Royal Society, 2017).

Michael James Lighthill, nacido en 1925, fue el autor cuyo apellido dio nombre al informe. A los quince años ganó una beca para estudiar en el Trinity College de Cambridge (Agar, 2020). Durante la Segunda Guerra Mundial, Lighthill trabajó en el laboratorio Nacional de Física en aerodinámica supersónica e hipersónica, y más tarde se integró a la Universidad de Manchester. Su visión académica y científica se centró en las matemáticas aplicadas y en lo que podría utilizarse en la realidad inmediata.

De acuerdo con Agar (2020), "en la UCL, destacó la necesidad de que las universidades produzcan graduados en materias útiles como la construcción." En 1973, después de que el gobierno británico le encargara hacer una evaluación científica y académica del estado de la IA, Lighthill entregó un documento en el que hizo serias críticas a los avances prácticos en el área de la robótica y los sistemas de procesamiento de lenguaje. Quizás su visión pragmática en torno a la aplicación de la ciencia en la realidad ayuda a explicar por qué Lighthill mostró una

postura tan crítica en torno a los avances en la IA, especialmente en su famoso informe.

Ahora bien, el informe nació en un contexto de creciente escepticismo sobre la viabilidad de la inteligencia artificial (IA) en la década de 1970. Durante los años 50 y 60, la IA había experimentado un notable auge gracias a los avances en algoritmos de búsqueda, redes neuronales y procesamiento de lenguaje. Sin embargo, a medida que las expectativas superaban los resultados tangibles, los gobiernos comenzaron a cuestionar la inversión en esta disciplina. Dado que, en el Reino Unido, la investigación en IA dependía en gran medida de fondos públicos, y el Consejo de Investigación Científica (Science Research Council, SRC) solicitó una evaluación objetiva de su progreso. En este contexto, Lighthill presentó un informe que enfatizaba las limitaciones técnicas y la falta de aplicaciones prácticas, lo que contribuyó a una drástica reducción en el financiamiento gubernamental y al inicio del primer “invierno” de la IA (Boden, 2018); es decir, el informe crítico la falta de aplicaciones productivas reales del naciente campo de la IA.

El informe fue muy crítico en el lenguaje y la robótica, es importante explicar cómo estaba organizado. Según Agar (2020), el informe constaba de tres partes (A, B, C). La parte A se centra en la ingeniería y el proceso de automatización; la parte B se enfocó en examinar los avances en la coordinación de ojos, sistemas de lenguaje, juegos y resolución de problemas de sentido común; finalmente, la parte C se centra en las redes neuronales, sistemas de reconocimiento de imágenes y la adquisición de conocimientos y habilidades. La parte que salió mejor favorecida en el diagnóstico de Lighthill fue la A, lo cual no resulta sorprendente considerando su postura académica y científica.

Así pues, el reporte Lighthill, publicado en 1973, fue un informe encargado por la Royal Society del Reino Unido para evaluar el estado y las perspectivas de la inteligencia artificial (IA) en ese momento. Redactado por el matemático James Lighthill, concluyó que los avances en IA eran limitados y que los enfoques existentes no eran capaces de resolver problemas complejos del mundo real de

manera eficiente. Además, crítico duramente las promesas excesivas de la IA y destacó tres áreas de investigación:

1. IA avanzada o “robusta”, que busca simular el comportamiento humano.
2. Sistemas de procesamiento de datos comerciales, utilizados en aplicaciones prácticas.
3. Sistemas de modelado del cerebro, con enfoques inspirados en la neurociencia.

Lighthill argumentó que solo el segundo grupo había logrado avances significativos, mientras que los otros dos carecían de aplicaciones viables. Como consecuencia, el informe llevó al recorte drástico de fondos para la investigación en IA en el Reino Unido, marcando el inicio de lo que se conoce como el primer invierno de la IA. Este periodo de desilusión afectó el desarrollo de la IA en las décadas siguientes, hasta el resurgimiento de nuevas técnicas en los años 80.

El informe fue muy influyente y redujo la financiación y el apoyo a la investigación en IA, especialmente en el lenguaje natural y la robótica, considerados poco prometedores o difíciles. La parte A, entonces, que se ocupaba de la ingeniería y la automatización, fue la única que recibió una valoración positiva por parte de Lighthill, mientras que las partes B y C, que trataban temas más relacionados con la cognición humana, fueron criticadas duramente.

Ahora bien, independientemente de cómo se hizo el análisis dentro del informe de Lighthill y cuáles eran las posturas de este en torno a las fronteras de la IA, las consecuencias que tuvo fueron dramáticas. Según Agar (2020), "La causa del primer invierno de la IA fue un informe crítico encargado por el SRC, escrito por el matemático James Lighthill. Específicamente, Crevier sostiene que la causa de la recesión fue la crítica de Lighthill a los grandiosos objetivos de la IA". Mientras tanto, Schuchmann (2019) sostiene que "...el recorte de la financiación tuvo un gran impacto en la investigación en ambos inviernos de IA. En el primer invierno de IA, el informe Lighthill provocó un recorte de fondos para todas las

universidades del Reino Unido, excepto dos, y provocó además recortes en Europa y EE. UU”.

Aun así, Agar (2020) sugiere que "si bien el informe de Lighthill hizo que la investigación en IA se detuviera, sugiero que la crítica de Lighthill se ha malinterpretado". Además, el autor plantea que "Lighthill adopta una postura a favor de los mundos laborales: considera que la orientación de la buena ciencia es la resolución de problemas y articula las responsabilidades y respuestas de los científicos". A pesar de las posturas de Agar (2020), se puede argumentar que el informe de Lighthill fue un factor clave en el proceso de decepción con respecto a esta área de innovación.

Cabe señalar que el "invierno de la IA" ocurrió durante la Guerra Fría. Según la Universidad de Washington (2006), el gobierno de Estados Unidos dejó de apoyar el proyecto de los traductores de máquina debido a que no demostraron éxito en los procesos de traducción del ruso al inglés y viceversa. Por consiguiente, disminuyeron los recursos y muchas universidades de Estados Unidos dejaron de impulsar el estudio de las máquinas.

Asimismo, hubo prácticas académicas cuestionables en torno a esta tecnología: "Desafortunadamente, el éxito inicial fue artificial: el experimento Georgetown-IBM fue manipulado para esa ocasión en particular. Las reglas de gramática y vocabulario se crearon específicamente para tratar las muestras de texto en particular, de modo que el sistema pudiera aparecer de la mejor manera" (Universidad de Washington, 2006). Así pues, con tal de conseguir recursos económicos, se generó un escenario que facilitase el espectáculo a costa de la realidad empírica en torno a la IA.

Además de los problemas técnicos que evidenciaron los problemas de la IA, hubo elementos sociales que también contribuyeron a su fracaso y el inicio de un largo "invierno". Uno de estos factores fue la excesiva expectativa en torno a la IA. La gente pensaba que en pocas décadas la IA sería parte de su vida cotidiana y con funciones sorprendentes. Las personas imaginaban que la IA asumiría el papel de

esclavos cibernéticos de robots que formarían parte de cada espacio social, pero esto no sucedió.

Sin embargo, la realidad científica y las expectativas excesivas en torno a la IA, ligadas a situaciones tecnológicas y contextos particulares, como la falta de información y datos, así como el lento avance en el proceso de cómputo, hicieron que estos escenarios se volvieran parte del reino de la ciencia ficción. Además, esto favoreció que el entusiasmo en torno a la IA, en especial el aprendizaje de máquinas se esfumara, incluyendo el apoyo del público y de los gobiernos.

De nuevo, es importante volver al concepto del ciclo de entusiasmo. Schuchmann (2020) explica que hay dos elementos que pueden explicar por qué la IA pasó de ser una ciencia de frontera que generaba entusiasmo, a una anécdota que no logró cumplir sus promesas durante la Guerra Fría. El autor resume la llegada del primer "invierno" en torno a tres elementos: primero, las expectativas y promesas en comparación con los resultados reales; segundo, la financiación de los gobiernos e industrias; y tercero, las limitaciones tecnológicas (Schuchmann, 2019).

Es importante detenerse en el segundo punto. Aunque las redes neuronales ligadas a la idea de las máquinas inteligentes dieron resultados prometedores, estaban limitadas por la tecnología de la época. Este elemento provocó que el factor de entusiasmo en torno a la tecnología de aprendizaje de máquina, en especial las habilidades de reconocimiento de voz, se desvaneciera, y con ello, el financiamiento que esta área de investigación recibía.

Además, hubo dos documentos importantes, el ya mencionado Informe de Lighthill y el libro de Perceptrones de Minsky y Papert, en los cuales "se demostró que los perceptrones solo se podían entrenar para resolver problemas lineales separables...plantando obstáculos que impedían el progreso de los perceptrones" (Schuchmann, 2019). Aunque estos documentos no sepultaron por sí mismos a la IA, sí enfocaron la luz sobre sus limitaciones, especialmente en las redes neuronales que estaban limitadas por problemas técnicos y por la situación tecnológica de la época.

Hay que explicar que según Latour (1993), el concepto de traducción se refiere a “las negociaciones, intrigas, actos de persuasión o violencia, por los cuales un actor consigue la adhesión de otros y teje una red” (p. 10). La traducción es el acto de crear relaciones e identidades entre actores humanos y no humanos, que pueden ser frágiles o inestables. Esto es muy importante, porque cuando la traducción fracasa, la estabilidad del programa y sus asociaciones se van abajo.

Y de acuerdo Callon (1986), el concepto de traducción se refiere a “las negociaciones, intrigas, actos de persuasión o violencia, por los cuales un actor consigue la adhesión de otros y teje una red” (p. 203). La traducción es el proceso de construir relaciones e identidades entre actores heterogéneos, humanos y no humanos, que pueden ser modificados o desplazados por otros actores. Y este término ilustra y explica claramente lo qué ocurrió en el invierno de la IA.

Con base en la teoría de actor-red, y retomando uno de sus conceptos más importantes, hubo un error en la traducción. El programa prometía una revolución en la IA, pero los actores involucrados no pudieron traducir claramente sus metas a otros miembros de la red que les proporcionaban financiación. Esto generó una disolución de las relaciones e identidades entre los actores heterogéneos que participaban en el programa, y provocó una resistencia y oposición por parte de los financiadores.

En resumen, el *invierno de la IA* (el primer invierno de la IA) y los inviernos subsecuentes, se pueden atribuir a tres pasos fundamentales. El primero es el nacimiento de una nueva tecnología que promete muchas cosas y que demuestra avances significativos. El segundo es que esta tecnología comienza a demostrar ciertas limitaciones cuando, después de sus éxitos iniciales, se enfrenta a ciertas limitaciones en lo que realmente puede hacer. El tercer elemento es la decepción, generada por una expectativa enorme de académicos, de la prensa, y de entusiastas en torno a esta tecnología de IA que provocó un sentimiento de decepción: las nieves del invierno cayeron en torno a esta promesa de innovación tecnológica.

C. Tercera parte, del segundo invierno a la reivindicación de Rosenblatt

La IA tuvo un segundo auge, una segunda vida después del periodo conocido como el primer "invierno" de la IA. Con la aparición de los sistemas expertos, surgieron promesas tecnológicas que permitieron que muchos laboratorios recibieran apoyo tanto de industrias armamentistas como de empresas. Por lo tanto, durante la década de 1980, la IA comenzó a tener un segundo aire, un segundo suspiro de vida.

Sin embargo, estos sistemas expertos comenzaron a mostrar limitaciones, su dependencia de reglas rígidas los hacía inflexibles y poco adaptables a situaciones nuevas; tenían dificultades para manejar incertidumbre y razonar con información incompleta o ambigua; su escalabilidad era limitada, ya que al aumentar la complejidad se volvía difícil agregar nuevas reglas sin afectar el rendimiento; el desarrollo y mantenimiento eran costosos, pues requerían expertos en el dominio y en programación; carecían de capacidad de aprendizaje automático, ya que toda su inteligencia dependía de reglas estáticas introducidas manualmente; no podían capturar ni utilizar el conocimiento tácito o intuitivo de los expertos humanos, lo que restringía su aplicabilidad; además, eran incompatibles con el uso de grandes volúmenes de datos. Estas limitaciones lentamente provocaron un sentimiento de decepción. Poco a poco, en las conferencias sobre IA comenzaron a mostrar signos de desinterés. Las antes llenas salas de discusión sobre esta tecnología no tardaron en mostrar un gran número de lugares vacíos (UNIR Revista, 2022).

Aunque no hay grandes personajes o figuras que ayuden a describir este periodo de entusiasmo, se pueden identificar algunos entornos tecnológicos históricos que ayudan a comprender por qué la IA llegó al segundo "invierno". Para ello, primero debemos definir qué son los sistemas expertos y, posteriormente, analizar las críticas en torno a estos sistemas y las promesas que simplemente no se cumplieron.

Pero es importante observar lo que pasa en otras partes del mundo. Si bien durante el primer invierno de la IA algunas promesas tecnológicas como las redes

neuronales pasaron al olvido en diferentes universidades y países, hubo otros quienes recogieron las piezas y siguieron construyendo en torno esos cimientos. En especial, dos países siguieron promoviendo el desarrollo del aprendizaje de máquina el primero de ellos fue Japón, y es una enorme inversión que quizás no logró cumplir la promesa; y el segundo fue Canadá, en especial la Universidad de Toronto, que logró plantar las semillas que más tarde se convertirían en los frutos que permitieron que el aprendizaje de máquina se pusieron la tecnología prácticamente universal.

De acuerdo con Schuchmann (2019):

“En el corazón de la comercialización de la IA se encontraban los sistemas expertos. Estos sistemas fueron hechos a mano por expertos en topografía y creando conjuntos de reglas *si entonces* en consecuencia. El método se denomina enfoque *de arriba hacia abajo* de la IA y muchos creen que el conocimiento experto es la mejor manera de crear IA” (Schuchmann, 2019).

Los sistemas expertos no tardaron en recibir críticas importantes en el ámbito de la IA. Por ejemplo, en 1984, John McCarthy criticó los sistemas expertos porque carecían de sentido común y conocimiento sobre sus propias limitaciones (Schuchmann, 2019). La crítica de McCarthy se debía a la naturaleza propia de los sistemas expertos, ya que estos solo podían llegar a una decisión si la respuesta a la pregunta se encontraba en la regla de posibles soluciones.

Según Gillies (en Coursera, s.f): "Los primeros sistemas se basaban en reglas lógicas. Por ejemplo, los sistemas expertos eran un tipo de sistema de IA que intentaba hacer cosas que los humanos expertos pueden hacer" (Gillies, en Coursera, s.f). Los sistemas expertos recibieron mucho financiamiento de empresas y de la Agencia de Proyectos de Investigación Avanzada de Defensa (DARPA). Sin embargo, a medida que aumentaba el entusiasmo en torno a los sistemas expertos, muchos investigadores comenzaron a temer que se estaba entrando en un nuevo ciclo de entusiasmo, sin embargo:

“Hoy en día, muchos creen que Rosenblatt ha sido reivindicado. Los principios subyacentes al perceptrón ayudaron a desencadenar la revolución de la IA moderna. El aprendizaje profundo y las redes neuronales, que pueden clasificar

imágenes en línea, por ejemplo, o permitir la traducción de idiomas, están transformando la sociedad” (Lefkowitz, 2015).

De manera paradójica, a medida que empezaba el segundo invierno de la IA que duró hasta mediados de 1990: “se produjo un cambio de paradigma en el campo de la IA. Los sistemas expertos que dominaban el campo dieron paso a más métodos basados en el aprendizaje automático” (Schuchmann, 2019). La IA empezó a avanzar desde la misma mirada que generó el primer ciclo de entusiasmo, más tarde de decepción, y quien se reivindicó con eso fue Rosenblatt.

D. Canadá, Hinton y al aprendizaje de máquina

Canadá surgió como uno de los pocos países interesados en desarrollar esta tecnología. Aunque muchos gobiernos, universidades y empresas entraron en el ciclo de entusiasmo, donde hubo emoción y después decepción en torno a la IA, primero a través de los modelos de neuronas encarnadas en el perceptrón y más tarde en los sistemas expertos, la nación del extremo norte de América continuó apoyando esta tecnología. Por lo tanto, Canadá se convirtió en un faro que atrajo el interés de diferentes investigadores que más tarde desarrollaron elementos que favorecieron el avance de la IA, especialmente en el campo del aprendizaje automático, de acuerdo con Kim (2019):

“Canadá fue uno de los primeros inversores en investigación de IA. Durante el llamado invierno de la IA, un período durante las décadas de 1980 y 1990 en el que el menor interés en la IA provocó una caída en la financiación de la investigación, las agencias de financiación canadienses continuaron apoyando la investigación básica de la IA, atrayendo a académicos de todo el mundo” (Kim, 2019).

Mientras tanto, en 1982, Japón anunció su enfoque en desarrollar redes neuronales más avanzadas, lo que motivó a la financiación estadounidense en el área y, por lo tanto, aumentó la investigación en el campo. El gobierno japonés invirtió significativamente en tecnología de redes neuronales y aprendizaje automático. Mientras la IA - ligada a sus apuestas por los sistemas expertos - pasaba por su segundo invierno, entre 1986 y 1993, la Sociedad Japonesa de IA vio un aumento en su membresía (Nishida, 2012).

Es importante mencionar que uno de los talentos atraídos por Canadá fue el Doctor Geoffrey Hinton. El científico británico es considerado uno de los fundadores de la IA profunda junto con Yoshua Bengio y Yann LeCun. Entre sus contribuciones más importantes se encuentra el desarrollo de los modelos de backpropagation, los cuales aumentaron la capacidad de las redes neuronales y, por lo tanto, la capacidad de aprendizaje de las máquinas.

Sin entrar en muchos detalles, se puede ver que la historia de Hinton y, por lo tanto, de Canadá en torno a la IA está relacionada con las situaciones históricas, culturales, políticas y económicas que se dieron en el segundo invierno de la IA. De acuerdo con Smith (2017), el investigador se mostró decepcionado con la falta de financiamiento del gobierno de los Estados Unidos, especialmente del Departamento de Defensa. Esto lo llevó a migrar a Canadá para continuar sus investigaciones en aprendizaje automático, un tema que le ha interesado para comprender cómo los seres humanos aprenden a ver.

Geoffrey Hinton trataba de comprender cómo funcionaba la mente humana y para ello emprendió una larga carrera académica. En una entrevista con Ng (s.f), el doctor Hinton comentó que para tratar de comprender cómo funcionaba la mente humana primero decidió estudiar filosofía, pero no encontró verdades muy claras en la filosofía, así que posteriormente decidió estudiar psicología, pero en la psicología encontró teorías muy simples que no explicaban realmente cómo funcionaba el proceso de la mente humana, en particular el aprendizaje. Más tarde, el investigador se dedicó a la carpintería por un rato; finalmente, el investigador se orientó hacia la IA donde encontró varias de las respuestas que estaba buscando al desarrollar mecanismos para que las máquinas aprendieran; así pues, en palabras de Hinton:

“...comencé a estudiar fisiología y física. Creo que cuando estaba en Cambridge, era el único estudiante universitario en fisiología y física. Y luego me di por vencido y traté de hacer filosofía, porque pensé que podría darme más información. Pero eso me pareció que en realidad carecía de formas de distinguir cuando decían algo falso. Y entonces pasé a la psicología.

Y en psicología tenían teorías muy, muy simples, y me pareció que era algo desesperadamente inadecuado para explicar lo que estaba haciendo el cerebro.

Entonces me tomé un tiempo libre y me convertí en carpintero” (Hinton en Ng y Coursera, s.f.).

El contexto en el que Geoffrey Hinton decidió mudarse a un ambiente favorable como el canadiense, no dejó de ser polémico. El campo del aprendizaje de máquina, tal como se ha expuesto, estaba atravesando un segundo invierno, un segundo periodo de decepción. Sin embargo, el académico británico decidió seguir apostando a esta área en aras de perseguir su curiosidad intelectual: “...cuando tienes lo que crees que es una buena idea y otras personas piensan que es una completa basura, eso es señal de una muy buena idea” (Hinton en Ng y Coursera, s.f.).

Quizás el punto de quiebre en la carrera de Hinton fue el estudio e implementación de diversos algoritmos en torno a la inteligencia artificial. Por ejemplo, la investigación en sistemas de redes neuronales multicapa junto con figuras como LeCun y Bengio (LeCun et al., 2015). Sin embargo, algunos de los hitos más importantes en su trayectoria fueron el empleo de redes neuronales convolucionales y el uso de GPUs de Nvidia en herramientas como ImageNet (Se, 2024). Además, su trabajo sirvió de inspiración para otras técnicas de inteligencia artificial, como los transformadores, en donde su influencia se refleja en artículos académicos tan trascendentales como *Attention Is All You Need* (Vaswani et al., 2017).

Ahora bien, se puede entender todo este contexto utilizando las herramientas teóricas y lógicas propuestas en el primer capítulo. Primero, se tiene una relación constante entre los actores humanos y no humanos, los actores humanos tienen un impulso en cuanto a recursos tecnológicos a partir de los objetos que estos relacionan, impulsados por el entusiasmo generado por estos objetos. Segundo, este impulso se manifiesta en recursos económicos, apoyo gubernamental y una serie de publicaciones y cobertura mediática, generando un escenario febril en torno a estos objetos.

Sin embargo, también se ve cómo estos objetos son influenciados por los actores humanos y otros objetos. Las limitaciones tecnológicas o la falta de la promesa de

gratificación inmediata que estos avances propusieron generan que los actores humanos, especialmente los encargados de instituciones dejen de apoyar el desarrollo de estos mismos objetos, lo cual afecta a otros factores humanos que pierden recursos económicos y esto se traduce en la última parte del ciclo de entusiasmo: la decepción.

Aun así, estos ciclos no fueron lineales y los procesos de innovación tampoco se pueden dibujar como una gráfica ascendente de fondo de forma constante. Tal como se ha planteado, la trayectoria de los sujetos en los procesos de innovación a veces está ligada a ciclos de fracaso, encarnados en una decepción con las tecnologías. Los objetos de contorno no atraviesan por traducciones adecuadas y pueden fracasar o estancarse en los sistemas de innovación.

Las traducciones no fracasan completamente. Si observamos el caso de Japón y el de Canadá existen dos escenarios bastante particulares dentro de lo que está ocurriendo en el segundo invierno de la IA. Por un lado, Japón invirtió una gran cantidad de recursos y logró generar avances quizás no espectaculares, pero sí importantes dentro del aprendizaje de máquina profundo. Por otro lado, mientras el mundo estaba viendo con recelo a la IA, Canadá apostó por invertir en este tipo de tecnología, lo cual más tarde resultó en una serie de tecnologías que le harían exitosa.

Así pues, después de dos inviernos, la IA como objeto de contorno resultó exitosa, pero tuvo que atravesar una serie de traducciones. Y estas traducciones también se unieron a mecanismos para que fueran comprensibles. También hay que señalar que dentro de las constelaciones de objetos que dieron lugar al escenario contemporáneo de la IA, hubo dos periodos de decepción ligados un ciclo de entusiasmo relacionado con subcampos de la IA encarnados en las redes neuronales y los sistemas expertos.

Esto es, porque después de dos inviernos, la IA como objeto de contorno resultó exitosa, pero tuvo que atravesar una serie de traducciones. Y estas traducciones también se unieron a mecanismos para que fueran comprensibles. También es necesario mencionar que dentro de las constelaciones de objetos que dieron lugar

al escenario contemporáneo de la IA, hubo dos periodos de decepción ligados un ciclo de entusiasmo relacionado con subcampos de la IA encarnados en las redes neuronales y los sistemas expertos.

En conclusión, la IA experimentó dos periodos de inviernos, en los que el entusiasmo y la inversión en esta tecnología disminuyeron por limitaciones y decepciones en su desarrollo. El primer invierno de la IA se relacionó con el uso de sistemas expertos y su incapacidad para adaptarse a nuevas situaciones, manejar incertidumbre, escalar eficientemente, aprender de datos, capturar conocimiento tácito y operar con información ambigua, lo que limitó su aplicabilidad y llevó a una disminución del interés y la financiación en el campo. El segundo invierno tuvo que ver con la incapacidad de los modelos neuronales y sus algoritmos para procesar y acceder a una gran cantidad de información y datos. Sin embargo, estos ciclos no fueron lineales y en algunos países como Japón y Canadá, se lograron procesos de traducción exitosos que ayudaron a salvarlos.

E. Watson

Es importante hablar de la plataforma Watson, ya que este objeto y el laboratorio que lo desarrolló han sido fundamentales para demostrar los avances del aprendizaje de máquina profundo. Por ello, este apartado se centra en analizar la trayectoria de Watson para descubrir cómo sus procesos de innovación afectaron las condiciones económicas, tecnológicas e históricas de sus actores.

Watson fue desarrollada en un contexto muy distinto al de Deep Blue, la máquina que usó sistemas expertos para derrotar al campeón mundial de ajedrez, ya que implementa herramientas de aprendizaje de máquina profundo. Esto se debe a que Watson se desarrolló dentro de un escenario donde había dos elementos importantes: una mayor capacidad de poder de procesamiento de computación y una mayor disponibilidad de datos.

Estos elementos permitieron a Watson romper las barreras tradicionales contra la implementación del aprendizaje profundo. Ya no había escasez de datos ni falta de oportunidades para tener una gran cantidad de información, lo que permitió a la

computadora, o en este caso, a la IA, aprender a través de sus redes neuronales. Además, ya había una gran capacidad y poder de procesamiento de datos, lo que le permitía a la máquina acceder y procesar toda esta gran cantidad de información para lograr lo que se pretendía con los modelos teóricos del aprendizaje de máquina: aprender.

Sin embargo, había retos importantes, problemas que debían ser señalados y circunstancias que debían ser superadas. No era lo mismo hacer lo que se hizo con el sistema de Deep Blue para vencer a los campeones de ajedrez a partir de sistemas expertos que contenían todos los posibles movimientos dentro del juego. El reto de Watson estaba en vislumbrar todas las posibilidades, también en acceder a la información, procesarla, aprender de los errores de distintos juegos y ganar.

Esto demostró que no sólo se debía programar una máquina para que siguiera todas las posibles rutas para ganar, ya que esto no bastaba en un juego como Jeopardy. Para ganar, la computadora debía aprender, y para aprender debía superar una gran cantidad de retos técnicos. Así podemos retomar el término objeto de contorno y encontrar el objeto de contorno en común que posibilitó las traducciones entre los distintos miembros del laboratorio que desarrollaron a Watson. Si se sigue la misma ruta, se observa que el objeto de contorno con respecto a la trayectoria de Watson es la IA; Watson no era un ser humano, no podía aprender igual que un ser humano adquiere conocimientos de forma natural, pero sí tenía ventajas en comparación.

El objeto de estudio de la trayectoria de Watson es la IA, entendida como la capacidad de aprender y no solo procesar información o imitar el comportamiento humano, sino de hacer acciones similares a las de los humanos, como propuso Alan Turing. Sin embargo, Watson no era un ser humano y no podía aprender de la misma forma natural que este, aunque sí tenía algunas ventajas. Por ejemplo, los seres humanos pueden tener dificultades para enfrentarse a situaciones completamente nuevas, mientras que los sistemas de IA adaptativos pueden superarlas con éxito (Ertel, 2019). Por otro lado, los seres humanos tienen una

habilidad intuitiva para procesar el lenguaje y las imágenes, que hasta hace poco era inalcanzable para las máquinas.

Pero durante el desarrollo de Watson ya existían teorías y algoritmos en torno al aprendizaje automático; por ello, había un cierto avance —por lo menos de manera hipotética— sobre cómo se podrían solucionar los desafíos en torno al juego de Jeopardy y las limitaciones con respecto al aprendizaje automático y a la IA que se asociaban con estos. Existían modelos de algoritmos para recrear redes neuronales, quizás no iguales a las humanas, pero basadas en la biología de los seres humanos. El problema estaba en cómo hacer que todas esas herramientas y teorías funcionaran, y ahí es cuando surgió Watson.

Watson la desarrolló el equipo de investigación de IBM. De acuerdo con el equipo de investigación: “El *IBM Research* buscaba un desafío de investigación importante que rivalizara con el interés científico y popular de Deep Blue, el campeón de ajedrez informático (Hsu 2002), que también tendría una clara relevancia para los intereses comerciales de IBM”. Para ello, decidieron emprender el reto de desafiar a los concursantes del popular juego de preguntas y respuestas (QA) conocido como Jeopardy.

“Con el control de calidad en mente, nos decidimos por el desafío de construir un sistema informático, llamado Watson, que pudiera competir a nivel de campeón humano en tiempo real en el programa de televisión estadounidense Jeopardy. El alcance del desafío incluye presentar a un concursante automático en tiempo real en el programa, no simplemente un ejercicio de laboratorio” (Ferrucci et al., 2010).

Las motivaciones del equipo eran similares a las del laboratorio que desarrolló *DeepBlue*: demostrar si un sistema podía derrotar a los mejores humanos en una tarea específica. Sin embargo, a diferencia de Deep Blue que utilizaba sistemas expertos anclados a una al poder de cómputo de 1990 (fecha en la que el sistema logró su objetivo de vencer al campeón mundial de ajedrez), el juego de Jeopardy implicaba desafíos más allá de aprender las reglas de un juego. El equipo que desarrolló Watson comentó:

“Jeopardy exige un juego estratégico para igualar el ingenio contra los mejores jugadores humanos. En un juego típico de Jeopardy, Watson se enfrenta a las siguientes decisiones estratégicas: decidir si entrar e intentar responder una

pregunta, seleccionar casillas del tablero y apostar en Dobles Diarios y Jeopardy Final”

Los dobles diarios son una pista oculta al final de cada ronda de Jeopardy, mientras que los Jeopardy Final son las últimas rondas de Jeopardy en las que los participantes deben hacer sus últimas apuestas. Asimismo, los concursantes deben responder una pregunta final de forma escrita en menos de treinta segundos. Este tipo de reglas, más el hecho de que deben responder la mayoría de las preguntas en forma de pregunta, hace de Jeopardy un juego que requiere técnicas más allá de los sistemas expertos para que una máquina gane.

Ferrucci et al. (2010) explican que para enfrentar esos desafíos: "impulsaron la construcción de modelos estadísticos de jugadores y juegos, análisis teóricos de juegos de escenarios y estrategias de juegos particulares, y el desarrollo y aplicación de técnicas de aprendizaje por refuerzo para que Watson aprendiera su estrategia para jugar Jeopardy". El equipo de Watson tenía una ventaja importante respecto a otros investigadores que desarrollaron técnicas de aprendizaje para resolver problemas: más datos. Sin embargo, Ferrucci et al (2010) consideran que, si bien se contaba con una buena cantidad de datos e información, esto no significaba que se dejara a un lado la construcción de modelos para procesar información:

“(se) requiere un modelado extremadamente cuidadoso y una evaluación de la teoría del juego, ya que el juego de Jeopardy tiene información incompleta e incertidumbre que modelar, límites de puntaje crítico que reconocer y jugadores competitivos inteligentes que tener en cuenta” (Ferrucci et al., 2010).

Asimismo, tal como se ha visto desde el trabajo de Star y Griesemer, la elaboración de Watson implicó un esfuerzo más allá del equipo organizado por IBM. Por ejemplo, el equipo colaboró con *Carnegie Mellon University*: “Nosotros también, en cooperación con CMU, comenzamos la iniciativa Avance Abierto de Respuesta a Preguntas (OAQA)” (Ferrucci et al., 2010). Además, el laboratorio encargado de Watson tuvo la visión de emplear una tecnología que fuera más allá de la aplicación de vencer a los mejores jugadores de Jeopardy y después aplicarla en escenarios más allá del juego: “Hemos aplicado con éxito DeepQA a las tareas Jeopardy y TREC QA. Hemos comenzado a adaptarlo a diferentes

negocios, aplicaciones y problemas adicionales de desafío exploratorio que incluyen medicina, búsqueda empresarial y juegos” (Ferrucci et al., 2010).

El desarrollo de Watson representó un momento importante en torno a la IA. Primero, logró establecer el modelo de aprendizaje automático con alianzas con herramientas tecnológicas y equipos de investigación. Segundo, aprovechó las capacidades de cómputo de la época, así como la gran cantidad de datos que disponía, para lograr aprender y generar estrategias en aras de ganar en un juego tan complejo como Jeopardy, que entre otras cosas requería habilidades de lenguaje.

Además, como tercer elemento hubo un aspecto económico. Watson fue apoyado por una gran empresa, IBM. Esto le permitió no sólo crear alianzas con otras universidades, sino también tener la seguridad de que el proyecto podría desarrollarse sin mayores inconvenientes y sin la necesidad de pedir recursos monetarios a instituciones gubernamentales.

Asimismo, IBM también encontró ciertos incentivos al apoyar el proyecto Watson. Por un lado, Watson demostró la enorme capacidad que esta compañía tenía en los procesos de investigación, especialmente en un área de frontera como el aprendizaje automático. Por otro lado, también logró desarrollar un producto, una herramienta con aplicaciones comerciales importantes o con potenciales aplicaciones importantes.

Sin embargo, Watson una vez más evidenció que la trayectoria de los objetos de contorno, esta idea de un proceso de innovación que de manera constante triunfa con base en las correctas traducciones, en ocasiones es ilusoria. De acuerdo con Hernández y Fitch: “*Watson Health* y algunos otros proyectos de IA de la industria tecnológica que han tenido problemas fueron demasiado ambiciosos, y trataron de responder preguntas amplias y complicadas relacionadas con la salud, dijeron los expertos” (Hernández & Fitch, 2021).

Muchas de las aplicaciones con las que se contemplaba Watson a corto plazo como la medicina, el diagnóstico de enfermedades, entre otras, fueron complicadas de implementar en la vida real. Bakator y Radosav (2018 p. 53),

mencionan que si bien hasta el año 2018 los sistemas de aprendizaje profundo solamente han sido una herramienta que puede ayudar los doctores, en un futuro existe la posibilidad de que tenga aplicaciones más prometedoras: en especial en área del diagnóstico médico. El sistema de aprendizaje profundo en la medicina representa una promesa, pero aún no reemplaza a profesionales médicos en esa área, aunque solo puede ser cuestión de tiempo y probablemente sea bajo el modelo colaborativo humano-máquina.

En síntesis, la trayectoria de Watson implicó la convergencia de muchas tecnologías que permitieron aplicaciones complejas del aprendizaje de máquina. Aun así, también se demostró de nueva cuenta este ciclo complejo de entusiasmo con respecto a las innovaciones tecnológicas como IA. Más allá de la decepción encarnada en las expectativas que Watson no cumplió, hay que resaltar los avances y elementos que luego servirían como los cimientos para desarrollar e implementar la IA en muchas aplicaciones cotidianas.

F. AlphaGo

AlphaGo recibió un gran apoyo de empresas privadas. En este caso el proyecto fue impulsado por Alphabet Inc, la empresa matriz de Google, como uno de los tantos programas de investigación y desarrollo que esta tenía. Esto remite a un escenario completamente distinto al que se enfrentaron los primeros investigadores y teóricos de la IA como Ada Lovelace y Adam Turing que buscaban apoyos gubernamentales o en círculos sociales, mientras que AlphaGo fue respaldada por una inversión millonaria que comenzó a ver al aprendizaje profundo como algo monetizable.

Pero el paralelo principal con los anteriores hitos en torno a la IA tuvo que ver con las traducciones que hicieron las personas que formaron parte del equipo de AlphaGo. Si se considera que el laboratorio de *DeepMind* (quienes crearon a AlphaGo) fue un equipo conformado con la intención de crear un mecanismo una plataforma de aprendizaje de máquina, entonces se puede asumir que no hubo tantos retos nuevos ni tantos enfrentamientos como en otras ocasiones; sin

embargo, esto no fue así. Sí, había una meta fija clara definida en torno a qué tipo de máquina se debe construir, pero esto no fue tan fácil.

De acuerdo con Steffen Hölldobler, Sibylle Möhle y Anna Tigunova (2017), AlphaGo se enfrentó a diferentes retos porque tenía que hacer la convergencia distintos tipos de tecnologías y herramientas. Esta variedad de tecnologías y herramientas iban desde modelos de aprendizaje de máquina, hasta herramientas de análisis de juegos como el Modelo Montecarlo, una técnica de simulación y optimización que utiliza muestreo aleatorio para estimar resultados numéricos. Se basa en la generación de múltiples simulaciones o escenarios posibles para aproximar soluciones a problemas complejos, especialmente aquellos que involucran incertidumbre o grandes espacios de búsqueda. Entonces hubo la necesidad de conjugar todo este tipo de sistemas para encontrar una solución práctica quisiera de la plataforma AlphaGo algo factible.

Además, las apuestas teóricas y los modelos de laboratorio se tenían que probar en la realidad, en un escenario donde las personas participaran de su construcción. Por eso el laboratorio que creó AlphaGo lo puso a prueba dentro de Internet y frente a varios participantes humanos. Esto llevó a pensar que un jugador misterioso de AlphaGo que estaba rompiendo las expectativas se volviera sumamente famoso cuando en realidad se trataba de AlphaGo.

“En diciembre de 2016, la comunidad de Go quedó desconcertada por un jugador desconocido Maestro en el servidor en línea de Go, que a finales de mes logró ganar a más de 30 profesionales de primer nivel sin perder. En enero de 2017 DeepMind admitió que era una nueva versión de AlphaGo” (Steffen Hölldobler, Sibylle Möhle y Anna Tigunova, 2017).

AlphaGo también estaba asociada con uno de los grandes desafíos a los que se enfrentaba la IA. Si bien buscaba vencer a un campeón a un humano en un juego con reglas más o menos definidas, a diferencia del ajedrez –donde DeepBlue resultó victoriosa con su modelo de expertos–, AlphaGo competía en un escenario mucho más demandante. Steffen Hölldobler, Sibylle Möhle y Anna Tigunova (2017) explican que:

“El problema con Go es el tamaño de la placa, que produce un espacio de estado de 10^{170} posiciones. En comparación, el espacio de estados de Ajedrez es aproximadamente 10^{43} . Se sabe que estos juegos tienen un factor de ramificación alto: el número de movimientos disponibles desde la posición actual. El número de posibles escenarios de juego en Go es mayor que el número de átomos en el universo” (Steffen Hölldobler, Sibylle Möhle y Anna Tigunova, 2017).

Steffen Hölldobler, Sibylle Möhle y Anna Tigunova (2017), *en su artículo Lessons learned from AlphaGo*, explican que “...Go es un juego de información perfecta. El resultado del juego depende solo de la estrategia de ambos jugadores, como se explicó antes. Sin embargo, a diferencia de otros juegos con reglas establecidas como las Damas chinas o el ajedrez, Go encierra un gran número de posibilidades que hacen complicado la implementación de sistemas expertos donde se programen todas las reglas del juego, así como sus posibilidades.

“Esto hace que sea atractivo resolver Go computacionalmente porque podemos confiar en una máquina para encontrar la secuencia óptima de movimientos. Sin embargo, esta tarea es extremadamente difícil debido al enorme espacio de búsqueda de posibles movimientos. Por lo tanto, Go se ha considerado una frontera deseada para la IA, que se predijo que no sería alcanzable en la próxima década” (Steffen Hölldobler, Sibylle Möhle y Anna Tigunova, 2017).

Asimismo, la naturaleza del AlphaGo hace que sea complicado usar únicamente una estrategia de aprendizaje automatizado. Por consiguiente, el equipo que desarrolló AlphaGo se apoyó en otras herramientas para lograr obtener los elementos que le permitieran ganar en el juego. El equipo que desarrolló AlphaGo explica: “Combinando la búsqueda de árboles con políticas y redes de valores, AlphaGo finalmente ha alcanzado un nivel profesional en Go, brindando la esperanza de que el desempeño a nivel humano ahora se pueda lograr en otros dominios de IA aparentemente intratables” (Silver et al., 2016).

De esta manera se tienen dos elementos importantes. Primero se tuvieron que hacer traducciones con tal de alcanzar un propósito: vencer al campeón mundial de AlphaGo. Segundo, el equipo de AlphaGo tuvo que buscar herramientas para hacer una segunda traducción, una traducción que pusiera prueba todos los avances y todos los insumos tecnológicos en aras de justificar la existencia AlphaGo, y para que AlphaGo pudiera aprender de otros jugadores –así como de

sí mismo—, y después poner a prueba sus herramientas de aprendizaje de máquina junto con el *Modelo Montecarlo* en una situación real; así pues, de acuerdo con el equipo que desarrolló AlphaGo:

“Una de las formas de abordar de manera eficiente el árbol de búsqueda del juego es aplicar el método Monte Carlo. Este método aproxima una función con el uso de muestreo aleatorio. La media de las muestras converge al valor real de una función. Monte Carlo se usa comúnmente cuando el problema es demasiado grande para ser resuelto analíticamente. Se ha aplicado ampliamente en matemáticas, estadística y otras disciplinas” (Steffen Hölldobler, Sibylle Möhle y Anna Tigunova, 2017).

Hay que señalar qué tipo de herramientas de aprendizaje de máquina empleó AlphaGo. A diferencia de modelos tradicionales de aprendizaje de máquina AlphaGo utilizó un modelo de reconocimiento de imágenes lo cual le permitió tener una perspectiva física real de todas las variantes del juego tal como un jugador humano lo puede hacer. Asimismo, empleó diferentes capas y técnicas de aprendizaje de máquinas profundo para aprender y a partir de este aprendizaje genera un resultado para ganarle.

AlphaGo también empezó a practicar contra sí mismo creando una especie de espejo. Esto es, para jugar varias veces contra sí misma, aprendiendo todas las variables de juegos que podrían existir y desarrollar una especie de intuición similar humana para luego usarla para ganar. Así pues, para mejorar el Modelo Montecarlo “AlphaGo utiliza redes neuronales convolucionales profundas y aprendizaje por refuerzo” (Steffen Hölldobler, Sibylle Möhle y Anna Tigunova, 2017).

La necesidad de emplear dos herramientas tiene que ver con la forma en los seres humanos participan en juego. Como se mencionó anteriormente, si bien las personas eligen un movimiento mientras intentan predecir el movimiento de sus adversarios, también hacen uso de su intención (Steffen Hölldobler, Sibylle Möhle y Anna Tigunova, 2017). Esto hace que se requiera de una herramienta como las redes neuronales que permitan la retroalimentación y con ello el aprendizaje por refuerzo.

Steffen Hölldobler, Sibylle Möhle y Anna Tigunova (2017) explican que en las Redes Neuronales Artificiales (ANN): “las neuronas están dispuestas en capas, las conexiones existen solo entre las capas consecuentes y no hay bucles en la red. Consiste en una sola entrada y capa de salida y varias capas ocultas”. Y de forma más elaborada explican la noción de otro tipo de redes neuronales artificiales, otra herramienta empleada en Alphago, de la siguiente manera:

“Una red neuronal convolucional se basa en dos conceptos: campos receptivos locales y reparto de peso. El primer concepto significa que cada neurona está conectada solo a una pequeña región de un tamaño fijo en la capa anterior. Cada neurona es responsable de su propia subárea en la capa anterior, pero el conjunto de pesos de conexiones, conocido como filtro, es el mismo para todas las neuronas (supuesto de reparto de peso)”.

Steffen Hölldobler, Sibylle Möhle y Anna Tigunova (2017) explican que: “Conceptualmente, el aprendizaje por refuerzo es un método para entrenar a un agente de IA al no darle explícitamente la respuesta correcta. En cambio, el objetivo del agente es maximizar la recompensa, que es una retroalimentación del entorno simbólico a sus acciones”. En el capítulo anterior ya se explicaron estos sistemas y estrategias tecnológicas. Pero es necesario conocer la perspectiva de quienes las emplean para desarrollar objetos, actores no humanos, que se pueden considerar como hitos en sus campos.

AlphaGo es un proyecto de IA concebido por DeepMind, un grupo de investigadores que pertenece a Google de Alphabet Inc, y que ha contado con el apoyo económico y tecnológico de otras empresas como Apple, Microsoft y Twitter, que han adquirido startups relacionadas con el reconocimiento de voz, la predicción de texto y el procesamiento de datos visuales (Royal Society, 2017 en BBC News, 2016). El objetivo de AlphaGo era vencer a un campeón humano en el juego de Go, un reto que se consideraba muy lejano debido a la complejidad y las posibilidades del juego. Sin embargo, AlphaGo logró alcanzar un nivel profesional en Go, combinando la búsqueda de árboles con políticas y redes de valores, y demostrando así el potencial de la IA para resolver problemas aparentemente intratables (Silver et al., 2016).

Finalmente, después de muchos años de desarrollo, el laboratorio de DeepMind logró que AlphaGo ganase “cuatro de cinco partidos contra Lee Sedol, quien ha sido el mejor jugador de Go del mundo durante más de una década” (Royal Society, 2017). Ese hecho representó un momento cumbre en el área de la IA. Asimismo, esto fue el resultado de una serie de alianzas y traducciones, de intereses que fueron atravesados por un lenguaje común que al final lograron avanzar al campo de la IA: “Combinando la búsqueda de árboles con políticas y redes de valores, AlphaGo finalmente ha alcanzado un nivel profesional en Go, brindando la esperanza de que el desempeño a nivel humano ahora se pueda lograr en otros dominios de IA aparentemente intratables” (Silver et al., 2016⁴).

Para concluir, AlphaGo es un hito en la historia de la IA debido a la convergencia de varias tecnologías y procesos de innovación. Es un triunfo en el campo de la IA que ha superado el estigma de los "dos inviernos" de la IA. Aunque pueda haber un exceso de entusiasmo alrededor de AlphaGo, sus aportaciones son importantes ya que representan la materialización de distintos esfuerzos y su aplicación en diferentes plataformas y aplicaciones.

4. Del laboratorio al mundo cotidiano: IA en redes sociales y economía gig

El uso de plataformas sociales y laborales sigue creciendo significativamente en México y Japón. En el caso de las redes sociales en México, datos recopilados por Statista y fuentes como Digital Portal, basadas en información gubernamental y no gubernamental, revelan tendencias claras. Facebook se mantiene como la red social más utilizada en 2023, con casi el 93% de los mexicanos empleando esta plataforma, seguido de WhatsApp con un 92.2%, e Instagram con más del 80%, mientras que TikTok ocupa el cuarto lugar (Statista, 2023a; DataReportal, 2024).

Por otro lado, en Japón, su población también utiliza activamente redes sociales, aunque las preferencias son diferentes. La aplicación de mensajería LINE lidera como la plataforma más empleada con 78.1% de los usuarios, seguida por YouTube, Instagram y X (anteriormente conocida como Twitter). Completan el

⁴ Son el equipo que desarrolló AlphaGo.

listado Facebook y TikTok, destacando cómo las plataformas reflejan hábitos culturales específicos (Humble Bunny, 2025; Statista, 2023b).

En cuanto a las plataformas digitales de trabajo, en México operan Uber, Mercado Libre, Rappi, Cabify, Didi, Jokr, InDrive, Didi Food y Uber Eats (Lagos, 2024)) entre otras. Mientras que, en Japón, las plataformas de la economía gig incluyen opciones como Lancers y CrowdWorks para freelancing en áreas como diseño y programación, mientras que servicios como Uber Eats, DiDi Food y Yamato Transport ofrecen oportunidades en entrega de alimentos y logística; además, plataformas como Coconala permiten a los usuarios realizar tareas específicas o brindar servicios especializados. Hay que resaltar que, si bien el servicio de taxi bajo demanda de Uber opera en Japón, este es básicamente manejado por los gremios y empresas de taxistas del país.

Ahora bien, cada una de estas plataformas utiliza modelos específicos de inteligencia artificial para optimizar sus operaciones y servicios. Facebook, propiedad de Meta, es reconocida por emplear algoritmos de aprendizaje de máquina para potenciar su publicidad personalizada (Horwitz, 2023). Además, ha integrado herramientas de IA generativa, como LLAMA, en sus ecosistemas de WhatsApp e Instagram, con el objetivo de mejorar la personalización del contenido, optimizar la experiencia del usuario y ofrecer respuestas automatizadas más fluidas y naturales (Meta, 2023).

Dentro del uso de sistemas de inteligencia artificial en plataformas digitales de trabajo, el caso de Uber destaca como un ejemplo clave que también servirá como marco de referencia más adelante. En su sistema de taxis bajo demanda, Uber utiliza Michelangelo, una plataforma de aprendizaje automático lanzada en 2016 que permite crear, implementar y gestionar modelos a gran escala (Wang et al., 2024). Este sistema integra componentes esenciales del aprendizaje automático (ML), como modelos, datos y funciones de pérdida, empleando técnicas como el aprendizaje supervisado, no supervisado y por refuerzo.

Con el tiempo, Uber ha avanzado hacia el uso de redes neuronales profundas (Purwono et al., 2023) y modelos generativos de lenguaje (LLMs) para mejorar

tanto la experiencia del usuario como la productividad interna (Wang et al., 2024). Desde la perspectiva de la Teoría del Actor-Red (ANT), estos sistemas actúan como actores no humanos que conectan conductores y clientes, facilitando pagos, ofreciendo recomendaciones y proporcionando soporte mediante chatbots (Morton, 2024). Además, reflejan dinámicas de poder y pueden presentar fallas sistémicas que impactan directamente a los usuarios (Wang et al., 2024).

Lo anterior implica que muchos modelos de inteligencia artificial ya salieron del laboratorio para integrarse a la vida cotidiana y con ello tiene implicaciones importantes. Ahora bien, a través del análisis de constelaciones, se ha encontrado que la IA está relacionada con el desarrollo de las primeras computadoras, los perceptrones y las redes neuronales, y con la creación de algoritmos que permiten el aprendizaje automático. Es importante mencionar que la trayectoria de la IA también estuvo ligada a un conjunto de alianzas cuyo resultado fue el inicio de la disciplina.

Después de hablar de la historia de la IA, sus actores importantes y sus proyectos más icónicos como Watson y AlphaGo, es importante mencionar algunos de los usos más cotidianos que tiene. La IA ya forma parte de la vida cotidiana de las personas, a menudo sin que estas sepan, y esto demuestra su efectividad al superar lo que se conoce como el "valle de la incertidumbre".

Así pues, utilizando el método de constelaciones, este apartado se centrará en dos casos empíricos importantes: las plataformas sociales (o redes sociales) y la economía gig. Ambos casos son ejemplos de aplicaciones de la IA y de sus herramientas, como el aprendizaje automático, para que estas constelaciones de plataformas y aplicaciones funcionen.

En este capítulo se puede apreciar que existen dos perspectivas en torno a la IA. Por un lado, están los entusiastas, quienes creen que la IA generará avances significativos en poco tiempo. Por otro lado, están los pesimistas, quienes creen que la IA no es realmente inteligente, sino solo un conjunto de herramientas técnicas que simulan la inteligencia humana o que intentan igualarla. Ninguno de estos dos grupos tiene la verdad absoluta sobre lo que es la IA.

Desde las ciencias sociales y humanidades, se puede observar que la realidad de la IA es más simple, especialmente con los avances que se están implementando en la vida cotidiana. Aunque no se alcanzó la IA general, una máquina capaz de pensar con capacidades similares o superiores a las humanas, la tecnología ha tenido avances trascendentes en lo que se conoce como IA acotada, especializada en realizar una sola tarea. Esto se ve en herramientas especializadas como redes neuronales de lenguaje que pueden tomar dictado o predecir intenciones.

Así pues, se tiene una constelación de herramientas que funcionan en torno a la IA. Por ejemplo, las redes neuronales del lenguaje especializadas son eficientes en la traducción, sugerencia de texto, producción de algunos materiales originales, y en reproducir patrones de habla; mientras que las redes neuronales profundas son excelentes para el reconocimiento de imágenes. Cada herramienta parte de principios y avances desarrollados a lo largo de la trayectoria de la IA, en especial la parte relacionada con el aprendizaje de máquina.

Dentro de estas herramientas generan aplicaciones específicas de la IA existen aquellas que ya se están implementando en plataformas que se usan todos los días. Dentro de las plataformas con uso extensivo están las llamadas redes sociales. De acuerdo con *17° Estudio sobre los Hábitos de los Usuarios de Internet en México 2021* realizado por la Asociación de Internet MX (AIMx) en colaboración con *The Competitive Intelligence Unit* (The CIU) “La contabilidad de usuarios de internet en México ascendió a 86.8 millones en 2020, esta cifra es equivalente a la adición de 8 millones de internautas en un año y un coeficiente de adopción de 76.3% entre los mexicanos en edades de 6 años o más” (Piedras, 2021).

De acuerdo con el estudio realizado por *Hootsuite* y *We are social*, el “77.2% de la población nacional (de México) accede a las redes sociales” (Ramos, 2021). Así pues, se puede apreciar que poco más de tres terceras partes de la población mexicana está conectada, y casi el mismo porcentaje usa redes sociales. Estas redes sociales, como Twitter, LinkedIn, e Instagram, funcionan usando muchas

herramientas de la IA que se fueron desarrollando con tecnología surgida de los casos empíricos que aquí se examinaron.

Es importante destacar que muchas veces las personas no se dan cuenta de que las plataformas que utilizan en su vida cotidiana funcionan mediante el uso de IA. Esta IA se puede utilizar para diferentes propósitos, como, por ejemplo, sugerir noticias, recomendar amigos a seguir o mostrar videos que les gustarán en plataformas sociales como TikTok. Estas plataformas sociales utilizan diferentes herramientas de IA, junto con herramientas especializadas en una sola tarea, para brindar un servicio eficiente en particular.

Para comprender la manera en la que se articulan las diferentes herramientas de la IA. De acuerdo con Van Dijck et, al (2018):

“Una *plataforma* en línea es una arquitectura digital programable diseñada para organizar las interacciones entre los usuarios, no solo los usuarios finales, sino también las entidades corporativas y los organismos públicos. Está orientada a la recopilación sistemática, el procesamiento algorítmico, la circulación y la monetización de los datos de los usuarios”.

Asimismo, las plataformas están compuestas de distintos elementos que permiten su funcionamiento. De forma similar a los videojuegos, una plataforma se compone de los elementos generales que constituyen su forma (diseño), sus reglas y funciones relacionadas con los límites tecnológicos (arquitectura), la forma en que las personas interactuarán con ella (interfaz de usuario) y las herramientas para interactuar con ella (diegéticas, dentro de la plataforma; y extradiegéticas, fuera de la plataforma, como el ratón, el teclado y la pantalla táctil). Cada uno de estos elementos trabaja en conjunto para permitir que las personas interactúen con la plataforma y para que las empresas obtengan beneficios con ella.

En este sentido, Bailey et al. (2005) comentan que el diseño de la plataforma es “un enfoque de diseño orientado a la integración que enfatiza la reutilización sistemática para desarrollar productos complejos basados en plataformas y componentes virtuales de hardware y software compatibles, con el objetivo de reducir los riesgos de desarrollo, los costos y el tiempo de comercialización”. El

diseño es la etapa previa en la que se planifican los elementos que tendrá la plataforma según los existentes en su contexto. Otra forma de ver el diseño de las plataformas es considerando algunos aspectos del diseño de los videojuegos, como los relacionados con la experiencia del usuario (Morton, 2021).

Otro elemento de las plataformas con el que la mayoría de las personas interactúa con la IA es la arquitectura. La arquitectura se puede describir como la forma en que se aplican las reglas en torno a una aplicación en función de las limitaciones del hardware o dispositivos en los que se usa. También se puede ver la arquitectura de la plataforma como modelos de sistemas complejos en los que algunos componentes, como la plataforma en sí, se mantienen estables, mientras que otros, como los complementos, pueden cambiar a lo largo del tiempo o en distintos contextos (Baldwin & Woodard, 2008).

Luego se tiene la interfaz de usuario. Estos son los elementos audiovisuales y alguna retroalimentación táctil que permiten al usuario, empresas y desarrolladores interactuar con la plataforma. También se puede definir la interfaz de usuario como:

Se puede definir como diegética o la actividad que hace el avatar o actante que controla el video jugador dentro del escenario del videojuego; mientras que la experiencia extradiegética es la actividad o las acciones que hace el jugador a través de los dispositivos de interacción con los que cuenta en el mundo real (Wolf y Perron 2019 en Morton, 2021).

Así pues, la IA se puede desarrollar dentro de la plataforma. Ya sea mediante la normalización y ejecución de códigos y reglas, la extracción e interpretación de datos, la creación de nuevos contenidos o la recomendación de nuevos contenidos para los usuarios y creadores. Esto se logra mediante una serie de códigos heterogéneos, que son el "conjunto de convenciones que gobiernan las comunicaciones reguladas entre personas, donde el equipo disponible requiere tal tratamiento de mensajes" (Thevenot, 1984), permitiendo que la IA, especialmente la que se aproxima al concepto de inteligencia general de Turing, tenga la capacidad de aprender y forme parte de la vida de las personas a través de sus pequeñas pantallas de luz azul que miran constantemente en sus manos.

Finalmente, es fundamental destacar que tanto las plataformas sociales como las de trabajo enfrentan tres grandes desafíos relacionados con los derechos humanos: privacidad, acceso a la información y no discriminación. Por ejemplo, Facebook ha enfrentado numerosos problemas vinculados a sus modelos de inteligencia artificial, especialmente en lo que respecta a la recolección masiva de datos personales, la forma en que comparte esta información y su uso para publicidad dirigida. Además, Meta, como empresa matriz, integra y comparte tanto sistemas de inteligencia artificial como datos de los usuarios a través de su ecosistema de plataformas, como Facebook, Instagram y WhatsApp (My Privacy Blog, 2024), las cuales son las más utilizadas en países como México. Estos mismos desafíos pueden replicarse en otras plataformas que emplean sistemas de IA, como Uber, aunque Meta se distingue por su enfoque particularmente agresivo en la recolección y explotación de datos personales (Bareebe, 2022).

A. TikTok

En este apartado se estudiarán dos plataformas sociales: Facebook y TikTok. Facebook es una plataforma que forma parte de una constelación de plataformas, que incluye Instagram y WhatsApp. TikTok es una plataforma desarrollada en China, que utiliza la IA para sugerir vídeos a los usuarios.

La relevancia de TikTok se da en dos niveles, el primero tiene que ver con ser una plataforma de entrenamiento y social que compite con los gigantes norteamericanos Facebook Inc y Twitter; el segundo se refleja en su cantidad de usuarios. Esto es, porque “...la aplicación se ha descargado más de mil millones de veces hasta ahora, con una presencia global que incluye India, Estados Unidos, Japón, Corea del Sur, naciones europeas, Brasil y gran parte del sudeste asiático. Y precisamente, durante el primer trimestre de 2019, fue la tercera aplicación más descargada del mundo después de *WhatsApp* y *Messenger*” (The Business School, 2019).

Según Galer (2020), la historia de TikTok no comenzó como un cuento de hadas. TikTok no surgió de la nada, sino que surgió a raíz de la convergencia de tres aplicaciones distintas. La primera de ellas fue *Musical.ly*, que se lanzó en

Shanghai en 2014 pero tenía fuertes vínculos comerciales en Estados Unidos y una audiencia sólida en ese mercado clave; la segunda aplicación fue Douyin, desarrollada por el gigante chino ByteDance. En 2018, ByteDance compró Musical.ly y las fusionó para crear una aplicación con las mejores características de ambas. En esta primera etapa de TikTok, el motivo principal para desarrollar la plataforma fue el incentivo comercial: Zhang Yiming declaró que "China es el hogar de sólo una quinta parte de los usuarios de Internet en todo el mundo. Si no nos expandimos a una escala global, estamos destinados a perder frente a nuestros pares que miran las cuatro quintas partes. Por lo tanto, globalizarnos es imprescindible" (Wang et al, 2020).

TikTok sigue operando bajo una aplicación, pero con dos nombres diferentes. La fusión de ByteDance Musical.ly se hizo con la intención de incorporar la función musical (hacer canto de imitación en videos de 15 segundos), con el resto de las aplicaciones de *Douyin*. Así pues, ambas plataformas integraron sus funciones, pero *Musical.ly* cambió su nombre a TikTok en el mercado internacional, mientras que Douyin conservó su nombre en el mercado chino.

Más tarde, la empresa comenzaría a invertir de forma agresivamente para desarrollar y crear un laboratorio que ayudaría a perfeccionar el corazón de su aplicación: la IA. Esto es, porque la IA, y en especial el aprendizaje máquina, es la base para el funcionamiento de casi todos elementos de la plataforma: desde su diseño, hasta la interfaz y experiencia de usuario. De acuerdo con The Bussiness School (2019), TikTok utiliza la IA en tres niveles: el primero tiene que ver con consumidor de contenido, al cual se le muestran videos con base en lo que le gusta o le puede llegar a gustar; el segundo, tiene que ver con los creadores de contenido, a los cuáles se les sugiere que tipos de videos hacer, a qué tipo de público deben de ser dirigidos, y cómo mejorarlos.

El tercer nivel, tiene que ver con los anunciantes, cuyo contenido se ancla a la audiencia adecuada. Además de que sí son creadores de contenido, se les sugiere cómo hacer mejores videos para que estos se vuelvan virales o altamente vistos como fue el caso de una empresa creadores de osos de gomitas, cuyo

video se volvió viral (Galer, 2020). Así, la IA no solo tiene función auxiliar para el funcionamiento de la plataforma, sino central de la misma.

En el caso de la experiencia de usuario, Zhao (2021) explica que “el contenido que ven los usuarios básicamente satisface sus necesidades e intereses, y no necesita gastar el costo para encontrarlos. Como resultado, esto puede llevar a los usuarios a ver videos cortos durante horas e incluso convertirse en una adicción”. Para sugerir contenido a tal grado que pueda generar adicción, la plataforma genera perfiles de usuarios a partir de datos o metadatos. Los datos son la información que el usuario pone en su perfil como qué tipos de videos le gustan, su edad, temas de interés, país de origen, entre otros; por otro lado, los metadatos son datos de datos y no se pueden cuantificar como los videos de que el usuario aprende.

En aras de lograr lo anterior, *ByteDance*, la empresa dueña de TikTok, patentó un algoritmo que emplea el método y dispositivo de recomendación de descubrimiento de interés global (Zhao, 2021). Zhao explica que este algoritmo funciona a partir de un árbol de jerarquías basado en los temas de interés y la cercanía de los temas de interés. Por ejemplo, si a una persona le interesa la ética, y la IA, entonces el algoritmo le sugerirá videos relacionados con ética en IA, ética en aprendizaje de máquina, cómo hacer una tesis sobre ética en IA, metodología de la investigación, y así sucesivamente.

“Los metas clasificadores en la primera capa a continuación son una categoría general como Deportes, Tecnología, Entretenimiento, etc. Supongamos que hay un nodo principal (es decir, un nodo que tiene una rama) que representa deportes, entonces sus nodos secundarios significan clasificaciones más detalladas en el campo. de deportes, como baloncesto, fútbol, natación. Cada rama se puede refinar aún más en un dominio más estrecho y específico. Douyin también anima a los creadores a especializarse en un solo dominio, en lugar de buscar una amplia cobertura del contenido. En cuanto al algoritmo de recomendación, dicho contenido puede recomendarse con mayor precisión a los usuarios que lo necesiten y aumentar la audiencia válida” (Zhao, 2021).

Hay que tomar en cuenta cómo funcionan los mecanismos para perfeccionar el perfil, y las sugerencias de usuario a partir de su actividad en la plataforma. Para ello, TikTok emplea una serie de algoritmos de IA para obtener mejor rendimiento.

Ahora la IA está en su periodo acotado, cada herramienta es buena haciendo una tarea específica, pero es importante que una plataforma emplee algoritmos para mejorar su tarea.

Ahora bien, en cuanto al uso del aprendizaje de máquina en TikTok, Zhao (2021) explica:

“La clave de todos los algoritmos de aprendizaje automático es que puede aprender interacciones de funciones sofisticadas detrás de los comportamientos de los usuarios. La mayoría de las interacciones de funciones están ocultas en los datos y es difícil identificar una regla, (por ejemplo, la regla de asociación clásica *pañal y cerveza* se extrae de los datos, en lugar de ser descubiertos por expertos), que solo se pueden capturar automáticamente mediante el aprendizaje automático”.

La función de recomendación y creación de perfil de usuario en una plataforma no se da por terminada una vez que el algoritmo aprende las preferencias del usuario. Hay un proceso de aprendizaje continuo en tiempo real después de que el algoritmo aprende las preferencias del usuario. Esto se ajusta a la visión de Turing y Rosenthal de que una máquina inteligente, al igual que los seres humanos, aprende de sus errores y experiencias, alejándose de las reglas fijas programadas en códigos rígidos.

Para que la IA pueda aprender los hábitos de los usuarios y cómo reaccionan a los videos y publicidad dirigida, el algoritmo debe pasar por dos etapas: entrenamiento y ejecución. Como en la mayoría de las herramientas de aprendizaje automático, incluyendo Watson, los instrumentos de aprendizaje automático en TikTok se entrenan con datos de prueba. Una vez que la máquina muestra resultados prometedores, como identificar que a alguien le pueden gustar los bailes de banda después de que le gustaron los de música gruper, las herramientas de IA pasan a la fase de aplicación, formando parte de la experiencia del usuario, el diseño y la arquitectura de la plataforma.

Sin embargo, el proceso de entrenamiento de las herramientas de aprendizaje automático no termina una vez que se integran en la plataforma. Al igual que no puedes aprender más sobre una persona después de leer su perfil en una plataforma social si no interactúas diariamente con ella, las herramientas de

aprendizaje automático mantienen la capacidad de aprender mientras tienen más interacciones con el usuario. Esto, junto con las técnicas de TikTok para dividir a sus usuarios en grupos identificables, hace que la IA sea eficiente y no necesite de complejos árboles de relaciones entre familiares y amigos para centrarse en los gustos y grupos de personas con gusto similar.

Como señala Zhao (2021), los algoritmos de aprendizaje automático no solo permiten a la plataforma aprender y conocer al usuario, sino también permiten al usuario conocerse mejor a sí mismo. Pueden descubrir cosas que no sabían que les gustaban, pero que estaban ocultas y que salen a la luz: “Por lo tanto, el algoritmo de recomendación no solo recomendará al usuario el contenido y el tema que le interesa o preocupa exactamente, sino también algo nuevo e inesperado” (Zhao, 2021). Esto se logra mediante las herramientas y técnicas de reconocimiento de gustos mencionadas anteriormente, que coordinan los gustos e intereses de las personas con grupos que pueden compartir intereses similares, lo que permite a la plataforma sugerir videos que puedan interesar tanto al individuo como a los miembros del grupo con los que se relaciona.

Finalmente, es importante destacar el aspecto tecno histórico en el que surgió TikTok. Como todas las plataformas que usan aprendizaje automático, necesita muchos datos y poder de procesamiento. Esto se ha vuelto posible gracias a Internet, la tecnología en la nube y la gran cantidad de usuarios que tiene y que la utilizan a través de la red. . Por lo tanto, la IA es el resultado de una convergencia de áreas de estudio de la ciencia de la computación que crearon varias aplicaciones que pasaron por diferentes etapas, con traducciones fallidas, pero que finalmente dieron sentido a la trayectoria de este objeto en constante evolución conocido como IA y su capacidad de aprendizaje.

B. Facebook

El caso de Facebook es interesante, ya que ha sido una plataforma en el mercado durante aproximadamente 12 años, un período relativamente largo en el mundo de las plataformas en línea. Facebook comenzó en 2003 a través de una plataforma conocida como FaceMash, en la que "los estudiantes de Harvard que se

inscribieron en el servicio podían publicar fotografías de sí mismos e información personal sobre sus vidas, como los horarios de clases y los clubes a los que pertenecían" (Facebook | Overview, History, & Facts, 2021). La plataforma, que en un principio se enfocó en la demografía estudiantil, creció después con la incorporación de elementos, actores humanos y tecnologías, en una especie de objetos cualitativos.

Facebook nació cuando internet pasaba al modelo 2.0. Es decir, el internet se estaba convirtiendo en medio multimedia capaz de reproducir formas rudimentarias de video, audio e imágenes. Asimismo, Facebook surgió cuatro años después de que reventó la burbuja.com a finales de 1999, un fenómeno caracterizado por la inversión en compañías de internet que prometían ser el próximo éxito pero que carecían de valor intrínseco.

Además, dentro del contexto en que surgió Facebook, la tasa de usuarios de internet en los Estados Unidos comenzó a aumentar considerablemente. Esto, junto con la cada vez mayor implementación de tecnologías de internet de alta velocidad que dejaron atrás las antiguas descargas de los modem de 56 kilobits, por descargas y cargas de datos que superan varios megas de velocidad. Acto seguido, la mayor velocidad de datos que se movían, cargaban, y descargaban en internet, así como nueva tecnología de codificación, encriptación y empaquetamiento de datos, permitió que los usuarios pudiesen producir y consumir más información y contenido.

Además, en el contexto en el que surgió Facebook, la tasa de usuarios de Internet en los Estados Unidos comenzó a aumentar significativamente. Esto, junto con la creciente implementación de tecnologías de Internet de alta velocidad que dejaron atrás las antiguas descargas de los módems de 56 kilobits, permitió descargas y cargas de datos que superan varios megas de velocidad. Esta mayor velocidad de datos que se movían cargando y descargando en Internet y las nuevas tecnologías de codificación, encriptación y empaquetamiento de datos permitieron que los usuarios pudieran producir y consumir más información y contenido.

En 2004, Mark Zuckerberg comenzó a escribir un código para una plataforma que expandiera las capacidades superficiales de FaceMash. Esta plataforma se conocería como Facebook, y al principio, su uso estaba restringido a un grupo selecto de universidades. Sin embargo, Zuckerberg, junto con un grupo de socios y colaboradores, expandió el uso de la ahora conocida como Facebook, a todos los usuarios mayores de trece años en 2006 (CBS News, 2009 en Wikipedia s.f).

La plataforma, conocida como una de las llamadas Redes Sociales, gradualmente implementaría funciones que ayudarían en gran medida a su funcionamiento. Una de ellas fue la creación del botón Me Gusta, que además de ser una forma simple de interactuar con el contenido sin responder con una respuesta escrita, permitió a la empresa tener un instrumento para recolectar información cuantificable en forma de datos estructurados (Cheddar, 2021). Más tarde, el botón Me Gusta daría paso a unos botones de reacciones que le dan mucha información en crudo a la empresa.

Es necesario hablar de dos elementos de los botones de reacción y la subida constante de contenido en forma de fotos y videos por parte de los usuarios. Asimismo, es necesario prestar atención al contenido al que reaccionan los usuarios (noticias, actualizaciones de estados de sus amigos, contenido patrocinado por empresas, entre otros). Dado que estos elementos son ahora la fuente de información que Facebook utiliza para hacer funcionar sus productos y las herramientas de IA ligadas a ellos, especialmente las relacionadas con el aprendizaje automático, que requieren una gran cantidad de datos para operar.

Facebook utiliza los datos de los botones de reacción para alimentar sus algoritmos, que alimentan herramientas de IA, especialmente las relacionadas con el aprendizaje automático. Según van Dijck et al (2018), "Las plataformas utilizan algoritmos para filtrar automáticamente enormes cantidades de contenido y conectar a los usuarios con contenido, servicios y anuncios".

Teniendo en cuenta cómo Facebook utiliza las herramientas de su interfaz de usuario para recolectar datos, es relevante conocer algunas de las herramientas de IA que utiliza con esos datos. Hazelwood et al (2018) del laboratorio de

Facebook, explican que "Las cargas de trabajo de aprendizaje automático de Facebook son extremadamente diversas: los servicios requieren muchos tipos diferentes de modelos en la práctica. Esta diversidad tiene implicaciones en todas las capas de la pila del sistema". Esta diversidad de cargas de trabajo se manifiesta en herramientas de IA específicas para realizar tareas específicas.

Con base en el texto de *Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective* (Hazelwood et al, 2018), Facebook aplica sus esfuerzos del aprendizaje de máquina en distintas áreas. En particular emplea las herramientas del aprendizaje máquina para mejorar la experiencia de usuarios, para que los anuncios comerciales lleguen al público que cumple su perfil, para la traducción del lenguaje y reconocimiento de voz, y para el reconocimiento de rostros. Para ello, emplea una serie de procesadores (CPU) y Unidades de procesamiento de gráficos (GPU), en aras de entrenar a sus sistemas para que puedan aprender y después producir un resultado.

A partir del trabajo de Hazelwood et al (2018), se puede apreciar que la fase de entrenamiento de las herramientas de Facebook se divide en dos niveles. El primero es la fase de entrenamiento, la cual se realiza fuera de línea, mientras que el segundo es "una fase de inferencia para ejecutar el modelo entrenado en producción y hacer un conjunto de predicciones en tiempo real. Esta fase se lleva a cabo en línea" (Hazelwood et al, 2018). La fase de entrenamiento, según la experiencia del autor al usar la plataforma Watson, consiste en mostrar una serie de datos a la máquina para que pueda aprender. Una vez que la máquina aprende a distinguir, por ejemplo, lo que es un gato, se le muestra una imagen fuera del juego de imágenes de entrenamiento. Si la reconoce con un margen de error bajo, se considera un éxito.

Para tener una idea más clara de lo anterior, hay que detenerse en *Facer*, la herramienta que Facebook emplea para el reconocimiento de imágenes:

"*Facer* es el marco de detección y reconocimiento de rostros de Facebook. Dada una imagen, primero encuentra todas las caras en esa imagen. Luego, ejecuta un algoritmo de reconocimiento facial específico del usuario para determinar la probabilidad de que esa cara pertenezca a uno de sus mejores amigos que haya

habilitado el reconocimiento facial. Esto le permite a Facebook sugerir a cuáles de sus amigos le gustaría etiquetar dentro de las fotos que carga” (Hazelwood et al, 2018).

En conclusión, Facebook junto con sus plataformas asociadas, como Instagram y Whatsapp, es uno de los ejemplos más evidentes de cómo la IA, especialmente el aprendizaje automático, ha llegado a los usuarios comunes y tiene aplicaciones comerciales importantes. Y se tiene que señalar que, a pesar de las expectativas iniciales y las críticas recibidas en el informe de Lighthill, el aprendizaje automático ahora forma parte de la vida cotidiana de muchas personas sin que se percatan de lo que sucede detrás de escena. Sin embargo, debido a la capacidad de aprendizaje de datos de la que dispone Facebook y los avances en el aprendizaje automático, la empresa ha utilizado estas herramientas en modelos y sistemas como el caso de la prevención del suicidio analizado en el capítulo anterior.

C. Didi y el aprendizaje de máquina en la economía gig

El siguiente caso relevante para examinar es el de la plataforma de economía gig llamada Didi. La trayectoria de Didi, que este autor observó en su tesis de maestría, está repleta de diferentes elementos económicos, históricos y tecnológicos importantes. Entre ellos, se encuentra el escenario en el que surgió, que regularmente está sujeto a la intervención del gobierno chino.

Didi es una compañía fundada por Cheng Wei en China en 2012 bajo el nombre de Didi Dache (Morton, 2021). La historia de la compañía, como muchas gigantes en el rubro de las tecnologías de la información y comunicación, es compleja. Por ejemplo, en 2015 se fusionó con Kuaidi Dache para crear Didi Chuxing (Fast company, 2019), la compañía que actualmente maneja la plataforma digital de trabajo. Más tarde, en 2016, Didi logró sacar del mercado a su mayor competidor, Uber, la empresa que simboliza a las plataformas digitales de trabajo de la economía gig.

La trayectoria de innovación de Didi está ligada a procesos tecnológicos, situaciones económicas, y elementos históricos que se han planteado como ejes del análisis de casos empíricos en este apartado. A causa de una necesidad, la de

un empresario que necesitaba un taxi en un momento puntual “y después de frustrarse en el intento por conseguir uno se le ocurrió la idea de un botón” (Morton, 2021), la idea de una plataforma dedicada a esto comenzó a surgir. Hay que resaltar que el contexto de ese empresario no era tan favorable, sí bien su idea coincide con la de los fundadores de Uber, el escenario, el momento en el que estaba plantado, no podía ser más diferente.

Mientras que los creadores de Uber estaban en medio de un país, de un estado como el de California, que apostaba a la innovación y la inversión en empresas de tecnologías de la información, el creador de Didi se encontraba en medio de un país cuya innovación era mediada fuertemente por el estado. Durante la fundación de Didi el gobierno chino intervenía de manera proteccionista en aras de apoyar a las empresas con o sin participación del estado. Aun así, “Sin la inversión directa del gobierno en su infraestructura, DiDi se convirtió en un gigante incomparable con una participación del 94.6% en el mercado de transporte” (Xiao, 2017 en Chen & Qiu 2019).

Didi se benefició y aprovechó entre otras cosas nuevas regulaciones. Por ejemplo, en 2016 el gobierno de China aprobó una ley que “...aclara la incertidumbre regulatoria y establece un nuevo marco bajo el cual las compañías de transporte pueden operar en el país” (Mozur en The New York Times, 2016). Uno de los elementos importantes que aprovechó Didi fue el gran avance en las ciencias de datos y en la IA, más allá de la tenacidad de su emprendedor, porque la empresa aprovechó distintos momentos jurídicos.

Hay que subrayar otro aspecto de la trayectoria de Didi: la situación económica de China. Muchos de sus avances y sus ocasos tuvieron lugar en países considerados potencias económicas, como se vio en la historia de la IA. Incluso Ada Lovelace, a pesar de sus restricciones de género o obligaciones como miembro de una familia burguesa, tuvo oportunidades y conexiones debido a que se encontraba en un país rico capaz de invertir en las ciencias y atraer talento en distintas disciplinas académicas.

Hasta antes de 1980, China era un país azotado por crisis económicas causadas por la caída de dos guerras del opio, la caída de su dinastía imperial, la invasión de Japón y la revolución socialista liderada por Mao Zé Dong. El gobierno de Mao también implementó el "Gran Salto Adelante", el cual causó malas cosechas y una crisis económica que resultó en la gran hambruna de 1959 a 1961 según Henry y Barragán (2013).

Entre 1978 y 1979, Deng Xiaoping, el sucesor de Mao introdujo una serie de reformas, incluyendo el abandono del sistema de responsabilidad rural y la creación de la ley de inversiones conjuntas, lo cual permitió la participación de inversión extranjera en los procesos de producción (Henry y Barragán, 2013). Esto llevó a un rápido crecimiento económico de China, primero como la fábrica del mundo y más tarde como una potencia de innovación independiente. Por ejemplo, en mayo de 2020 China anunció planes de invertir más de un billón de dólares en la economía a través del desarrollo de tecnologías clave, como redes inalámbricas y IA (Bloomberg News, 2020).

El plan de China consistía en obtener mayor independencia tecnológica con relación a las potencias occidentales. Y es en medio de este periodo donde DiDi comenzó a formarse como compañía. Así pues, por un lado, tenía la oportunidad de estar en un mercado emergente con una posible masa de usuarios; y en segundo lugar estaba en un país que comenzaba a hacer posible la respuesta a la pregunta de su creador Cheng Wei: ¿Cómo puedo conseguir un taxi en todo momento?

El plan de China consistía en obtener mayor independencia tecnológica con respecto a las potencias occidentales. Y es en medio de este periodo donde DiDi comenzó a formarse como compañía. Así pues, por un lado, tenía la oportunidad de estar en un mercado emergente con una posible masa de usuarios; y en segundo lugar, estaba en un país que comenzaba a hacer posible la respuesta a la pregunta de su creador Cheng Wei: ¿Cómo puedo conseguir un taxi en todo momento?

Cabe señalar que Didi representa la convergencia de herramientas de la IA desarrolladas. Desde las redes neuronales de lenguaje, modelos de sistemas expertos, redes neuronales profundas, entre otras. Cada herramienta, igual que otras plataformas como Uber, permite a la plataforma de servicios de taxi bajo demanda realizar tareas que facilitan la interacción entre conductores y usuarios (Chen y Qiu, 2019 en Morton, 2020).

La manera en la que DiDi ocupa la información es mediante distintas herramientas de IA, especialmente de aprendizaje automático. Por ejemplo, DiDi utiliza el aprendizaje por refuerzo para tareas como predecir y recomendar rutas a sus conductores, la tecnología de redes neuronales profundas para recolectar e interpretar la información, y redes recurrentes profundas (como los modelos ReLu propuestos por Hinton) para emparejar a los conductores con los usuarios (Wang y Qin en DiDi, 2019).

Otro aspecto importante de DiDi son los mecanismos con los que desarrolla sus herramientas de IA. Estos tienen un enfoque en la integración de talento a través de iniciativas que promueven alianzas con la sociedad civil y el sector educativo. Una de las mayores iniciativas es Didi Engine, que es "una iniciativa para innovadores digitales, cuyos objetivos son entrenar el talento tecnológico joven, generar capacidades de I+D y desarrollar proyectos de innovación y diseño de IA para el bienestar social en los mercados internacionales en expansión de DiDi" (The Wire, 2019 en Morton 2020).

En suma, DiDi surgió por la convergencia de factores históricos, tecnológicos y sociales que favorecieron la innovación. Además, en el desarrollo de la IA, especialmente en el aprendizaje automático, dos elementos clave contribuyeron a la creación de la plataforma DiDi: la gran cantidad de datos y el poder de cómputo disponible en los teléfonos inteligentes y ordenadores utilizados por la empresa. Finalmente, el papel de los objetos encarnados en diferentes herramientas de IA, que actúan como instrumentos en una orquesta, permiten llevar a cabo tareas específicas que contribuyen al funcionamiento sincronizado de la plataforma.

CAPÍTULO TERCERO

TAXONOMÍA Y PROBLEMAS DE LA IA

“Hoy, para lanzar nuestras plataformas, todas tienen una pieza de IA que les dice rápidamente cuál es el anuncio en el que es más probable que hagan clic” (Ng, en Coursera, s.f).

La IA se ha vuelto cada vez más presente en nuestra vida cotidiana, ya sea a través de aplicaciones de reconocimiento de voz o de entrenamiento. Sin embargo, el lenguaje y la terminología relacionados con la IA pueden resultar complejos para aquellos que no son expertos en el tema. Por ello, se requiere una explicación con respecto a los conceptos y nociones de esta tecnología.

Así pues, en este capítulo, se discutirán los conceptos, nociones y problemas relacionados con la IA. Se creará un marco teórico sobre los conceptos de las ciencias sociales y humanidades que se utilizarán para responder a la pregunta de investigación. Es necesario tener un conocimiento sólido de estos conceptos para entender de manera clara y precisa el fenómeno de estudio y su naturaleza interdisciplinaria.

La pregunta de investigación busca responder ¿En qué medida las plataformas digitales de Alphabet-Google, Uber y los sistemas de videovigilancia (C5) que utilizan IA en México y Japón vulneran derechos humanos relacionados con los principios éticos de transparencia y *explicabilidad*, rendición de cuentas y responsabilidad, y derecho a la intimidad y protección de datos, en la toma de decisiones automatizadas y los efectos sociales derivados de esta disrupción tecnológica en todas las etapas del ciclo vital de estos sistemas que utilizan la IA?

Para responder a esta pregunta, se requiere tanto una comprensión teórica de las ciencias sociales y humanidades, como un conocimiento de las herramientas y construcciones teóricas relacionadas con las ciencias y tecnología, ya que esta investigación tiene una naturaleza interdisciplinaria.

En este apartado se abordarán tres conceptos clave relacionados con la IA. El primero es el concepto de IA en sí misma; el segundo es el aprendizaje automático profundo o *Deep Learning*; y el tercero es el de las redes neuronales y los algoritmos asociados a ellas. Es importante destacar que la IA ayudará a comprender el fenómeno de estudio en una perspectiva más amplia, el impacto de sus herramientas y plataformas en el ámbito social y en el ecosistema general de aplicaciones.

Entender qué son las redes neuronales y los algoritmos asociados a ellas es fundamental para explicar cómo funciona el aprendizaje automático. Sin embargo, no se trata simplemente de hacer un resumen de términos y conceptos relacionados con la IA. También se utilizarán algunas herramientas planteadas en el marco teórico para comprender cómo estos fenómenos afectan a la realidad. Así, no solo se entenderá el funcionamiento de las herramientas de IA, sus implicaciones sociales y en la organización de los procesos de innovación y su impacto en los derechos humanos.

Finalmente, en este último apartado se analizarán los derechos y consecuencias éticas relacionadas con las herramientas de IA. Esto es relevante porque son el tema central de esta tesis y porque son problemas surgidos conforme se implementa esta tecnología. En resumen, este capítulo buscará esclarecer los conceptos de IA, aprendizaje automático, redes neuronales y sus algoritmos, así como los aspectos sociales y las consecuencias en los derechos humanos relacionadas con ellos.

1. El aprendizaje de máquina

La IA es considerada como "la nueva electricidad" (Ng, en Coursera, s.f). Dentro de la IA, especialmente en la IA acotada, donde se ha logrado el mayor avance, hay herramientas esenciales para su funcionamiento. Una de las herramientas más importantes en el campo de la IA es el aprendizaje automático, un campo de estudio que puede generar controversias entre quienes no están del todo familiarizadas con su funcionamiento.

Según Valverde (2019), el aprendizaje automático es "un subcampo de la IA que encuentra soluciones aproximadas a través de un proceso de generalización basado en los datos disponibles". Además, se puede definir como "la capacidad de un algoritmo para aprender a partir de datos anteriores para producir un comportamiento inteligente y tomar decisiones correctas en situaciones que nunca antes había enfrentado" (Saeed y Al-Mushayt, 2020). Por último, Gillies (en Coursera, s.f) explica que "el aprendizaje automático consiste en crear programas estadísticos llamados modelos. Un modelo toma una entrada y devuelve una salida" (Gillies, en Coursera, s.f).

Hay que destacar que, según O'Neill (2016), un modelo *no* es "una representación abstracta de algún proceso, ya sea un juego de béisbol, la cadena de suministro de una compañía petrolera, las acciones de un gobierno extranjero o la asistencia a un cine" (O'Neill, 2016). La autora argumenta que el modelo utiliza lo que conocemos para predecir varios escenarios. Por ejemplo, una persona que tiene dos gatos puede crear un modelo que le ayude a determinar cuánta comida darles, sus preferencias alimenticias y cómo evitar que engorden.

Según los ejemplos de O'Neill, se puede deducir que el input de mi modelo serían la personalidad de mis gatos, la cantidad de comida disponible y sus preferencias alimenticias, mientras que el output sería la hora de comer y cuántas raciones les proporciona. Al evaluar el éxito del modelo, es decir, cómo están comiendo los gatos y si están engordando, se pueden realizar ajustes y actualizaciones para mejorar su desempeño, lo que se conoce como un "modelo dinámico" (O'Neill, 2016).

Por consiguiente "para crear un modelo, tomamos decisiones sobre lo que es lo suficientemente importante para incluir, simplificando el mundo en una versión de juguete que se puede entender fácilmente y de la cual podemos inferir hechos y acciones importantes" (O'Neill, 2016). Asimismo, la autora afirma que un modelo se debe actualizar constantemente. Así, se puede imaginar al modelo como un instrumento musical que se debe afinar cotidianamente para desempeñar su función.

La IA, especialmente en el campo del aprendizaje automático, busca encontrar soluciones a través de un proceso de generalización basado en los datos disponibles. Los algoritmos son la herramienta clave para llevar a cabo este proceso. En el pasado, se creía que cuantos más datos se tuvieran, mejor sería la eficiencia del aprendizaje automático; sin embargo, con el avance tecnológico, se ha descubierto que no necesariamente es así.

Para que una IA pueda aprender y generar un resultado específico, como la creación de una pintura renacentista, es necesario entrenarla con una gran cantidad de datos. Esto genera una relación entre entrada y salida, donde la entrada son los datos y la salida es el resultado. El ejemplo del reconocimiento de un gato es un buen ejemplo de cómo se puede entrenar una máquina para generar un resultado específico.

Además, los datos también pueden clasificarse en dos tipos: estructurados y no estructurados. Los datos estructurados tienen un significado bien definido, como el tamaño de una casa o la edad de un usuario (Ng, en Coursera, s.f). Por otro lado, los datos no estructurados son aquellos que no tienen un significado bien definido, como imágenes o canciones (ídem): en estos casos, la IA debe reconocer lo que hay en la imagen o en el texto.

Asimismo, el aprendizaje de máquina requiere de redes neuronales artificiales para entrelazar los distintos modelos. Para comprender qué son las redes neuronales, hay que hacer un simple ejercicio de comparación con las humanas. Según Andrew Ng, "el tipo de aprendizaje automático más utilizado es un tipo de IA que aprende de A a B, o asignaciones de entrada a salida...A esto se le llama aprendizaje supervisado: entrada + salida" (Ng, en Coursera, s.f). Además, menciona que "con la IA moderna, con redes neuronales y aprendizaje profundo, lo que vimos fue que, si entrena una red neuronal pequeña, el rendimiento se ve así, donde a medida que los alimenta con más datos, el rendimiento mejora durante mucho más tiempo" (Ng, en Coursera, s.f).

El aprendizaje de máquina funciona con una serie de herramientas, como los datos, los modelos y las redes neuronales, que le permiten a la máquina aprender

y dar un resultado. Además, a medida que las redes neuronales son más complejas, es decir, tienen más capas, son más eficientes y pueden realizar tareas que requieren mayor complejidad, como el reconocimiento de imágenes, esto se conoce como aprendizaje de máquina profundo, que está ligado a redes neuronales profundas debido a su complejidad, y en aras de matizar esto Gilles (en Coursera, s.f) explica:

“La palabra profundo en el aprendizaje profundo significa que hay muchas, muchas capas de neuronas que se alimentan de otras neuronas. Esto significa que, aunque los cálculos básicos son muy simples, una red neuronal profunda puede aprender características muy complejas” (Gillies, en Coursera, s.f).

Para concluir, según Andrew Ng, las redes neuronales han revolucionado el aprendizaje supervisado y han generado un gran valor económico a pesar de haber existido durante décadas, como se explicará en el próximo capítulo. Además, el académico menciona que, gracias al aprendizaje profundo y las redes neuronales, las computadoras son capaces de interpretar mejor los datos no estructurados en comparación con hace unos años. Con base en el desarrollo de la IA, esta tecnología finalmente está cumpliendo con las expectativas de hace décadas.

A. Redes neuronales

Es fundamental dedicar un apartado a la explicación de las redes neuronales. Según el Dr. Gillies (s.f., en Coursera), una neurona en este contexto es “una unidad de red neuronal, pero una red puede contener cientos de neuronas, cuya salida se alimenta a las entradas de muchas otras”. Respecto al aprendizaje automático, Gillies destaca que “muchos algoritmos de aprendizaje se basan en matemáticas bastante simples, pero funcionan porque estos cálculos pueden escalar hasta miles de características en las neuronas”.

Las redes neuronales pueden entenderse como funciones logarítmicas que generan resultados en función de una entrada o comando (input). Están compuestas por neuronas artificiales organizadas en una arquitectura específica para procesar datos y obtener una salida significativa (Marini, 2022). En las redes

neuronales artificiales (artificial neural networks, ANN), las neuronas se estructuran en capas, donde las conexiones solo se establecen entre capas consecutivas y no existen bucles dentro de la red. Su estructura consta de una capa de entrada, una de salida y varias capas ocultas (Hölldobler & Möhle, 2017).

A medida que las redes neuronales crecen en tamaño y complejidad, la inteligencia artificial adquiere la capacidad de realizar tareas cada vez más avanzadas, como el reconocimiento de patrones en música, voz, lenguaje e imágenes. En este sentido, Gillies (s.f., en Coursera) señala que “dado que es una red neuronal, la salida de un filtro puede alimentarse a otros filtros para crear nuevas características complejas, que a su vez se procesan en capas sucesivas para generar características aún más sofisticadas. Esta es la tecnología fundamental detrás del reconocimiento de imágenes en el aprendizaje automático”.

No obstante, los algoritmos de aprendizaje profundo plantean desafíos importantes debido a su complejidad y falta de transparencia. En muchos casos, los modelos aprenden de manera autónoma, lo que dificulta que los diseñadores comprendan plenamente su funcionamiento. Esto no solo complica la detección y corrección de errores, sino que también puede generar sesgos en la toma de decisiones. De hecho, algunos de los algoritmos más avanzados de IA, como los de aprendizaje profundo, son difíciles de interpretar para los humanos, ya que se basan en técnicas de cálculo que no siguen el razonamiento simbólico-lingüístico propio del pensamiento consciente (Gavilán, 2020).

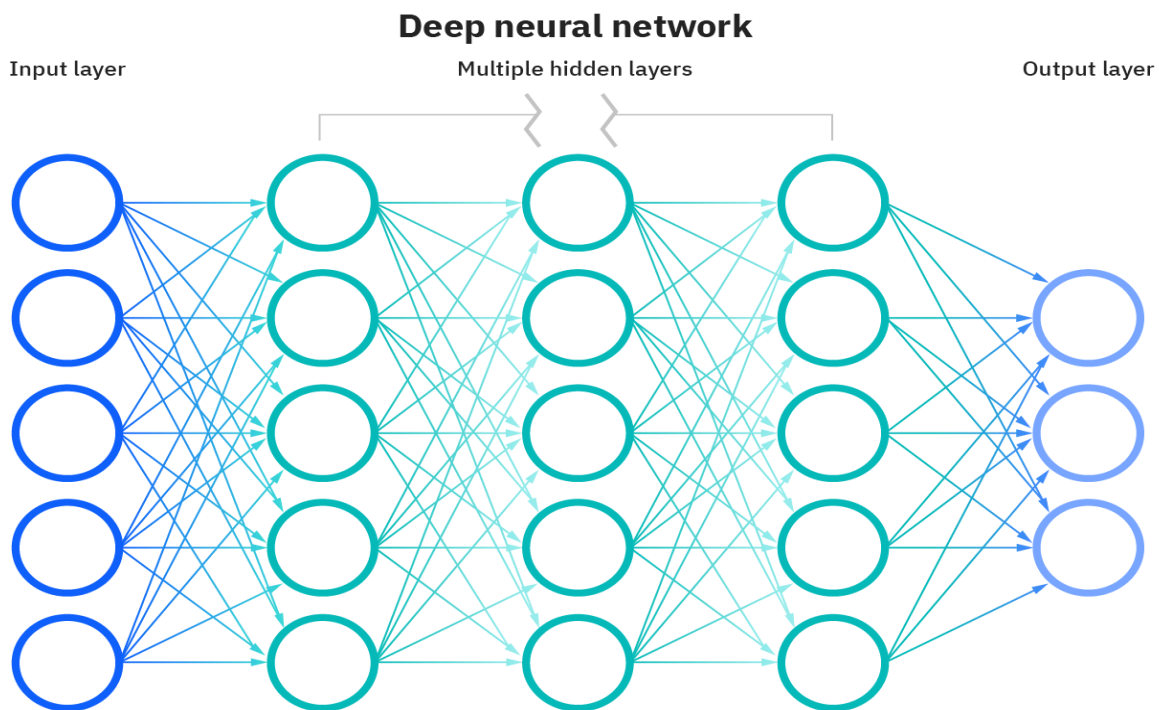
Este problema, conocido como falta de explicabilidad, se refiere a la dificultad de comprender cómo y por qué un algoritmo llega a una determinada conclusión o predicción (IBM, s.f.). La falta de transparencia en estos sistemas puede tener consecuencias significativas en ámbitos como la salud, la justicia o la educación, donde la confianza y la rendición de cuentas son esenciales.

Las redes neuronales profundas, utilizadas en el aprendizaje automático, permiten realizar tareas cada vez más complejas a medida que aumenta la cantidad de capas neuronales. Este tipo de aprendizaje, conocido como deep learning, se

basa en el entrenamiento de redes de gran escala y se emplea en aplicaciones cotidianas como el reconocimiento de voz e imágenes (Ng, s.f., en Coursera). Además, su capacidad para procesar grandes volúmenes de datos ha impulsado avances en campos como la medicina, la automatización y la inteligencia artificial generativa.

Para comprender mejor el concepto de red neuronal, es crucial diferenciarlo de las redes neuronales profundas. Andrew Ng, en su curso de aprendizaje automático, enfatiza que “cuando hablamos de redes neuronales artificiales, es solo para subrayar que no estamos refiriéndonos al cerebro biológico, sino a un software” (Ng, s.f., en Coursera). También señala que “cuando representamos gráficamente una red neuronal artificial, existe una analogía muy vaga con el cerebro” (Ng, s.f., en Coursera).

Por su parte, la Royal Society (2017) define las redes neuronales como “un enfoque del aprendizaje automático en el que las capas de unidades computacionales están conectadas de manera inspirada en las conexiones entre las neuronas del cerebro”. Esta definición resalta dos aspectos clave: si bien las redes neuronales artificiales toman inspiración de las redes neuronales biológicas, no son equivalentes; además, su funcionamiento se basa en funciones matemáticas y estadísticas, lo que introduce un nivel de abstracción necesario para su comprensión (imagen 9, en Google).



(imagen 9, en Google)

Según Ng (en Coursera, s.f), en la figura de arriba, los pequeños círculos se utilizan para abreviar capas y funciones. Por ejemplo, si se quiere conocer el precio de un automóvil, los dos primeros círculos azules representan el color y la marca del coche, lo que da como resultado una nueva función que se representa en el círculo verde subsecuente. Más abajo, un tercer círculo representa un nuevo valor, como la antigüedad, y se combina con la primera nueva función, y así sucesivamente.

“Esta gran red neuronal artificial es solo una gran ecuación matemática que le dice, dadas las entradas A, cómo se calcula el precio B”, y “...entonces esa es una red neuronal, es un grupo de neuronas artificiales, cada una de las cuales calcula una función relativamente simple”;...pero cuando apila suficientes de ellos juntos como ladrillos de Lego, pueden calcular funciones increíblemente complicadas que le brindan asignaciones muy precisas desde la entrada A hasta la salida B” (Ng, en Coursera, s.f).

Las redes neuronales, como se puede ver en la imagen ocho, pueden crear varios filtros uno tras otro, en un proceso conocido como red neuronal convolucional.

Este proceso es esencial para el funcionamiento actual de las herramientas de IA. Una neurona convolucional se da cuando el resultado de un filtro genera otro filtro, que a su vez se alimenta con otra variable, creando características más complejas. Según el doctor Gillies (en Coursera, s.f) “Una neurona convolucional es básicamente un filtro de Photoshop que transforma una imagen” (otro ejemplo son los filtros de imágenes populares en aplicaciones).

Más arriba se mencionó que las redes neuronales profundas, una de las herramientas del aprendizaje de máquina, pueden tener distintas aplicaciones. Entre ellas se encuentra el reconocimiento de voz, pero es pertinente retomar una explicación de cómo funcionan para este propósito:

“Después de clasificar el orden de las neuronas en la primera capa por su respuesta de frecuencia pico sintonizada, el análisis de la segunda capa sugiere que se está aprendiendo un conjunto de estructuras de filtro en forma de peine; esto demuestra la capacidad de la red para converger en una estructura casi idéntica a los métodos tradicionales de procesamiento de señales de seguimiento de tono” (Tan & Lim, 2018).

Con respecto a la forma abstracta de cómo la IA emplea redes neuronales como herramienta del aprendizaje de máquina, Geoffrey Hinton, uno de los pioneros de las redes neuronales, explica lo siguiente:

“Creo que la idea de que los pensamientos deben estar en algún tipo de lenguaje es tan tonta como la idea de que entender el diseño de una escena espacial debe ser en píxeles, los píxeles entran...Entonces creo que los pensamientos son solo estos grandes vectores, y que los grandes vectores tienen poderes causales. Causan otros grandes vectores, y eso es completamente diferente a la visión estándar de la IA de que los pensamientos son expresiones simbólicas” (Hinton en Ng y Coursera, s.f).

Las redes neuronales profundas son una colección de capas de funciones matemáticas que surgieron para representar al cerebro humano. Estas redes neuronales están presentes en varias aplicaciones que utilizan el aprendizaje automático, como la recomendación de contenido, el reconocimiento de imágenes y el procesamiento y reproducción de voz y lenguaje. Sin embargo, las redes neuronales profundas son solo una de las muchas herramientas de IA y

aprendizaje automático, los cuales están divididos en varias técnicas y categorías que es importante explicar.

B. Tipos de aprendizaje de máquina

El aprendizaje automático es diverso. Esta herramienta se divide en categorías y funciones. Cada subcategoría de aprendizaje automático usa muchas técnicas y algoritmos para funcionar; por lo tanto, sus aplicaciones son bastante diversas y no todas son accesibles para uso comercial, industrial o común.

La primera subcategoría del aprendizaje automático es el aprendizaje supervisado. Como su nombre indica, el aprendizaje supervisado necesita del apoyo constante de un ser humano en su proceso de entrenamiento y aprendizaje. Regularmente, el aprendizaje supervisado se divide en dos fases: la primera, el entrenamiento, que consiste en seleccionar datos para que la máquina aprenda de ellos; la segunda, la fase de prueba, en la que la máquina utiliza lo aprendido para obtener resultados con datos no pertenecientes a la fase de entrenamiento.

El ingeniero Andrew Ng menciona que hay dos formas para procesar datos y para implementar herramientas de sistemas de aprendizaje profundo. La primera, son las Unidades de procesamiento de cómputo o CPU; la segunda, son las Unidades de procesamiento gráfico o GPU. Normalmente los CPU se han usado en el aprendizaje de máquina ya que están diseñados para hacer tareas generales, pero, en los últimos diez años, se observó que los GPU también eran bastante eficientes para realizar operaciones de aprendizaje de máquina y procesamiento de datos: —inclusive los GPU puede ser más prácticos debido a que se pueden construir enormes servidores a base de grandes cantidades de estos.

Un ejemplo práctico del *aprendizaje supervisado* es el que se emplea en algunas plataformas sociales como Facebook. La empresa de Facebook, ahora llamada Meta, emplea el aprendizaje supervisado en distintas herramientas de su plataforma, y sí bien no mencionan directamente que emplean el aprendizaje supervisado, con base en una de las técnicas que mencionan en el artículo

Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective se observa que en efecto este puede ser el caso:

- “1. Una fase de formación para construir el modelo. Esta fase generalmente se realiza sin conexión.
2. Una fase de inferencia para ejecutar el modelo entrenado en producción y hacer un conjunto de predicciones en tiempo real. Esta fase se realiza online” (Hazelwood et al., 2018b).

Otra subcategoría del aprendizaje de máquina es el *aprendizaje no supervisado*. En este caso, la IA no necesita pasar por una etapa previa de entrenamiento con base en categorías y datos seleccionados por los seres humanos. Así, los modelos del aprendizaje no supervisado permiten a la máquina generar categorías y con ello aprender de datos sin que un actor humano necesite un entrenamiento previo.

El aprendizaje no supervisado es un método de aprendizaje automático que analiza y agrupa conjuntos de datos sin etiquetar, es decir, sin un conocimiento previo de las categorías o clases a las que pertenecen los datos (Chen et al., 2022). Estos algoritmos descubren patrones ocultos o estructuras en la información, lo que permite explorar datos desconocidos, comprimirlos o reducir su dimensionalidad (Inesdi, s.f.). El aprendizaje no supervisado se puede clasificar en diferentes tipos según el objetivo y la técnica utilizados, como el agrupamiento, la asociación o la reducción de dimensionalidad (IBM, s.f.). Así pues, el aprendizaje no supervisado tiene aplicaciones en diversos campos, como el reconocimiento de imágenes, la segmentación de clientes, la minería de textos o el análisis de señales (TIBCO Software, s.f.).

Quizás el ejemplo más cercano de aprendizaje no supervisado se encuentra en las herramientas para detectar correo electrónico basura. En este caso, el aprendizaje no supervisado busca los vecinos más cercanos, aquellos datos que se relacionan entre dos objetos, para determinar qué es un correo útil y que es una basura; además busca agrupar los datos en categorías; pescados azules, pescados rojos, pescados verdes, entre otros. Para ello, el aprendizaje no

supervisado utiliza tres instrumentos, las cuáles son agrupación –que ya se mencionó–, asociación y deducción dimensional.

1. “La agrupación en clústeres es una técnica de minería de datos para agrupar datos sin etiquetar en función de sus similitudes o diferencias. Por ejemplo, los algoritmos de agrupación de *K-means* asignan puntos de datos similares en grupos, donde el valor de K representa el tamaño de la agrupación y la granularidad. Esta técnica es útil para la segmentación del mercado, la compresión de imágenes, etc.
2. La asociación es otro tipo de método de aprendizaje no supervisado que usa diferentes reglas para encontrar relaciones entre variables en un conjunto de datos determinado. Estos métodos se utilizan con frecuencia para el análisis de la cesta de la compra y los motores de recomendación, según las recomendaciones de "Clientes que compraron este artículo también compraron".
3. La reducción de dimensionalidad es una técnica de aprendizaje que se utiliza cuando la cantidad de características (o dimensiones) en un conjunto de datos dado es demasiado alta. Reduce el número de entradas de datos a un tamaño manejable y preserva la integridad de los datos. A menudo, esta técnica se utiliza en la etapa de preprocesamiento de datos, como cuando los codificadores automáticos eliminan el ruido de los datos visuales para mejorar la calidad de la imagen” (Delua, et al, 202).

El aprendizaje no supervisado representa algunas ventajas frente al supervisado. Quizás la mayor ventaja del aprendizaje no supervisado es el tiempo que se requiere para que aprenda y pueda manifestarse en una herramienta útil. Sin embargo, dado que no aprende a partir de etiquetas de categorías predeterminadas, es más susceptible a tener mayores márgenes de error; asimismo, dada su complejidad requiere de una mayor cantidad de recursos y de poder de cómputo que el de su par supervisado.

Un ejemplo común de aprendizaje no supervisado es el reconocimiento de gatos. Andrew Ng explica el proceso cuando menciona que: “YouTube aparentemente

tiene muchos videos de gatos. Pero fue un resultado notable que, sin decirle de antemano que debía encontrar gatos, el sistema de IA, el algoritmo de aprendizaje sin supervisión pudo descubrir por sí mismo qué es un gato” (en Coursera, s.f). Aunque el ejemplo puede parecer divertido o anecdótico, plantea una de las formas en las que el aprendizaje no supervisado ha encontrado un espacio en aplicaciones cotidianas.

La siguiente categoría del aprendizaje automático es el aprendizaje por refuerzo. Según Steffen Hölldobler, Sibylle Möhle y Anna Tigunova (2017): “Conceptualmente, el aprendizaje por refuerzo es un método para entrenar a un agente de IA sin darle explícitamente la respuesta correcta. En cambio, el objetivo del agente es maximizar la recompensa, que es una retroalimentación del entorno simbólico a sus acciones”. Es decir, en el aprendizaje por refuerzo, el agente no humano trata de encontrar elementos que le permitan obtener un resultado que le genere un incentivo dentro de un ambiente que se analiza a nivel simbólico, como traducir la melodía de una canción a elementos matemáticos que más tarde puedan ser clasificados.

Quizás el ejemplo más claro y cotidiano de aprendizaje por refuerzo es la plataforma Netflix. La plataforma de streaming Netflix ofrece video a través de internet mediante una suscripción mensual. El modelo de negocio de Netflix se basa en el crecimiento constante y retención de suscriptores a través de la creación y recomendación de contenido. El sistema de recomendación de contenido es esencial, ya que ayuda a los suscriptores a navegar una gran cantidad de series, películas y documentales, mientras los retiene en la plataforma.

Sin embargo, con base en la extensa y variada librería de la plataforma Netflix, además de su gran cantidad de usuarios, es prácticamente imposible crear un mecanismo de recomendación manual o automatizado (por medio de mera programación). Por ello, la plataforma de contenido audiovisual ha decidido emplear el aprendizaje de máquina. De acuerdo con el sitio de Investigación de Netflix:

“El aprendizaje automático tiene un impacto en muchas áreas interesantes de nuestra empresa. Históricamente, la personalización ha sido el área más conocida, donde el aprendizaje automático impulsa nuestros algoritmos de recomendación. También utilizamos el aprendizaje automático para ayudar a dar forma a nuestro catálogo de películas y programas de televisión mediante el aprendizaje de las características que hacen que el contenido sea exitoso” (Netflix Research, s. f.).

En cuanto al uso del aprendizaje de máquina por refuerzo, la plataforma Netflix afirma que es una de varias herramientas que utiliza para ofrecer y mejorar sus servicios. Por la naturaleza del modelo de recomendación, se requiere un aprendizaje refuerzo a partir de las actividades observadas de los usuarios. Sin embargo, se tiene que conocer la postura del equipo de investigación de Netflix: “Nuestra investigación abarca muchos enfoques algorítmicos diferentes, incluidos modelos causales, bandidos, aprendizaje por refuerzo, conjuntos, redes neuronales, modelos gráficos probabilísticos y factorización matricial” (Netflix Research, s.f.).

El aprendizaje por refuerzo también ha encontrado un nicho importante en los videojuegos. Cuando un jugador se enfrenta a un personaje no jugable que funciona a través del aprendizaje de máquina que alimenta la IA. De esta manera, el personaje no jugable (que no controla el jugador y que muchas veces es un el enemigo controlado por la máquina) aprenderá del jugador y responderá a sus acciones de manera más realista.

Aun así, en el caso de los videojuegos, el aprendizaje por refuerzo se debe de emplear con cuidado. Esto es, en aras de evitar que el jugador caiga en lo que Freud llamó el *Valle inquietante*⁵:

Ahora bien, a pesar de que con los avances tecnológicos en los videojuegos se pueden implementar NPC o enemigos con acciones impredecibles, ese tipo de personajes no necesariamente ayudan a la experiencia del jugador por lo que, a

⁵ El concepto del valle inquietante fue retomado por otros psicólogos como Lacan quien lo describe como “El campo donde no sabemos distinguir el bien del mal, el placer del disgusto” (Lacan, 2005). Más adelante se usará ese concepto para describir las implicaciones éticas de la inteligencia artificial.

pesar de que existen máquinas con modelos de IA como el aprendizaje profundo, en el diseño del juego se busca crear un balance entre lo predecible o impredecible que busque enriquecer la experiencia del video jugador (Statt, 2019 en Morton, 2021).

Otro tipo de IA es la generativa. Este tipo de plataformas está ganando popularidad entre la comunidad académica y también entre muchos grupos de personas después de que el modelo Chat GPT, capaz de generar, corregir y elaborar textos, se hizo público. Como su nombre indica, la IA es capaz de generar contenido a partir del análisis de datos, utilizando "sistemas expertos que contienen bases de conocimiento y un motor de inferencia que genera contenido a través de una base de datos de reglas if-else" (Gozalo-Brizuela & Garrido-Merchán, 2023). Sin embargo, según Gozalo-Brizuela y Garrido-Merchán (2023):

La IA generativa moderna contiene un modelo discriminador o transformador entrenado en un corpus o conjunto de datos que es capaz de mapear la información de entrada en un espacio latente de alta dimensión y un modelo generador, que es capaz de generar un comportamiento estocástico creando contenido novedoso en cada nuevo ensayo incluso desde el mismos avisos como una entrada, realizando aprendizaje no supervisado, semi-supervisado o supervisado, dependiendo de la metodología particular de la IA convulsiva.

Es importante destacar que, a pesar de lo impresionantes que pueden ser los modelos naturales de lenguaje con sistemas de IA generativa como el ChatGPT de OpenAI o GLaM, LaMDA y PaLM de Alphabet-Google, estos son básicamente matemáticas que generan contenido a partir de datos existentes. Por lo tanto, estos sistemas solo pueden funcionar, como se explicará en el capítulo tres, con una gran cantidad de datos y una gran cantidad de cómputo. Esto es mencionado por Gozalo-Brizuela y Garrido-Merchán (2023), quienes explican que solo existen unos pocos desarrolladores de modelos de IA generativa (OpenAI, Google Research, DeepMind, Meta AI, entre otros) debido a la gran cantidad de recursos que estos sistemas requieren.

Así pues, existen tres categorías principales del aprendizaje de máquina: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo. Cada una de estas categorías tiene aplicaciones en la vida cotidiana, desde el

reconocimiento de imágenes hasta los sistemas de recomendación. Sin embargo, es importante mencionar que también existen categorías intermedias o híbridas, como el aprendizaje semi-supervisado, que también tienen aplicaciones en la vida cotidiana y son parte del gran ecosistema de plataformas y herramientas utilizadas en la IA.

C. Aprendizaje de máquina profundo

El aprendizaje de máquinas profundo es uno de los conceptos clave de la IA, y su importancia se ha vuelto cada vez más evidente en las últimas décadas. Desde su aplicación en tecnologías como reconocimiento de voz y visión, hasta su uso en la creación de nuevas herramientas y algoritmos, el aprendizaje profundo ha tenido un impacto significativo en el campo de la IA. Sin embargo, para entender plenamente el estado actual de la IA, es esencial analizar este concepto con mayor profundidad.

La relación entre las redes neuronales y el aprendizaje profundo es otro aspecto crucial en el campo de la IA. Aunque estos términos a menudo se usan indistintamente, es importante comprender las diferencias y similitudes entre ambos. Según Andre Ng "las redes neuronales se inspiraron originalmente en el cerebro, pero los detalles de cómo funcionan no tienen casi ninguna relación con cómo funcionan los cerebros biológicos" (Ng, en Coursera, s.f.). Se tiene que seguir investigando y discutiendo estas relaciones para entender mejor la ciencia de frontera que es la IA.

Según Valverde (2019), en el campo de la IA, existe una tendencia hacia el uso de aprendizaje profundo de extremo a extremo (*end-to-end*) para evitar la ingeniería de características (*feature engineering*) y extraer características relevantes directamente de los datos brutos. Esto tiene como objetivo agilizar el procesamiento de datos y mejorar el aprendizaje y la generación de resultados. Además, esta metodología se diferencia de la idea clásica de programación, donde se insertan reglas y contingencias a partir de un código.

El aprendizaje profundo (deep learning) se fundamenta en el aprendizaje automático (machine learning), el cual opera a partir de tres componentes esenciales: modelo, datos y función de pérdida. Este proceso sigue un principio similar al método científico de “prueba y error”, ajustando continuamente un modelo en función de la diferencia entre sus predicciones y los datos reales del fenómeno que analiza (Jung, 2022). En cierto modo, este mecanismo se asemeja al aprendizaje humano, que no depende de rutas preprogramadas, sino que se adapta mediante la experiencia y la corrección de errores. Sin embargo, a diferencia del aprendizaje humano, el aprendizaje automático se basa principalmente en funciones matemáticas y estadísticas para optimizar resultados.

Desde una perspectiva técnica, el aprendizaje profundo puede definirse como una función que mapea datos de entrada sin procesar (como una imagen médica) hacia una salida deseada (como un diagnóstico), minimizando una función de pérdida mediante técnicas de optimización como el descenso de gradiente estocástico (SGD) (Saeed y Al-Mushayt, 2020). Además, este enfoque se caracteriza por emplear múltiples capas de procesamiento que interpretan datos en distintos niveles de abstracción, lo que permite extraer patrones complejos y generar representaciones más precisas (Mishra et al., 2021).

Un ejemplo práctico de su aplicación es el uso de aprendizaje profundo en plataformas de comercio electrónico, como Amazon, donde la IA analiza continuamente los hábitos de consumo de los usuarios para ofrecer recomendaciones personalizadas basadas en patrones de compra. Este proceso se logra mediante el aprendizaje de representaciones, construyendo múltiples capas que establecen conexiones entre los datos de entrada y las salidas deseadas. La profundidad de estas capas determina la capacidad del modelo para representar y procesar información de manera eficiente (ídem).

El aprendizaje profundo es un subcampo del aprendizaje automático que ha logrado resultados sobresalientes en varios dominios, como la visión por computadora y el procesamiento del lenguaje natural (Saeed y Al-Mushayt, 2020). Esto se refleja en aplicaciones de uso cotidiano como los procesadores de texto o

plataformas de mensajes que pueden predecir desde la próxima oración que la persona va a escribir hasta aprender de su voz para tomar dictado. Además, permite comprender el funcionamiento intuitivo de las cámaras de los teléfonos inteligentes como ajustar la luz, enmarcar a las personas para generar contraste o etiquetar y hacer categorías de las fotos.

El aprendizaje profundo tiene varias ventajas, pero una de las más relevantes es su bajo costo energético en el procesamiento de datos. Según Deng, Yunbin (2019), una vez que se construye un modelo de sistema de aprendizaje profundo, se puede implementar en un dispositivo móvil con mucho menos poder de cálculo. Esto significa que, a pesar de que diseñar un sistema de aprendizaje profundo implica un costo significativo, la implementación de este es relativamente barata, especialmente debido a la disponibilidad de unidades de procesamiento gráfico (GPU) eficientes y de bajo costo en muchos dispositivos.

Es importante señalar los usos cotidianos del aprendizaje de máquina profundo, ya que estos ejemplos empíricos permiten hacer una traducción entre el universo técnico de los científicos de datos (aquellos que se dedican a la IA) y las personas que estudian los fenómenos de las ciencias sociales y humanidades (CSH). Estos ejemplos y la noción de que la IA tiene un impacto en la vida cotidiana más allá de los laboratorios crean un puente entre el lenguaje técnico de la rama y los casos empíricos y problemáticas de los investigadores de las CSH.

De acuerdo con Gillies (en Coursera), el aprendizaje de máquina tiene varias aplicaciones, como el reconocimiento de imágenes y la traducción de textos. El investigador plantea que uno de los usos más importantes del aprendizaje de máquina tiene que ver con la lengua hablada y escrita: "El motor de búsqueda de Google se basa en gran medida en el aprendizaje automático del texto de las páginas web para ofrecer buenos resultados" (Gillies, en Coursera, s.f). Esto, junto con el uso común de herramientas de Google y Alphabet, demuestra que el aprendizaje profundo e IA se han convertido en una parte común de nuestra vida cotidiana.

El aprendizaje de máquina depende de una gran cantidad de capas. Según Saeed y Al-Mushayt (2020), "La arquitectura de red de un modelo de aprendizaje profundo se refiere a la organización general de sus capas. Estas capas se pueden organizar en diferentes estructuras con diferente número de neuronas en cada capa y un conjunto distinto de parámetros." Se aprecia que no existe una forma de aprendizaje máquina profundo, sino herramientas que se ajustan a las necesidades de investigadores, desarrolladores, plataformas y empresas.

Hay que resaltar que uno de los elementos que han hecho posible el uso de herramientas de aprendizaje profundo es la cantidad de datos. Según Gillies, "la recomendación de Amazon parece increíblemente precisa, es porque se aprendió de sus cientos de millones de clientes, y esa es una de las características del aprendizaje automático moderno" (Gillies, en Coursera, s.f.). La gran cantidad de datos está ligada al gran número de usuarios que suben información y contenido a internet a través de dispositivos fijos y móviles.

La cantidad de datos producida y el acceso a ellos es uno de los elementos que hicieron que la IA pasará de ser una promesa de laboratorios de investigación a una realidad práctica en la vida cotidiana. El segundo elemento es la gran cantidad de cómputo y el desarrollo de nuevas herramientas, algoritmos y técnicas de aprendizaje de máquina. En suma, el aprendizaje de máquina es la convergencia de muchas tecnologías.

Anteriormente se explicó el concepto de datos y su importancia en el aprendizaje de máquina. Andrew Ng (en Coursera, s.f) comenta que la forma en que la IA interpreta los datos se "puede pensar en los cuatro pasos del algoritmo como estos cuatro pasos, donde el primer paso es la detección de palabras de activación, el segundo paso, el reconocimiento de voz, luego el reconocimiento de intenciones y, finalmente, la ejecución de los comandos que el usuario le pidió al altavoz inteligente" (Ng, en Coursera, s.f).

Se ha mencionado que existen datos estructurados y no estructurados, y que la IA tiene la capacidad de interpretarlos, pero que no necesita una gran cantidad para funcionar. Según se producen muchos datos diariamente, las empresas e

investigadores acceden a una fracción de estos a través de sus plataformas o ecosistemas ligados a otras organizaciones. Según Bulau (2021), "la persona promedio producirá 2,5 trillones de bytes de datos por día en 2020". Estos datos se convierten en el combustible que impulsa el aprendizaje de máquina.

D. Aplicaciones del aprendizaje de máquina

La IA tiene muchas aplicaciones debido a los avances que ha tenido esta tecnología, en especial el aprendizaje de máquina profundo. Por ello, es pertinente dedicar un apartado para explicar cuáles son las distintas aplicaciones, cuál el contexto realista en que se encuentra la tecnología, y cuáles son las aplicaciones de frontera que se están desarrollando e investigando.

De acuerdo con Gilles (en Coursera, s.f): "La IA se divide en una amplia gama de subcampos que abordan desafíos particulares como la visión por computadora, computadoras que pueden ver y comprender. Procesamiento del lenguaje natural, computadoras que entienden el lenguaje humano. Robótica e incluso creatividad computacional". A veces se usan estos subcampos sin que la persona se dé cuenta, como al sugerir una palabra en un procesador de texto. El hecho de que la IA no se perciba en las aplicaciones diarias tiene que ver en parte con el avance de esta tecnología, así como en su aplicación:

"Muchos de estos subcampos han dado lugar a sistemas informáticos que pueden hacer cosas increíblemente útiles. Robots que pueden hacer funcionar fábricas. El motor de búsqueda de Google, que se basa en un sofisticado procesamiento del lenguaje natural, y computadoras que incluso pueden diagnosticar enfermedades observando escaneos de rayos X" (Gillies, en Coursera, s.f).

Ahora bien, como lo planteó Gilles (en Coursera, s.f), no todas las aplicaciones de la IA salen del laboratorio. Y como se verá en el siguiente capítulo, la trayectoria de la IA es todo menos constantemente ascendente. Aun así, es relevante seguir remarcando qué puede hacer y qué no la IA para conocer mejor la tecnología y comprender el lenguaje que se maneja alrededor de está, y más importante aún, sus implicaciones éticas.

Más allá de los casos empíricos que se han mencionado, “el aprendizaje automático es el que procesa los datos de compras anteriores y las compras de otros, y lo usa para detectar patrones y hacer predicciones” (Royal Society, 2017). Asimismo, tiene implicaciones para el reconocimiento de imagen, ya sea en cámaras de videovigilancia y plataformas sociales como Facebook. Además, cada vez tiene mayores capacidades de reconocimiento de lenguaje, lo cual hace más flexibles a los asistentes de voz como Siri y Alexa, además de que permite efectuar tareas cotidianas como la redacción de texto por medio del dictado si se le enseña a la IA a reconocer la voz.

En este sentido Andrew Ng (s.f) explica: “En lugar de solo reconocer gatos, he visto algoritmos de IA capaces de reconocer tipos específicos de flores, IA capaz de reconocer tipos específicos de alimentos y la capacidad de tomar como entrada una imagen y clasificarla en qué tipo de objeto, y esto está siendo utilizado en todas las aplicaciones” (Ng, en Coursera, s.f). Las otras aplicaciones que Ng menciona se encuentran en todo tipo de dispositivos y plataformas. Entre las aplicaciones que más resaltan son las API; por ejemplo, en los teléfonos inteligentes que buscan enfocar automáticamente el rostro de un ser humano al momento de tomar una fotografía.

Otro tipo de aplicación cotidiana son los procesadores de texto y aplicaciones de mensaje persona a persona. Como se ha mencionado arriba, el aprendizaje de máquina tiene la capacidad de aprender y conocer el lenguaje humano. En este sentido Ng (en Coursera explica: “La IA y el aprendizaje profundo específicamente también están progresando mucho en el procesamiento del lenguaje natural”; “El procesamiento del lenguaje natural o PNL se refiere a la IA que comprende el lenguaje natural, es decir, el lenguaje que usted y yo podríamos usar para comunicarnos” (Ng, en Coursera, s.f).

Hay que resaltar que existen más de un tipo de redes de reconocimiento de lenguaje. Por ejemplo, Andrew Ng explica “Un segundo tipo de procesamiento del lenguaje natural o PNL es la recuperación de información” (Ng, en Coursera, s.f). Asimismo, el profesor y fundador de la plataforma de Coursera menciona que el

ejemplo más claro de Procesamiento de lenguaje natural está en la búsqueda web, la cual “es quizás el ejemplo más conocido de recuperación de información en el que escribe la consulta de texto y desea que la IA lo ayude a encontrar documentos relevantes” (Ng, en Coursera, s.f).

A finales de 2022, *OpenAI* lanzó una versión de su plataforma de redes neuronales de lenguaje (un subcampo del procesamiento del lenguaje natural). La plataforma, conocida como ChatOpenAI, deriva del modelo *ChatGPT* y permite a los usuarios corregir textos, redactar ideas para ensayos desde cero, explicar estadísticas de manera sencilla, elaborar y corregir código de programación, entre otras funciones. A pesar de los sesgos y limitaciones de la plataforma, como la creación de referencias bibliográficas falsas, su interfaz es muy amigable, ya que es similar a conversar con texto con una persona.

De acuerdo con la pregunta —¿Puedes describirte en un párrafo para una tesis sobre IA y derechos humanos? — que se le hizo a la plataforma en 2023 (marzo, 2023):

“ChatGPT es un modelo de procesamiento del lenguaje natural desarrollado por OpenAI. Es una red neuronal de gran escala entrenada con una gran cantidad de texto de la web. A medida que recibe una pregunta o una serie de datos de entrada, genera una respuesta coherente y coherente. Puede realizar tareas como generación de texto, traducción automática, respuesta a preguntas y clasificación de texto. Aunque es una herramienta poderosa para la automatización de tareas y la generación de contenido, es importante tener en cuenta que, como con cualquier tecnología basada en datos, puede estar sujeta a sesgos y limitaciones, y es importante aplicar un enfoque ético en su desarrollo y uso, especialmente en temas relacionados con los derechos humanos” (Sic).

En cuanto a las aplicaciones en la ciencia, la Real Sociedad (2017), plantean de manera puntual que:

“El aprendizaje automático está estrechamente relacionado con los campos de la estadística y la ciencia de datos, que proporcionan una variedad de herramientas y métodos para el análisis de datos y la inferencia de datos. También está relacionado con la robótica y la automatización inteligente” (Royal Society, 2017), y “El aprendizaje automático también respalda el progreso en estos campos, como tecnología subyacente tanto para la IA como para la ciencia de datos” (Royal Society, 2017).

En el ámbito de la educación, la IA también es implementada. Por ejemplo, de acuerdo con la Real Sociedad: “El aprendizaje automático ya se utiliza en los sistemas de educación en línea; en los cursos masivos abiertos en línea (MOOC) se utiliza para analizar las aportaciones de los estudiantes, calificar pruebas u otras tareas basadas en computadora, y en algunas funciones de visión por computadora” (Royal Society, 2017). Esto agiliza muchas de las tareas que suelen ser repetitivas para los profesores, lo que permite que el personal docente se concentre en las labores pedagógicas más no en la gestión de tareas burocráticas y repetitivas ligadas a su función.

En el ámbito del transporte, la IA y el aprendizaje automático pueden crear vehículos no tripulados y autónomos utilizando diversas técnicas. Por ejemplo, el nivel más alto de automatización de trenes es conocido como Grade-of-Automation 4 (GoA4), el cual permite a los trenes moverse de estación a estación, abrir puertas, sortear obstáculos, entre otras actividades. Ahora bien, estos sistemas de transporte son altamente automatizados, pero no necesariamente usan aprendizaje de máquina para realizar sus operaciones. Por otro lado, los vehículos autónomos son capaces de aprender de su entorno a través de sensores y utilizar diferentes técnicas de aprendizaje profundo en modalidad de aprendizaje automático “*on edge*”; en este sentido, La *Royal Society* señala que “con la información adecuada, también sería posible evaluar los flujos de tráfico en tiempo real y realizar ajustes dinámicos para mejorar el flujo de tráfico” (Royal Society, 2017).

En medicina, un ejemplo claro de cómo se usan la IA y el aprendizaje automático en el diagnóstico de enfermedades. Esto se logra mediante la técnica de reconocimiento de imágenes, donde la máquina aprende patrones a partir de una gran base de datos, lo que le permite detectar ciertos padecimientos. La Royal Society explica que esto se puede ver en el caso del diagnóstico de cáncer de mama: “Un sistema de aprendizaje automático entrenado en imágenes de tejidos logró una mayor precisión que los patólogos al encontrar y utilizar características de la imagen que eran predictivas pero que no se habían utilizado previamente en las evaluaciones de patología” (Beck et al., 2011 en Royal Society, 2017).

Pero en aplicaciones de medicina más recientes, la IA ha resultado sumamente útil. En especial, la aplicación más importante que ha tenido la IA durante el año 2020 fue en el desarrollo de las vacunas contra el virus SARS-Cov-2 causante de la pandemia del Covid-19. De acuerdo con Dave Jhonson, del laboratorio que desarrolló la vacuna de Moderna (en Massachusetts Institute of Technology, 2021):

“Sabíamos que, si podía conseguir uno en el mercado, podía sacar cualquier número de ellos al mercado. Entonces, todas las decisiones que tomamos sobre cómo diseñamos la empresa y cómo diseñamos la infraestructura digital giraron en torno a esta noción de plataforma de que no vamos a construir esto para empezar: vamos a construir una solución que atienda esto. toda la plataforma. Creamos este primer equipo preclínico en el que podemos analizar bastantes de estos. Por eso creamos estos algoritmos para automatizar actividades. Cada vez que vemos algo en lo que sabemos esa escala y hacerlo paralelo va a mejorar las cosas, implementamos este proceso”.

En términos estrictos, el desarrollo de vacunas mediante la tecnología de ARN mensajero (mRNA) y el diagnóstico de enfermedades a través del reconocimiento de imágenes, ilustran las fortalezas y debilidades de la IA. Aunque la IA actualmente está limitada a realizar un conjunto de tareas específicas, una vez que aprende, como en el caso del aprendizaje automático profundo, puede realizar estas tareas con gran precisión, lo que reduce la carga de trabajo para los investigadores y médicos humanos, permitiéndoles enfocarse en tareas más creativas o en actividades donde la combinación de intuición y experiencia es esencial.

La IA y el aprendizaje automático profundo también están penetrando en áreas como el derecho, lo que demuestra que los ejemplos de automatización ya no están limitados exclusivamente a áreas de trabajo no calificado. La IA está invadiendo sectores donde el personal tiene un nivel de educación que va desde licenciatura hasta especializaciones.

Una de las aplicaciones más evidentes de la IA se encuentra en el proceso de búsqueda de expedientes. Antes, estas tareas monótonas, pero requerían de conocimientos técnicos, jurídicos y experiencia, eran asignadas a abogados jóvenes o recién graduados. Sin embargo, gracias a la capacidad de la IA y el aprendizaje automático de buscar y catalogar información en bases de datos, el

trabajo de búsqueda de expedientes se ha convertido en una de las primeras áreas automatizadas por esta tecnología. Un ejemplo de esto es Blue J Legal, una empresa con sede en Toronto que utiliza software para analizar fallos de tribunales canadienses sobre impuestos y empleo, brindando a sus clientes una lista de casos similares, citas relevantes y una evaluación de las probabilidades de ganar o perder su caso (Dittmar y Wu en Coursera, s.f).

Una de las consecuencias de la implementación de la IA es la alteración de las funciones de los abogados. Esto quiere decir, que un abogado debe desarrollar nuevas capacidades conforme la IA penetra en su área de trabajo en aras de mantenerse competitivo o relevante en su tarea de litigante. De acuerdo con Dittmar y Wu en Coursera (s.f):

“Hoy en día, un abogado que no sabe lo que es administrar una base de datos, clasificar información o diseñar un proceso agrega muy poco valor a sus clientes. Por supuesto, ese abogado resuelve lo cotidiano. Pero en la medida en que las empresas se empiezan a sofisticar y a requerir servicios más integrales, esas capacidades se vuelven muy limitadas”.

Teniendo en cuenta que la IA está cambiando las funciones del abogado, es importante examinar su papel en el gobierno. La penetración de la IA, especialmente el aprendizaje automático, en las funciones de gestión, suministro y administración de servicios públicos, ha dado lugar a lo que se conoce como gobierno electrónico o e-gobierno. El término e-gobierno es una expresión amplia que abarca todos los aspectos de la gobernanza en los que la IA tiene un papel activo, como la gestión de documentos, servicios asistidos por IA, transporte público, suministro de servicios básicos y justicia.

De acuerdo con la Organización para la Cooperación y el Desarrollo Económico (OCDE, 2003) en Villoria y Ramírez (2013), el gobierno electrónico se refiere a la operación del gobierno a través de las tecnologías de información y comunicación, especialmente Internet. Esto incluye servicios como pagos, solicitudes de servicios y denuncias a través de plataformas en línea. Sin embargo, con el aumento del aprendizaje automático profundo y la ciencia de datos asociada, el término de e-gobierno debería redefinirse para abarcar la automatización y la toma de decisiones basadas en datos.

Una propuesta para redefinir el concepto de gobierno electrónico es la de la de Saeed y Al-Mushayt (2020). Los académicos proponen que el gobierno electrónico es: “es la aplicación del empleo de técnicas electrónicas avanzadas y servicios web para presentar, intercambiar y promover los servicios del gobierno para ciudadanos y empresas con el objetivo de mejorar la productividad y reducir el costo” (Saeed y Al-Mushayt, 2020). En este sentido, también sugieren el uso de la IA como una herramienta para auxiliar en las funciones de ese tipo de gobierno; pero advierten:

“La IA ha avanzado recientemente los resultados de vanguardia en un número cada vez mayor de dominios. Sin embargo, aun enfrenta varios desafíos que dificultan su implementación en las aplicaciones de gobierno electrónico, tanto para mejorar los sistemas de gobierno electrónico como para las interacciones entre el gobierno electrónico y los ciudadanos” (Saeed y Al-Mushayt, 2020).

Ahora bien, para implementar las tecnologías de IA en los servicios gubernamentales, es esencial que la población tenga acceso a internet, ya que muchas herramientas de aprendizaje profundo requieren plataformas conectadas a la red. Según Pérez, Camacho, Mena y Arroyo (2016), los principales desafíos para que la gente realice trámites y servicios gubernamentales en línea (e-gobierno) son la creación de un marco regulatorio sólido y la reducción de la brecha digital. Además, es importante impulsar la cultura informática y ofrecer servicios en todos los dispositivos y plataformas utilizados por la población.

Otra forma en que se puede utilizar la IA en el gobierno es en programas sociales focalizados para identificar a las personas en situación de riesgo o precariedad que pueden beneficiarse de apoyos gubernamentales específicos. Según la Royal Society (2017), si se identifica tempranamente a jóvenes en alto riesgo de abandonar la educación o no encontrar empleo, se pueden realizar intervenciones para evitar que estas personas se desvíen del sistema educativo o de formación.

Por otro lado, en el ámbito de los servicios gubernamentales, los sistemas de aprendizaje profundo pueden ayudar a comprender el rendimiento de algún servicio gubernamental con el fin de encontrar oportunidades para mejorar su eficiencia (Al-Mushayt, 2019). Según Araya (2019): “Basándose en herramientas de aprendizaje profundo, la IA podría mejorar significativamente la prestación de

servicios a los ciudadanos y elevar el trabajo de los profesionales del servicio público, al mismo tiempo que inspira a una nueva generación de tecnócratas a ingresar al gobierno". En este sentido, las herramientas como el sistema de aprendizaje profundo pueden hacer obsoletas las tareas de recopilación y trámites de documentos que normalmente requieren interacción cara a cara.

Como consecuencia de la implementación de la IA en el gobierno, han surgido otras variables como la justicia electrónica. El término justicia electrónica se refiere a la participación de las tecnologías de la información y comunicación en los procesos jurídicos, incluyendo la asistencia de herramientas de IA en este proceso. De esta manera, la IA puede ayudar a localizar expedientes, agilizar el proceso e incluso participar en la dictaminación de sentencias.

Al igual que el término comercio electrónico, el término gobierno electrónico nació del boom de Internet. No se limita al uso de Internet o sistemas de acceso público para uso directo de clientes o ciudadanos. e-Gov comenzó como un campo de práctica, básicamente convocando a los profesionales que luchan por enfrentar los nuevos desafíos del medio de Internet mediante la implementación de nuevos sistemas de manera creativa (Grönlund y Horan, 2004).

Según Grönlund y Horan (2004), el gobierno electrónico surgió para abordar los desafíos y aprovechar las oportunidades que presentaba internet. Entre estos desafíos estaba la necesidad de aprovechar los recursos tempranos de internet, y entre las oportunidades se podía crear plataformas y recursos para agilizar las funciones del gobierno, desde el nivel usuario a gobierno, gobierno a empresa o gobierno. Con el avance de la IA, el gobierno electrónico ha incorporado esta tecnología, como se puede ver en ejemplos como el reconocimiento facial, que se utiliza en sistemas de vigilancia para garantizar la seguridad de los ciudadanos en distintas urbes y espacios.

Otro de los usos de la IA en el formato de e-gobierno es la toma de decisiones. Según Zuiderwijk et al (2021), "los beneficios del uso de la IA en el gobierno tienen que ver con los beneficios de la toma de decisiones. El aprendizaje automático podría ayudar a los tomadores de decisiones gubernamentales (Alexopoulos et al., 2019) y conducir a una toma de decisiones mejor y más precisa (Ben Rjab y Mellouli, 2019)" (Zuiderwijk et al., 2021). En este sentido, la IA puede ayudar a

darle sentido a la gran cantidad de datos con los que trabaja el gobierno, para después ofrecer sugerencias sobre un problema específico.

Esto se puede ver en otras funciones de la IA en el gobierno, como el transporte público y el tránsito. En el caso de la Ciudad de México, Alphabet inició en 2021 un proyecto para mostrar rutas del transporte público de esta entidad a través de su plataforma Google Maps (Olivares, 2021). Esto sigue el proyecto iniciado en 2015 de mostrar algunas rutas del transporte público de la capital mexicana en *Google Transit*.

El caso de Google Maps resulta interesante en dos sentidos. Es un servicio gratuito ofrecido a través de una plataforma privada. Por otro lado, este servicio requiere de una gran cantidad de datos recolectados del sistema de transporte público en aras de que pueda funcionar; por consiguiente, se aprecia una relación cuasi horizontal en donde una entidad como el sistema de transporte y sus usuarios proveen datos y una empresa datos facilita un servicio o una API⁶ a partir de estos datos con los que opera su IA.

También hay que señalar que la información recolectada por usuarios y gobiernos puede servir para que la IA asista a la planeación de mejores sistemas de tránsito. Esta es la meta actual de muchas compañías como *DiDi transit* (Morton, 2021), que buscan hacer alianzas con gobiernos locales en aras de ofrecerles su tecnología para mejorar el transporte público y movilidad a cambio de datos relacionados con el uso de estos sistemas de transporte. Así pues, la IA facilita la solución a problemas logísticos que requieren de un gran capital humano.

Teniendo en cuenta que la IA puede navegar entre una gran cantidad de datos, se mencionó su función para asistir en las tareas monótonas de búsqueda de documentos. Esto se puede ver en los mecanismos diseñados para encontrar o catalogar archivos. Además, junto con otras herramientas de IA como el reconocimiento de voz y de imagen, ayudan a hacer trámites burocráticos que de otra manera requerirían una gran cantidad de tiempo en espera, traslado o obstáculos burocráticos.

⁶ API se refiere a Interfaz de programación de aplicaciones (Education IBM, 2021).

De acuerdo con Zuiderwijk et al (2021), "El uso de la IA en el gobierno podría allanar el camino para una mejor interacción y comunicación entre el gobierno y los ciudadanos (Androutsopoulou et al., 2019)". En este sentido, los autores se refieren a que la IA puede asistir en el trato con los ciudadanos. Este tipo de aplicaciones ya se ven en distintos niveles, como los Bots, o asistentes de dudas usando aplicaciones de mensaje de texto, o servidores públicos virtuales más complejos que emplean una interfaz gráfica con la que los ciudadanos se pueden comprometer.

Uno de los asistentes virtuales que actualmente está en uso es Emma:

"Emma, utilizada por los Servicios de Ciudadanía e Inmigración de EE. UU. Para ayudar a los visitantes del sitio web a obtener respuestas y encontrar información en inglés o español; y la Señora Landingham, un chatbot de la Administración de Servicios Generales de EE. UU. que trabaja con la aplicación Slack y guía la incorporación de nuevos empleados, dice Meisel" (Desmond, 2019).

En resumen, la IA en el gobierno permite reducir tareas, un mejor trato a los ciudadanos, reducir tiempos en trámites burocráticos, mejorar el transporte público y el tránsito, administrar recursos para generar un gobierno más sustentable, reconocer el estado de ánimo de la ciudadanía y ayudar en tareas de seguridad y vigilancia. Al igual que la IA, muchas de estas tareas están en perfección continua. Cada tarea en la que se puede usar la IA en el gobierno está ligada a desafíos, especialmente en torno a los derechos humanos.

E. Futuras aplicaciones

Antes de abordar los problemas éticos en torno a la IA, es importante destacar las próximas aplicaciones de esta tecnología. En lugar de enfocarnos en cuestiones futuristas, esta sección se centrará en aplicaciones que se encuentran en una etapa muy avanzada de desarrollo o incluso en un escenario temprano de implementación que aún no ha llegado a la mayoría de la población.

De acuerdo con Raso "los sistemas basados en IA ya están superando a los especialistas médicos en el diagnóstico de ciertas enfermedades, mientras que el uso de IA en el sistema financiero está ampliando el acceso al crédito a los

prestatarios que alguna vez fueron rechazados" (Raso, Hilligoss, Krishnamurthy, Bavitz y Kim, 2018); no obstante, esta afirmación puede ser algo optimista ya que en cualquier momento se puede resquebrajar el ciclo de entusiasmo con respecto a la IA. En este sentido, la IA acotada está demostrando grandes avances, aun estando lejos de la visión de IA generalizada o fuerte propia de la ciencia ficción. Con base en esto, es necesario explicar de forma breve los campos en los que esta tecnología operará con mayor incidencia.

La medicina es una de las áreas más prometedoras para la aplicación de la IA, especialmente el aprendizaje automático. Esta tecnología puede asistir en el diagnóstico y el tratamiento de diversas enfermedades, como el cáncer, mediante el análisis e interpretación de imágenes médicas, como los rayos X. El aprendizaje automático se basa en entrenar a una máquina con un conjunto de datos e imágenes previamente etiquetados, para luego hacerla capaz de reconocer y diagnosticar casos nuevos a partir de imágenes desconocidas. Si la tecnología sigue avanzando y perfeccionándose, la medicina podría contar con una herramienta sin precedentes para mejorar la salud y la calidad de vida de las personas.

De acuerdo con Ertel (2019), "Los diagnósticos en casi todas las áreas de la medicina y la tecnología ahora se realizan mejor, más rápido y con mayor eficacia por computadora". Por otro lado, "...en medicina nuclear, la red neuronal artificial (ANN) es la columna vertebral de ML y DL. Las entradas pueden ser características radiómicas que se han extraído de los archivos de imagen o, si se utiliza una red neuronal convolucional (CNN), pueden ser las propias imágenes" (Currie & Rohren, 2021). En este sentido, el aprendizaje automático profundo promete ser una herramienta eficiente para resolver problemas, asistir en diagnósticos, y reducir costos.

Así pues, la tecnología ligada al aprendizaje automático no solo aprende de imágenes de enfermedades, sino que también genera diagnósticos a partir de las mismas. En primer lugar, se puede ahorrar mucho tiempo para detectar una enfermedad, especialmente en enfermedades no crónicas degenerativas. En

segundo lugar, el médico puede dedicar más tiempo al tratamiento en lugar del diagnóstico, además de contar con una herramienta importante que asista en este proceso.

Otra aplicación importante dentro de la medicina puede ser el desarrollo de tratamientos y fármacos. Dado que el aprendizaje automático no solo se limita a aprender a interpretar datos, sino también a proponer soluciones, esta tecnología puede servir para desarrollar nuevas medicinas, sugerir tratamientos para enfermedades y diseñar instrumentos para estos tratamientos. Por ejemplo, la IA puede utilizarse para diseñar un nuevo corazón artificial o ayudar en el diseño de éste, una nueva droga contra una enfermedad crónica degenerativa, y tratamientos específicos para enfermedades raras o con enormes variables individuales.

La IA puede servir para diagnosticar posibles enfermedades con base en la historia del paciente. El aprendizaje automático puede recolectar muchos datos, como los antecedentes clínicos de la persona y su familia, interpretar el ADN de este individuo, otros factores de riesgo como la edad, el medio ambiente y la alimentación, y así generar un plan para los síntomas presentes y los futuros. Además, si el paciente comparte constantemente sus datos relacionados con hábitos de sueño, alimentación, entre otros, estos pueden ser utilizados por el médico para sugerir nuevas rutinas y nuevos hábitos con el fin de evitar enfermedades y complicaciones relacionadas con las mismas.

En el área de transporte, la IA puede construir y diseñar sistemas de vehículos completamente automatizados o más avanzados en automatización. El proyecto más ambicioso es desarrollar e implementar automóviles autónomos que transporten a una persona sin conductor humano, desde un punto A hasta un punto B. Muchos de estos vehículos ya están en construcción y pruebas, pero enfrentan desafíos por la cantidad de datos y la capacidad de procesamiento necesarios para operar.

La Real Sociedad plantea que el aprendizaje automático admite una variedad de sistemas de uso común, pero su potencial va más allá. El aprendizaje de máquina

tiene el potencial de desarrollarse en áreas como el transporte, la salud, la educación a distancia y presencial, y en algunos casos, incluso en el gobierno. Esto se logra mediante la utilización de complejos algoritmos que dan vida a las redes neuronales profundas y hacen funcionar los dispositivos.

La afirmación es reforzada por van Dijck et al (2018) cuando mencionan que, aunque no está seguro de que los vehículos autónomos se apoderen de las carreteras en el corto plazo, muchas plataformas de transporte colectivo parecen estar apostando por ese futuro. Los autores se refieren a sistemas de transporte que operan bajo el modelo GoA 4, que son completamente automatizados y no requieren la asistencia de un humano. Aunque quizás los sistemas de transporte se sofisticen aún más utilizando el aprendizaje de máquina en la forma de vehículos de punto a punto no tripulados.

Otra de las aplicaciones de la IA es en el sistema de justicia; por ejemplo, en las etapas previas al proceso jurídico y la evaluación de daños.

“En busca de equidad y eficiencia, los sistemas de justicia están empleando cada vez más herramientas automatizadas para la toma de decisiones en todas las etapas procesales...la última generación de herramientas de evaluación de riesgos aprovecha las técnicas de aprendizaje automático para reequilibrar continuamente los factores de riesgo en respuesta a nuevos insumos” (Raso et al., 2018).

La IA se está convirtiendo en una parte de varios aspectos de la vida humana. Aunque las personas pueden no darse cuenta de su presencia, esta tecnología está cada vez más presente en áreas como la salud, el transporte y la justicia. Por lo tanto, es esencial considerar los riesgos y limitaciones asociados tanto con las aplicaciones actuales como futuras de la IA para poder entender mejor su impacto en nuestras vidas y seguir haciendo explicaciones accesibles sobre los conceptos técnicos relacionados con la IA, el aprendizaje automático y la ciencia de datos.

F. La caja negra

Quizás uno de los problemas más complejos del aprendizaje de máquina es la llamada caja negra. Es decir, los secretos de operación y funcionamiento de muchas aplicaciones y algoritmos a los que está relacionada este tipo de

tecnología. Las cajas negras son: “Una de las limitaciones de la IA es que es difícil de explicar y muchos sistemas de IA de alto rendimiento son cajas negras. Lo que significa que funciona muy bien, pero la IA no sabe cómo explicar por qué hace lo que hace” (Ng, en Coursera, s.f). Otra definición de la caja negra tiene que ver con la propiedad intelectual: “El modelo en sí es una caja negra, su contenido es un secreto corporativo ferozmente guardado” (O’Neill, 2016).

Ahora bien, el problema de la caja negra en la IA se refiere a la dificultad o imposibilidad de comprender cómo y por qué un algoritmo de IA produce una determinada salida o decisión a partir de una entrada o conjunto de datos (Bathae, 2018). Este problema afecta tanto a los diseñadores como a los usuarios de los sistemas de IA, ya que limita su capacidad de verificar, explicar, controlar y responsabilizar a los algoritmos de IA. El problema de la caja negra se agrava cuando los algoritmos de IA son complejos, dinámicos y opacos, como ocurre con el aprendizaje profundo o las redes neuronales artificiales (Chen et al., 2022); así pues, el problema de la caja negra plantea desafíos legales, éticos y sociales, especialmente cuando los algoritmos de IA se aplican a ámbitos sensibles como la medicina, la justicia o la seguridad.

De acuerdo con O’Neill, las cajas negras generan un enorme problema de responsabilidad. Es decir, si alguna herramienta de IA provoca alguna falla jurídica, ética y administrativa, es muy difícil encontrar o discernir quién es el actor responsable. Así pues, “...si los detalles están ocultos, también es más difícil cuestionar la puntuación o protestar contra ella” (O’Neill, 2016).

Asimismo, O’Neill explica tajantemente que “sobre todo, los programas emiten veredictos inquebrantables, y los seres humanos que los emplean solo pueden encogerse de hombros, como diciendo: “¿Qué puedes hacer?”. (O’Neill, 2016). Por consiguiente, el problema de la caja negra es que impide rastrear responsabilidades o encontrar el por qué y cómo llegaron a tomar algunas decisiones. Y hay que resaltar que esas decisiones a menudo afectan la vida de los actores humanos.

La falta de entendimiento provoca fenómenos como el sesgo de la IA. Esto se debe a dos problemas: la caja negra como secreto comercial y la caja negra como producto de la complejidad de la IA. En la segunda vertiente, el hecho de que no se sabe a menudo cómo una herramienta toma una decisión demuestra un problema subyacente más profundo: la falta de explicabilidad. Según la Real Sociedad:

Este enfoque para procesar datos significa que la información a través de la red neuronal está muy dispersa, con patrones complicados de fuerza de conexión entre unidades y potencialmente con muchos miles de capas de unidades. El resultado es el llamado problema de la *caja negra*: estos sistemas pueden generar resultados muy precisos, pero es difícil explicar por qué se ha obtenido un resultado (Royal Society, 2017).

Por ello, el problema de las cajas negras radica en la capacidad del aprendizaje automático. Considerando que la IA puede generar funciones a partir de otras, nuevos elementos en su proceso de aprendizaje, se complica comprender cómo y por qué toma ciertas decisiones. Según Yu y Ali (2019): "Por esta razón, las redes neuronales se describen comúnmente como una caja negra: sistemas cerrados que reciben una entrada, producen una salida y no dan ni idea de por qué".

Asimismo, los autores comentan que los riesgos y temores relacionados con las nuevas capacidades de la IA yacen en el problema en torno a la forma de cómo funcionan estas cajas negras y el uso malicioso que puede generar:

"El miedo a los sistemas de IA como cajas negras, así como las preocupaciones con respecto a las posibles implementaciones de la tecnología de IA en armas automatizadas y ataques cibernéticos, han provocado llamadas para garantizar una mayor confiabilidad, transparencia y alineación ética de la IA" (Yu & Ali, 2019).

El problema de las cajas negras también está ligado a riesgos de privacidad. A consecuencia de ello, tal como se mencionó, la gran problemática en torno a la responsabilidad que ocasiona el fenómeno de la caja negra es importante hablar de los riesgos a la privacidad ligada a esta modalidad operativa. Además, en este mismo sentido, es importante hablar de la privacidad en torno a la IA como un problema ético particular.

Ahora bien, antes de ello, es necesario plantear algunas soluciones en torno a la caja negra. En este sentido, Raso et al., (2018) preguntan: "¿Qué pruebas y análisis de expertos serían necesarios para abrir la caja negra del algoritmo, especialmente cuando está protegido por secretos comerciales y leyes de propiedad intelectual?". Una posible respuesta es generar legislaciones que favorezcan y propicien el desarrollo de herramientas de IA que tengan la capacidad de explicarse a sí mismas.

Dentro de las soluciones jurídicas, también resalta la de hacer auditorías. Yu y Ali (2019) exponen que "Los expertos han propuesto la idea de un servicio de calificación de auditoría para validar y certificar la calidad y precisión de los sistemas de IA". Y en cuanto a los modelos que faciliten mecanismos para que la IA se explique a sí misma, Andrew Ng explicó en su curso que: "si un sistema de IA no funciona, su capacidad para explicarse también nos ayudaría a descubrir cómo entrar y hacer que el sistema de IA funcione mejor" (Ng en Coursera, s.f).

Otras propuestas están relacionadas en entrenar o permitir que la IA aprenda de datos que sean comprensibles para los seres humanos. Por ejemplo, emplear imágenes que los seres humanos puedan distinguir y comprender. En este sentido:

“un conjunto suficientemente grande de imágenes de caras etiquetadas con la identificación de las personas cuyas caras aparecen en las imágenes, se puede entrenar una red neuronal para extraer un descriptor de vector de características de caras, y los algoritmos de clasificación pueden trabajar con los vectores como puntos en un espacio de alta dimensión” (Tan y Lim, 2018).

El problema de las cajas negras en la IA se refiere a la falta de comprensión y explicabilidad de cómo y por qué un algoritmo toma ciertas decisiones. Esto puede causar problemas éticos y legales en cuanto a responsabilidad y privacidad. Además, esto también puede conducir a sesgos en las decisiones tomadas por la IA. Para abordar este problema, se han propuesto soluciones como la generación de legislaciones que promuevan herramientas de IA capaces de explicarse a sí mismas y también la realización de auditorías de calidad y precisión de los sistemas de IA; por lo tanto, es importante investigar y desarrollar mecanismos

para afrontar estas preocupaciones y garantizar la confiabilidad y transparencia en el uso de la IA.

2. Los falsos profundos y los algoritmos adversarios

Las herramientas de la IA y aprendizaje de máquina son falibles. Esto es, están sujetas a ciertas técnicas y herramientas que pueden arruinar su eficiencia. Además, dado su complejidad y capacidad, crean nuevos instrumentos capaces de afectar distintos espacios de la sociedad como lo es suplantación de identidad más allá de simples datos; esto es, a partir de la suplantación de imagen y voz de una persona en lo que se conoce como falsos profundos o *deep fakes*.

Los *deepfakes* son videos hiperrealistas creados con inteligencia artificial que imitan la voz, expresiones y gestos de una persona para hacer que parezca decir o hacer algo que nunca ocurrió (Westerlund, 2019). Utilizan redes neuronales y mapeo facial para intercambiar rostros en videos, lo que los hace difíciles de detectar y fáciles de viralizar en redes sociales (Westerlund, 2019). Esto puede generar problemas grandes como fraudes, suplantación de identidad, y estafas.

Ahora bien, dentro de estos dos fenómenos es importante comenzar con los ataques a los que está susceptible la IA. Esto es, debido a que, como muchas herramientas tecnológicas, los seres humanos tienden a depositar su confianza en las decisiones generadas por una máquina. Por ende, es necesario hablar y explicar el ataque más común y peligroso que puede sufrir una herramienta de aprendizaje automático: los algoritmos adversarios y la forma en la que estos funcionan.

Los ataques adversarios son los mecanismos con los que se atacan a los algoritmos de la IA utilizando otros algoritmos o técnicas que pueden engañarlos. Según Andre Ng (en Coursera, s.f): "La tecnología de IA también es susceptible a ataques adversarios, lo que significa, por ejemplo, que nuestros filtros de correo no deseado pueden ser engañados para que el sistema de correo electrónico deje pasar mensajes indeseados". Estos ataques adversarios engañan a las

herramientas del aprendizaje automático con el objetivo de afectar sus resultados o generar resultados distintos.

En este sentido, se ha demostrado que los algoritmos de aprendizaje profundo son susceptibles a ataques a través de ejemplos adversarios. Los adversarios utilizan datos modificados de una manera no reconocible para que los humanos interfieran drásticamente en el comportamiento de los modelos de aprendizaje profundo (Schuchmann, 2019). Con base en esto, es importante destacar que existen dos tipos de ataques adversarios. Según Liu & Haohui (2020), los ataques adversarios se clasifican en dos categorías: ataques dirigidos y ataques no dirigidos.

Haohui, con base en la conferencia de Liu titulada *Adversarial Attacks in Machine Learning and How to Defend Against Them*, explica:

“El ataque dirigido tiene una clase de destino, Y , que quiere que el modelo de destino, M , clasifique la imagen I de la clase X como. El objetivo del ataque dirigido es que $-M-$ clasifique erróneamente al predecir el ejemplo adversario, I , como la clase objetivo–deseado y en lugar de la clase X . Por otro lado, el ataque no dirigido no tiene una clase objetivo que quiera que el modelo clasifique la imagen como. En cambio, el objetivo es simplemente hacer que el modelo objetivo se clasifique erróneamente al predecir el ejemplo adversario, I , como clase, distinta de la clase original, X ” (Liu & Haohui, 2020).

Los ataques adversarios son mecanismos con los que se le hace frente a los algoritmos de IA mediante otras técnicas para engañarlos. Según Andrew Ng en Coursera (s.f), "un ataque contra un sistema seguro significa un intento de hacer que haga algo diferente a lo que se pretendía hacer". Los sistemas de aprendizaje profundo son vulnerables a estos ataques porque los adversarios usan datos modificados para interferir drásticamente en el comportamiento de los modelos.

Existen dos tipos de ataques adversarios: dirigidos y no dirigidos. El objetivo de un ataque adversario es engañar a la máquina, lo que pone de manifiesto las ventajas y desventajas de la inteligencia actual. Andrew Ng proporciona ejemplos de cómo se llevan a cabo estos ataques mediante estudios empíricos como colocar pegatinas para engañar a un sistema de IA para que no vea una señal de tráfico. La forma de generar un ataque adversario está basada en comprender

cómo funcionan las herramientas de aprendizaje automático a las que se desea atacar.

Así pues, los ataques adversarios demuestran la fragilidad del aprendizaje profundo de máquina. Por lo tanto, este tipo de ataques exponen las debilidades de la tecnología y crean un ambiente de monitoreo constante y herramientas para combatirlos. Según Schuchmann (2019), "hay varios ejemplos de por qué tal descripción sería precisa, incluidos los ataques adversarios, la falta de capacidad para generalizar y la falta de datos. Se prosigue una discusión detallada de estas fallas y eventuales mecanismos de prevención".

Sin embargo, el problema de los ataques adversarios también tiene otras implicaciones. Según Andrew Ng (en Coursera, sf), "...muchos sistemas de IA están tomando decisiones económicamente importantes y algunos sistemas de IA están abiertos a ataques adversarios, si alguien más está deliberadamente tratando de engañar a su sistema de IA" (Ng, en Coursera, sf). En este sentido, la posibilidad de los ataques adversarios desnuda la perfecta fiabilidad de las herramientas del aprendizaje de máquina, así como las áreas de oportunidades para cubrirlas y el rol interdisciplinario que deben tener distintas empresas, organizaciones y gobiernos para mitigarlas.

Andrew Ng (en Coursera, s.f) menciona que: "...no deberíamos ser demasiado optimistas sobre las tecnologías de IA y tener una visión poco realista de las tecnologías de IA puede hacer que la gente piense que la sensibilidad o la superinteligencia, la IA y general llegará pronto...". Esto es debido a que muchas personas piensan que la IA es la IA general que suelen ver en las películas como *Terminator*. Por otro lado, existen personas –muchas personas– que usan las herramientas de la IA sin darse cuenta de ello y sin detenerse a pensar en las posibles implicaciones y riesgos que ésta encierra.

Otro punto central en torno a los ataques a los adversarios es su relación con las cajas negras de la IA. Esto ya se tocó brevemente, pero hay que aunar más para comprender cómo desarrollan y desdoblan los complejos problemas de la IA y, en especial, del aprendizaje de máquina profundo. Así pues, "...los ataques

adversarios demuestran un problema subyacente que es más profundo: la falta de explicabilidad. Debido a la naturaleza de caja negra de los modelos de aprendizaje profundo, predecir lo que está haciendo la red es difícil” (Schuchmann, 2019).

Además, los *deep fakes* son un problema importante relacionado con el aprendizaje automático. Aunque la capacidad de falsificar y suplantar identidad siempre ha sido una parte de la sociedad, las herramientas de aprendizaje automático han aumentado el nivel de complejidad e intangibilidad. Antes, una persona podía falsificar una firma, pero ahora puede falsificar una voz, un rostro o una imagen en vivo. Según Andrew Ng, la tecnología de IA se ha utilizado para crear deep fakes, como el video de BuzzFeed en el que se hace que el expresidente Obama diga cosas que nunca dijo. Otros ejemplos de deep fakes incluyen la tecnología de redes neuronales de lenguaje, que permite imitar la voz de una persona.

Nguyen et al. (2020) señalan que los algoritmos pueden crear imágenes y videos falsos que los humanos no pueden distinguir de los auténticos. Asimismo, es necesario desarrollar tecnologías que puedan detectar y evaluar automáticamente la integridad de los medios visuales digitales. Sin embargo, los académicos también proponen usos positivos para los deep fakes, como la generación de prótesis de voz para personas que han perdido la suya. De esta forma la tecnología tiene tanto riesgos como beneficios.

Aunque el objetivo de estas tecnologías de IA no es malicioso, existe el peligro constante de que sean utilizadas con fines disruptivos o para beneficios personales. Esto plantea preguntas sobre cómo regular y controlar estas tecnologías, así como la necesidad de educar a la población sobre cómo protegerse y detectar ataques adversarios y falsos profundos. Para enfrentar ataques adversarios, se pueden utilizar algoritmos de seguridad. Según el profesor Andre Ng, "las defensas tienden a ser muy técnicas, pero hay formas de modificar las redes neuronales y otros sistemas de IA para hacerlos más difíciles de atacar"

(Ng, en Coursera, s.f): una de las soluciones técnicas es entrenar a la IA para que se defienda de estos ataques.

Ren et al. (2020) explican que el entrenamiento adversario es un método de defensa intuitivo contra muestras adversarias, que busca mejorar la solidez de una red neuronal entrenando con muestras adversarias. Además, los investigadores sugieren perturbar el ruido o por medio de ruido (errores o funciones no precisas) aleatorias. Otras soluciones técnicas incluyen redes neuronales regenerativas, por ello los académicos Ren et al. (2020) exponen que:

La eliminación de ruidos es un método muy sencillo en términos de mitigar las perturbaciones / efectos adversos. Trabajos anteriores señalan dos direcciones para diseñar una defensa de este tipo: eliminación de ruido de entrada y eliminación de ruido de características. La primera dirección intenta eliminar parcial o totalmente las perturbaciones adversas de las entradas, y la segunda dirección intenta aliviar los efectos de las perturbaciones adversas en las características de alto nivel aprendidas por los DNN.

La técnica de generación de ruido aleatorio se refiere a la adición de perturbaciones a las redes neuronales mediante el uso de una técnica llamada autoensamblaje al azar. Esta técnica agrega una capa de ruido antes de cada capa de convolución en las fases de entrenamiento y prueba, y agrupa los resultados de la predicción sobre los ruidos aleatorios para estabilizar las salidas de las redes neuronales profundas (DNN). De esta manera, se crea ruido dentro de la misma plataforma para evitar ser atacada por una herramienta que intenta generar ruido.

La defensa contra el fenómeno de los falsos profundos se realiza con herramientas de IA. Por ejemplo, se puede entrenar una IA para identificar contenido o fenómenos falsos creados por otra IA. Por ejemplo, una herramienta de aprendizaje automático puede examinar la transmisión de un video en vivo para determinar si es generado a partir de herramientas como redes neuronales profundas o si se trata de contenido orgánico, donde el actor humano es quien dice ser.

De acuerdo con Nguyen et al (2020): “Hsu et al. propusieron un método de aprendizaje profundo de dos fases para detectar imágenes *deep fake*. La primera

fase consiste en un extractor de características basado en una red común de características falsas" (Nguyen et al, 2020). En otras palabras, la técnica para reconocer *deep fake* es entrenar a la IA con imágenes para que aprenda la diferencia entre una verdadera y una falsa, y luego superponer las imágenes y distinguir entre ellas. Por lo tanto, las herramientas de aprendizaje profundo son esenciales para aprender a identificar lo real de lo falso.

Sin embargo, las herramientas de aprendizaje profundo no son las únicas para combatir el problema de los *deep fake*. Según Nguyen et al (2020), "otra dirección de investigación es integrar métodos de detección en plataformas de distribución, como redes sociales, para aumentar su efectividad en hacer frente al impacto generalizado de las falsificaciones profundas". Es decir, se pueden integrar herramientas dentro de las plataformas para evitar las falsificaciones profundas antes de que se difundan y resulten dañinas.

El problema de los *deep fake* es relevante debido a la naturaleza de internet y el mundo hiperconectado en el que vivimos. Las falsificaciones profundas pueden generar problemas que van desde la suplantación de identidad hasta la generación de evidencias falsas para acusar a alguien de un delito, y debilitar la confianza de las personas en las fuentes periodísticas. Además, el fenómeno de los *deep fake* plantea la necesidad de crear mecanismos que aborden sus repercusiones sociales, que a menudo requieren una respuesta institucional y jurídica, como lo plantean Nguyen et al (2020):

Los *deepfakes* han comenzado a erosionar la confianza de las personas en los contenidos de los medios, ya que verlos ya no es proporcional a creer en ellos. Podrían causar angustia y efectos negativos a los destinatarios, aumentar la desinformación y el discurso de odio, e incluso podrían estimular la tensión política, inflamar al público, la violencia o la guerra. Esto es especialmente crítico hoy en día, ya que las tecnologías para crear *deep fakes* son cada vez más accesibles y las plataformas de redes sociales pueden difundir esos contenidos falsos rápidamente.

En conclusión, tanto los algoritmos adversarios como los *deep fakes* son problemas relacionados con el aprendizaje profundo de máquinas. Estos problemas requieren técnicas y herramientas específicas para mitigar o combatir

sus consecuencias. Es necesario que tanto empresas privadas, como gobiernos, instituciones, organizaciones y académicos de las ciencias sociales y humanidades, estén conscientes de las implicaciones en la vida cotidiana de las personas y sobre todo, cómo crear mecanismos para minimizar estas consecuencias negativas.

3. El aprendizaje de máquina y el trabajo

Es fundamental comprender el impacto que los sistemas de inteligencia artificial tienen en el ámbito laboral. Conocer estos cambios y su ritmo de evolución permitirá garantizar que sus aplicaciones se implementen de manera equitativa y beneficien a toda la población (Xu, David, Kim, 2018, p. 90). Este aspecto cobra especial relevancia con el auge de la economía de plataformas digitales o la economía gig, donde cada vez se integran más sistemas de inteligencia artificial en el entorno laboral, transformando la forma en que se organizan y ejecutan las tareas.

De acuerdo con Webb (2020), las herramientas relacionadas con la IA, especialmente el sistema de aprendizaje profundo, están generando un alto grado de ansiedad debido al reemplazo laboral causado por la IA. Este tipo de ansiedad, aunque no es nueva y ha ocurrido en otros espacios como la revuelta campesina del siglo XVII, se puede explicar por el tipo de puestos de trabajo que están en riesgo. Si antes los procesos de automatización relacionados con las primeras tres revoluciones industriales afectaban principalmente a funciones laborales altamente repetitivas, la IA y sus herramientas como el SAP ponen en riesgo funciones y puestos de trabajo que requieren habilidades más humanas como el aprendizaje, la interpretación de información y la ejecución de tareas basadas en lo que se ha aprendido.

Morgan (2018) explica que "...la razón por la cual el aprendizaje profundo está recibiendo tanta atención es porque está superando a casi todos los demás algoritmos de aprendizaje automático cuando se trata de clasificar imágenes, procesamiento de lenguaje y procesamiento de datos de series temporales". Por ejemplo, la tarea de diagnosticar una enfermedad tradicionalmente ha sido

realizada por la experiencia humana, permitiendo diagnosticar enfermedades complejas como el cáncer. Sin embargo, parece que falta tiempo para que el SAP transforme áreas como la medicina.

La parte burocrática, aquella relacionada con trámites y servicios, está especialmente en riesgo debido a la rápida evolución de las tecnologías ligadas a la IA. Según Araya (2019), "Basándose en herramientas de aprendizaje profundo, la IA podría significar una mejora significativa en la prestación de servicios a los ciudadanos y en el trabajo de los profesionales del servicio público, al mismo tiempo que inspira a una nueva generación de tecnócratas para ingresar al gobierno". Aunque los sistemas de aprendizaje profundo y otras herramientas de IA pueden mejorar algunas áreas del gobierno, también pueden hacer obsoletas tareas que requieren interacción persona a persona, como la recopilación y trámites de documentos.

Sin embargo, para implementar las tecnologías de IA en los servicios del gobierno, es esencial que la población tenga acceso a Internet, ya que muchas herramientas de aprendizaje profundo requieren plataformas conectadas a la red. Según Pérez, Camacho, Mena y Arroyo (2016), los principales desafíos para que la gente realice trámites gubernamentales y servicios públicos en línea (conocido como e-gobierno) son la necesidad de crear un marco regulatorio sólido y reducir la brecha digital. Además, los autores sugieren que es importante el apoyo gubernamental en la cultura tecnológica y que estos servicios estén disponibles en todos los dispositivos y plataformas utilizados por la población.

De acuerdo con la Encuesta sobre Disponibilidad y Uso de Tecnologías de la Información y la Comunicación en los Hogares (ENDUTIH, 2018), existe evidencia de que más de la mitad de las personas en México se conectan a internet a través de dispositivos móviles y la tasa de adopción de estos dispositivos ha sido significativa. Según los datos de esa encuesta, el número de usuarios con teléfonos inteligentes en México pasó del 65.1% en 2015 al 83.6% en 2018, lo que demuestra que más de cuatro quintas partes de la población cuentan con acceso a esos dispositivos. Estos datos son de gran importancia, ya que un teléfono

inteligente suele ser mucho más económico y permite el acceso a internet a través de conexiones de wifi públicos.

Lo anterior facilita mayor acceso a las plataformas que emplean el Sistema de Aprendizaje Profundo (SAP) y herramientas relacionadas como el Deep Neural Network (sistema neuronal profundo). Por ende, se abre la posibilidad de hacer trámites con modelos de autenticación ligados a lectores biométricos integrados a los dispositivos móviles ("Machine Learning y Deep Learning, el futuro en presente", 2018). A pesar de que existe una brecha digital en México, ya que solo el 68% de los mexicanos cuenta con servicio privado de internet (INEGI, 2018), hay medios para que la población pueda hacer trámites con herramientas que emplean el SAP.

Por su parte, Japón, hasta 2020, ha alcanzado una tasa de adopción de internet del 92.7 por ciento (Statista, 2021). Al igual que en México, el teléfono inteligente es el dispositivo más utilizado por las personas para conectarse a la red. Esto permite que la población japonesa tenga una conexión a internet prácticamente ubicua, lo que permite la implementación de aplicaciones y recursos infraestructurales que no serían posibles con una población que depende de *hotspots* o puntos de conexión estáticos de wifi, como es el caso del internet municipal en México.

Ahora bien, estas estadísticas sobre la adopción de internet y el uso de dispositivos para conectarse a la red implican que una proporción creciente de la población puede acceder e incorporarse a las plataformas digitales de trabajo. Sin embargo, estas plataformas, junto con sus sistemas de inteligencia artificial basados en aprendizaje de máquina, como Uber, no son neutrales para los trabajadores. Según Toxtli et al. (2023), estas dinámicas generan desequilibrios de poder, donde los trabajadores enfrentan salarios por debajo del mínimo y son sometidos a prácticas intimidatorias para preservar su reputación o evitar la terminación de su empleo.

Además, las plataformas digitales de trabajo, impulsadas por sus sistemas de inteligencia artificial, pueden dar lugar a problemas sustanciales relacionados con

la privacidad, el acceso a la información y el derecho a la no discriminación (Covington, 2023; Chau, 2023). Esto se debe a que los sistemas de IA dependen de datos generados por humanos, lo que inevitablemente introduce sesgos presentes en quienes producen esos datos, incluso cuando se implementan filtros y contrapesos en el diseño y desarrollo de dichos sistemas (Morton, 2024). Así, las plataformas no solo perpetúan desigualdades estructurales, sino que también plantean serios desafíos éticos y sociales que requieren atención urgente.

En conclusión, se puede ver cómo comienzan a vislumbrarse y a implementarse aplicaciones del sistema de aprendizaje profundo en áreas que tradicionalmente habían sido competencia. Además, con base en sus avances, se abre el camino para que estas tecnologías penetren e impacten en cada vez más funciones y puestos de trabajo. Por lo tanto, la IA tiene repercusiones importantes en muchos sectores laborales que aún no han sido afectados por las primeras oleadas de automatización o revoluciones industriales.

4. Facebook y el uso de los algoritmos para la prevención del suicidio

Facebook es una plataforma social con más de tres mil millones de usuarios hasta principios de 2021 (Facebook Revenue and Usage Statistics, 2021). La gran cantidad de usuarios y los datos que generan hacen que sea muy atractiva para diferentes compañías que buscan expandir sus negocios a través de ese público cuyos datos pueden ser utilizados para generar un perfil personal y, con ello, publicidad dirigida específicamente a ellos. Sin embargo, más allá de las aplicaciones comerciales y de entretenimiento, la gran cantidad de usuarios, las herramientas en su interfaz gráfica y las relaciones entre los usuarios posicionan a la plataforma como un campo donde se pueden prevenir acciones como el suicidio.

Gomes De Andrade et al. comparte esta posición. (2018), investigadores de Facebook Inc. quienes plantean que "Hace 10 años, Facebook sentía firmemente que nuestro alcance y recursos nos colocan en una posición única para ayudar con ciertos problemas sociales y personales por los que las personas en nuestra plataforma pueden estar atravesando". Los investigadores de la plataforma social

explican que, durante más de una década, la plataforma ha desarrollado herramientas para ayudar en problemas como el duelo que enfrentan las personas cuando un ser querido fallece, a través de cuentas de legado: cuentas de Facebook inactivas a las que los familiares y amigos pueden acceder y administrar una vez que la persona fallece.

Otra de las herramientas tiene que ver con el duelo romántico que ocurre cuando se termina una relación sentimental y amorosa importante. Gomes De Andrade et al (2018) comentan que: “Esta función permite a las personas gestionar mejor una experiencia de ruptura romántica difícil al ofrecer herramientas específicas que gestionan cómo interactúan con amores anteriores, ayudando a las personas a terminar sus relaciones en Facebook con mayor facilidad, comodidad y sentido de control”. Por consiguiente, la plataforma, a partir de herramientas *in situ*, genera mecanismos e instrumentos para ayudar a los usuarios sobrellevar –y hasta sanar–, cuando se enfrentan a momentos complicados en sus vidas.

De acuerdo con Gomes De Andrade et al (2018), Facebook lleva implementando herramientas contra el suicidio por más de una década. Muchas de estas herramientas fueron desarrolladas con la ayuda de organizaciones como *Forefront Suicide Prevention*, y *National Suicide Prevention Lifeline* en los Estados Unidos; además de personas que han tenido experiencias muy cercanas en torno al suicidio. Por consiguiente, se puede apreciar que la plataforma social intenta generar estrategias para la prevención del suicidio, más allá de los laboratorios en los que laboran sus ingenieros.

En torno a la literatura y gama de herramientas para la prevención del suicidio, el equipo de Gomes De Andrade et al (2018) comentan que existen otras herramientas como la aplicación Radar que emplea algoritmos de IA, en aras de prevenir el suicidio. Asimismo, explican que los asistentes de voz en los dispositivos móviles e inteligentes como Cortana, Siri, Alexa, entre otros, tienen funciones que –con base en los comandos del usuario– activan funciones para dirigirlos a centros de prevención del suicidio. Y es importante destacar que

muchas de las herramientas de asistencia con voz, trabajan en sí empleando algoritmos de IA.

Tomando en cuenta al uso de la IA para prevenir el suicidio, un artículo publicado por D'Hotman y Loh, menciona que:

Los algoritmos se pueden diseñar para corregir y aprender de los errores (entrenamiento) para agregar la precisión de la confianza de un modelo predictivo de IA; esto se llama aprendizaje automático.⁶ Como tal, la IA, y el aprendizaje automático más específicamente, está bien posicionada para abordar el desafío de navegar por big data para la predicción de suicidios.

Ahora bien, antes de analizar cómo Facebook utiliza herramientas de la IA para la prevención del suicidio, hay que conocer cuáles fueron sus primeras estrategias. En un principio la plataforma social empleaba el método reactivo en que un amigo o familiar reportaba una publicación con contenido sospechoso de conducta suicida a la plataforma. Si la plataforma lo consideraba relevante, entonces, le mandaba información de prevención del suicidio a la persona sospechosa de tener intenciones suicidas.

Sin embargo, el método reactivo demostró ser limitado. Gomes De Andrade et al (2018), explicaron que:

Con este enfoque reactivo, pudimos brindar apoyo y recursos de manera regular a un número considerable de personas. Pero, después de más investigaciones sobre casos de suicidio, nos dimos cuenta de que muchas publicaciones que indicaban ideación suicida no estaban siendo reportadas, incluso cuando esas publicaciones habían sido vistas por un gran número de personas.

Con base en los problemas del método reactivo, el laboratorio de Facebook se preguntó si existía alguna forma de hacer mejor las cosas. Para afinar de sus herramientas para prevenir el suicidio, la plataforma social consideró emplear el aprendizaje de máquina. De acuerdo con Gomes De Andrade et al (2018) “Primero aplicamos ML como una forma de priorizar los informes que fueron revisados por nuestros revisores especialmente capacitados en el equipo de Operaciones Comunitarias para asegurarnos de que revisaran la información más urgente o susceptible de contener información procesable”.

Más tarde, continuaron usando los reportes de los usuarios como base para los datos de entrenamiento de la herramienta del aprendizaje de máquina. Para ello, se centraron en el análisis de texto de las publicaciones de interés, así como de sus comentarios asociados. Posteriormente “A través de AI, evaluamos las publicaciones primero cuando se publican y nuevamente cada vez que se agrega un comentario” (Gomes De Andrade et al, 2018).

En cuanto a la pregunta de por qué el equipo usó la herramienta de aprendizaje de máquina, Gomes De Andrade et al (2018), explican:

Decidimos utilizar *Machine Learning* porque proporcionaba los mayores matices para determinar la intención suicida de una publicación. Nuestros primeros intentos de clasificar el contenido utilizaron una lista simple de palabras que podrían relacionarse con la ideación, pero rápidamente quedó claro que necesitábamos un enfoque más matizado, y esto nos llevó al aprendizaje automático.

Es importante señalar que el uso del aprendizaje de máquina no implicó olvidar la participación de los seres humanos. En su metodología el laboratorio que se encargó de desarrollar el informe sobre el uso de herramientas para prevención de suicidio en Facebook, afirma que la participación de los seres humanos para examinar las publicaciones de interés en torno al suicidio es importante. Sin embargo, dentro de las publicaciones existen muchas que dan señales falsas y que pueden consumir demasiado tiempo de los curadores (personas que examinan la relevancia de las publicaciones). Por ello: “Exclure dichos informes de la revisión manual permitiría a nuestros revisores humanos dedicar menos tiempo, atención y energía a los informes no procesables, lo que solo aumenta su fatiga y afecta su productividad y motivación” (Gomes De Andrade et al, 2018).

El equipo de Facebook consiente de las implicaciones éticas del aprendizaje de máquina para prevenir el suicidio en una plataforma social. Una de ellas gira en torno a la privacidad, que según Gomes De Andrade et al (2018) la primera tiene que ver con el dilema de eficiencia versus privacidad. Esto es, entre mayor transparencia de datos, mayor eficiencia; sin embargo, a mayor privacidad (menos transparencia de datos), menor eficiencia. Gomes De Andrade et al (2018), explican:

Un enfoque que favorezca la eficacia apuntaría a notificar de manera proactiva a los amigos sobre publicaciones que expresan ideas suicidas al alentarlos a interactuar con la persona en peligro que fue la autora de dicha publicación. Pero las consideraciones de privacidad nos obligaron a no emitir ningún tipo de notificaciones proactivas y confiar en nuestros mecanismos de informes estándar. Las publicaciones marcadas como que expresan ideas suicidas por nuestro clasificador de aprendizaje automático aun generarán una recomendación sobre materiales y líneas de ayuda que los autores de la publicación, y solo ellos mismos, pueden aprovechar; pero no alertará de ninguna manera a nadie más.

Lo anterior también se aplica en la forma en que funciona el aprendizaje de máquina. En particular, en torno al aprendizaje de máquina que requiere datos claros e identificables para ser eficiente; sin embargo, esto puede poner en riesgo la privacidad de la persona. Por consiguiente, Gomes De Andrade et al (2018) plantean que: “Nuestros sistemas de IA están diseñados para detectar expresiones de suicidio sin almacenar y conectar esa información a personas específicas. El enfoque de nuestro trabajo de aprendizaje automático está en el contenido expresado y no en la persona que expresa el contenido”.

Otra pregunta ética gira en torno a “Argumentados y efectivos versus rigurosos e ineficaces (¿Cómo establecemos el umbral?)”. El aprendizaje de máquina es probabilístico en naturaleza; es decir, entre más información, mejor precisión adquiere. Esto complica la forma en que la plataforma puede implementar sus herramientas de prevención del suicidio, ya que para ser completamente eficiente tendría que escanear todas las publicaciones de los usuarios en aras de asumir de manera sumamente precisa si tienen algún intento suicida.

Para solucionar el dilema de eficiencia de efectividad versus eficiencia, la compañía propuso estas respuestas:

1. “Si bajamos el umbral, más publicaciones deberán ser revisadas por más personas que tengan menos probabilidades de ser procesables; esto plantea el riesgo de tener un número desproporcionado de revisores humanos mirando publicaciones no preocupantes.
2. Si elevamos el umbral, más precisas serán estas publicaciones y menos personas necesitaremos para hacer la revisión humana del contenido; pero

esto corre el riesgo de que falte contenido que debería haber sido marcado y revisado” (Gomes De Andrade et al, 2018).

Las preguntas éticas más importantes son las que el equipo de Facebook se hace a sí mismo. Esto abarca los límites de la privacidad, acceso a la información y el papel de una empresa privada en un tema tan delicado como el suicidio. El equipo de Gomes De Andrade et al (2018) se pregunta si es correcto que Facebook participe en la prevención del suicidio y si es relevante dedicar recursos a esta tarea. Además, plantean la pregunta más relevante en un mundo rodeado de plataformas y aplicaciones cotidianas que utilizan la IA: "¿Estamos cruzando líneas y sobrepasando límites, yendo más allá de lo que se supone que debemos hacer como empresa? (Gomes De Andrade et al, 2018).

En suman, las herramientas de prevención del suicidio de Facebook plantean preguntas éticas sobre los derechos humanos. Estas cuestiones forman parte de la discusión sobre la IA, ya que no solo se trata de una herramienta en desarrollo, sino también de un objeto de consumo que comienza a generar escenarios que requieren la atención de personas con una visión distinta. El sistema de Chat GPT, por ejemplo, no sólo plantea problemas en torno a las fuentes que cita, sino que también reproduce sesgos discriminatorios (Getahun, 2023). Es importante analizar cómo cada país y sus instituciones responden a los desafíos de esta tecnología para comprender cómo se hace frente a los problemas de la IA con la regulación existente y cómo las propuestas regulatorias buscan contener los desafíos.

CAPÍTULO CUARTO

TRAYECTORIA DE LOS PROBLEMAS ÉTICOS DE LA IA Y SU RELACIÓN CON LOS DERECHOS HUMANOS

La IA es una tecnología que busca que las máquinas puedan realizar tareas que normalmente requieren inteligencia humana, como el razonamiento, el aprendizaje, la percepción o la toma de decisiones. La IA tiene un gran potencial para mejorar la vida de las personas y contribuir al desarrollo de la sociedad en diversos ámbitos, como la educación, la salud, la economía o la seguridad. Sin embargo, el uso de la IA también plantea importantes desafíos y riesgos para los derechos humanos, como el derecho a la privacidad, a la no discriminación, a la libertad de expresión o el derecho al olvido, por mencionar solo algunos. Estos desafíos y riesgos se derivan de factores como la opacidad, la complejidad, la falta de control o la responsabilidad de los algoritmos de IA, así como de los posibles sesgos, errores o abusos que puedan cometer o sufrir los usuarios o los diseñadores de los sistemas de IA.

Por ello, es necesario establecer un marco normativo basado en principios éticos que garantice el respeto y la protección de los derechos humanos en la IA. Sobre todo, con base en las herramientas teóricas de la teoría de actor-red que permiten explicar la forma en que se deben abordar esos problemas desde el enfoque de derechos humanos. Antes de analizar en el capítulo quinto algunos casos empíricos, hay que conocer cómo se articulan los principios éticos y los objetos regulatorios más importantes en torno a la IA.

Ahora bien, es importante definir tres conceptos clave: ética, transparencia algorítmica o ética de inteligencia artificial, y qué se entiende por derechos humanos. La ética se puede considerar un concepto sombrilla ligado a distintas tradiciones filosóficas. Por ejemplo, el eudemonismo aristotélico, la teología moral cristiana y las diversas éticas hedonistas, incluyendo las versiones del utilitarismo, son formas de ética material que enfatizan el contenido psicológico del querer (felicidad, salvación, placer o bienestar) sobre su forma racional (Bilbeny, 2012);

mientras que Kant y otros autores plantean nociones éticas, como el imperativo categórico, basadas en la idea de que debemos actuar según principios universales (Sullivan, 1989), es decir, actuar de tal manera que nuestra acción pueda convertirse en una ley universal sin contradicciones.

Asimismo, se entiende por transparencia algorítmica el hecho de que las empresas deben ser claras sobre los algoritmos que emplean (Ibarra, 2020). Esto implica no solo divulgar cómo funcionan estos algoritmos, sino también asegurar que sus procesos sean justos, no discriminatorios y alineados con valores éticos. Además, es crucial que las empresas proporcionen mecanismos para la rendición de cuentas y permitan auditorías independientes para verificar la integridad y equidad de sus sistemas algorítmicos.

En cuanto a la ética algorítmica o ética en inteligencia artificial, esta debe girar en torno a estándares de responsabilidad, estándares éticos y estar alineada con derechos humanos (Huang et al., 2022), esto es, su dimensión con respecto a factores humanos. A nivel de responsabilidad, es importante reconocer quién asumirá las consecuencias en caso de que un sistema o modelo falle. Además, el objetivo final de la ética en inteligencia artificial es desarrollar una IA ética que pueda seguir principios éticos y comportarse de manera ética (idem), lo cual también implica los algoritmos que componen los sistemas. Finalmente, la ética en inteligencia artificial debe estar anclada y girar en torno a los derechos humanos, e idealmente el diseñador y los ingenieros de software deben tomar cursos de derechos humanos (idem) o trabajar con personas de otras disciplinas que tengan especialidad en el tema tanto a escala global como local, considerando las necesidades y particularidades éticas y morales a nivel local y regional.

Con base en esto, es importante definir brevemente qué se entiende por derechos humanos. La teoría de los derechos humanos parte del principio de que cada ser humano es un ser racional con autonomía moral y dignidad (Ledesma, 2023). Así pues, estos derechos son inherentes a todos los individuos, sin distinción de raza, sexo, nacionalidad, etnia, idioma, religión o cualquier otra condición, y su posible transgresión por parte de modelos y sistemas de inteligencia artificial implica no

solo perjuicio para los usuarios, sino la evidencia de un anti-programa o una relación de poder con respecto a estos mismos usuarios.

A partir de lo anterior, este capítulo examina los valores y principios éticos, así como los marcos normativos internacionales y nacionales de México y Japón en materia de IA y de derechos humanos a la luz de las teorías del programa y antiprograma de Latour, así como de la equivalencia planteada por Boltanski. Esto es relevante porque, antes de discutir cómo la IA afecta los derechos a la intimidad, acceso a la información y no discriminación, es necesario entender su origen, los mecanismos y los objetos que permiten su funcionamiento. Para ello, seguiremos la trayectoria de los derechos humanos, analizaremos las regulaciones más importantes en torno a la IA y cómo estos mecanismos regulatorios, en particular la Recomendación sobre la ética de la IA de la UNESCO de 2021 y la Ley de la Unión Europea en torno a la IA aprobada en mayo de 2024.

Basándose en las herramientas propuestas en el análisis de constelaciones, este capítulo examinará la trayectoria de los problemas éticos de la IA en los derechos humanos. También analizará cómo la regulación ha surgido a partir de esos principios. Por lo tanto, se examinará la trayectoria de los derechos humanos, así como la de algunos actores relevantes en este tema. Más adelante, se analizarán los problemas éticos de transparencia y *explicabilidad*, rendición de cuentas y responsabilidad, y derecho a la intimidad y protección de datos transparencia, privacidad, acceso a la información y no discriminación que se encuentran reconocidos en diferentes documentos internacionales.

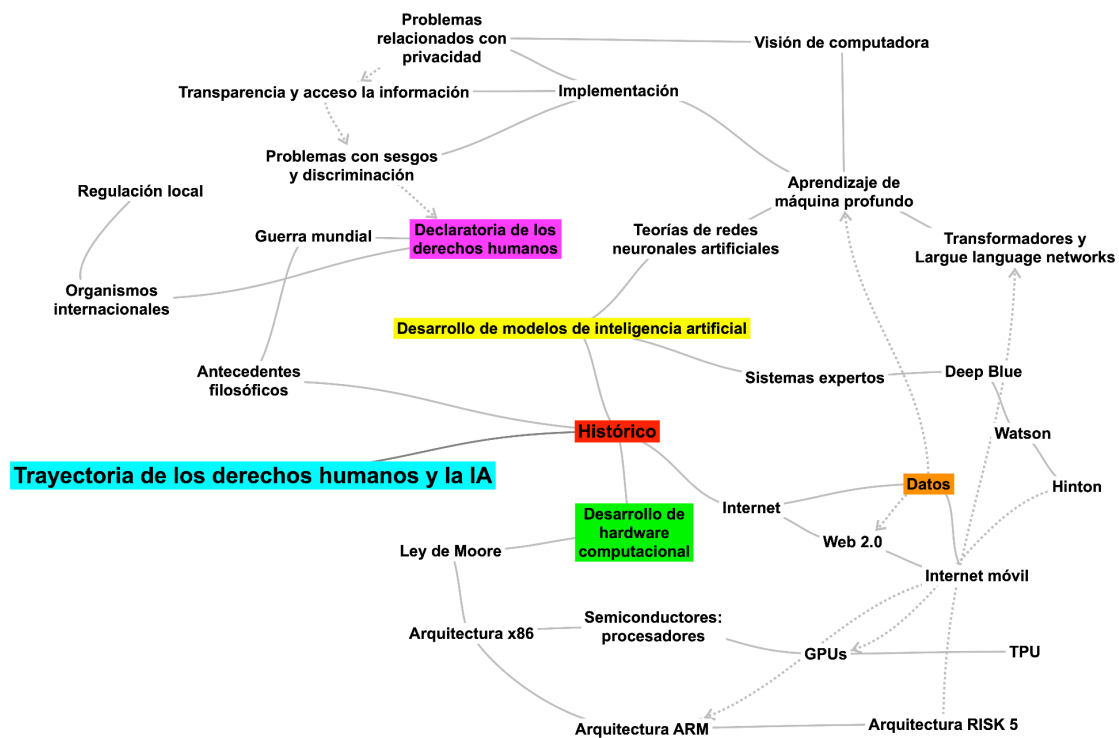
La regulación en torno a la IA se asignará como objeto de contorno. De la misma forma, los derechos humanos relacionados con la IA y las nociones generales de esta tecnología formarán parte de los sistemas de conocimiento. Asimismo, los derechos humanos de intimidad, acceso a la información y no discriminación propuestos en esta tesis y algunas herramientas teóricas relacionadas con el marco teórico serán asignados como conceptos clave.

En este capítulo se analizarán los derechos humanos reconocidos por la UNESCO y la ONU como marco fundamental que guía para desarrollar marcos normativos

nacionales en torno a la inteligencia artificial. Para efectos de este trabajo se entenderá por derechos humanos aquellos

Para entender los principios fundamentales de los derechos humanos, se analizará la Convención de los Derechos Humanos de 1948 y algunos otros compromisos internacionales relativos a estos derechos, entre ellos, la resolución del Consejo General de Naciones Unidas sobre *El derecho a la privacidad en la era digital* de 2019. Asimismo, se examinarán algunas de las regulaciones más relevantes en torno a la IA, incluyendo la Recomendación de la UNESCO sobre principios éticos de la IA aprobada en 2021; el compromiso alcanzado en la Asamblea de las Naciones Unidas para regular la IA en marzo de 2024; y la última legislación de la Unión Europea aprobada por el Consejo Europeo en mayo de 2024. Finalmente, se analizará la regulación existente en materia de IA en México y Japón.

Así, se distinguen una serie de asociaciones en torno a la IA y los derechos humanos que se pueden apreciar en la imagen 9 donde se muestra cómo la declaración de los derechos humanos y el desarrollo de modelos de inteligencia artificial están interconectados con diversos factores históricos, filosóficos y tecnológicos. Entre los temas abordados se incluyen problemas relacionados con privacidad, transparencia y acceso a la información, así como sesgos y discriminación. Se destacan también elementos históricos como la guerra mundial y la regulación local e internacional. Además, se presenta el desarrollo de hardware computacional, incluyendo la Ley de Moore, arquitecturas de procesadores y avances en semiconductores, GPUs y TPUs. La implementación de IA abarca teorías de redes neuronales, aprendizaje de máquina profundo, sistemas expertos y tecnologías específicas como Deep Blue, Watson y transformadores de lenguaje.



(Imagen 9, de autoría propia)

Los derechos humanos y los problemas éticos de la IA están estrechamente relacionados. Los problemas éticos de la IA surgen en el contexto social y también son moldeados e implementados a través de actores no humanos, como declaraciones, constituciones y otras representaciones. Estos actores no humanos han desempeñado un papel crucial en la promoción y defensa de estos principios y han ayudado a movilizar a los actores humanos para reconocer y defender sus derechos. Por lo tanto, es necesario seguir perfeccionando y ampliando la protección de estos principios para garantizar que la tecnología de IA se utilice sin vulnerar derechos humanos.

Sin embargo, existe un problema de traducción en el concepto de derechos humanos. Este problema se presenta a menudo entre diferentes áreas de las ciencias sociales y entre los aspectos filosóficos de varias sociedades. En particular, hay una tradición naturalista (fundamentación iusnaturalista) que

sostiene que los seres humanos nacen con derechos y que “además, consideraba que el origen de la sociedad y del Estado se hallaba, según sabemos, en el 'contrato social” (Álvarez Ledesma, 2023); y una tradición jurídica (iuspositivista), que afirma que estos derechos se adquieren a través de un marco regulatorio (Álvarez Ledesma, 2023). El problema de estas traducciones es que siguen reduciendo el mundo a un ente social donde los derechos humanos parecen un miasma intangible que sólo se manifiesta con discusiones filosóficas, teológicas y tradiciones jurídicas.

Se tiene que pensar en los derechos humanos como un objeto de contorno. Esto sigue cuidadosamente la propuesta metodológica planteada en la traducción. Ya que examinar a los derechos humanos como un objeto de contorno no solo permite analizar su trayectoria bajo el método constelaciones, sino describir e indagar su impacto a partir de las traducciones que han tenido y de cómo al enraizarse en actores humanos son capaces de producir un programa que asimismo genera respuestas a los antiprogramas.

Los derechos humanos se convierten en un actante con contenido. Esto nos permite analizar el contenido de los derechos humanos más allá de aspectos puramente sociales y humanos, y cómo se transformó mediante la necesidad de traducciones de distintos actores. Además, incluye actores no humanos en el análisis, como la propia Declaración de Derechos Humanos de 1948, que puede desempeñar un papel importante.

Es importante señalar que esta Declaración de Derechos Humanos de 1948 se ha ido complementando con otros documentos que reconocen y promueven derechos humanos, entre ellos: la Convención sobre el Estatuto de los Refugiados (1951), el Convenio sobre la Discriminación (Empleo y Ocupación) (1958), la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial (1965), el Pacto Internacional de Derechos Civiles y Políticos (1966), el Pacto Internacional de Derechos Económicos, Sociales y Culturales (1966), la Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer (1979), la Convención sobre los Derechos del Niño (1989), la Convención

sobre los Derechos de las Personas con Discapacidad (2006), la Convención relativa a la Lucha contra las Discriminaciones en la Esfera de la Enseñanza (1960) y la Convención sobre la Protección y Promoción de la Diversidad de las Expresiones Culturales (2005), así como cualesquiera otros instrumentos, recomendaciones y declaraciones internacionales pertinentes. Sin embargo, todos estos documentos no serán objeto de estudio para la presente investigación, ya que exceden los límites de este trabajo.

Al analizar el contenido de la Declaración de Derechos Humanos, las legislaciones locales y las normativas de regulación, y la discusión de los derechos humanos en relación con la IA, se verificará si estos actores reconocen los principios éticos planteados al inicio de la tesis. De esta manera, se podrá examinar la trayectoria de los derechos humanos y cómo se han creado objetos de contorno y se han realizado traducciones para contrarrestar los dilemas éticos de la IA. Y se podrá examinar de manera holística la formación de programas, antiprogramas y traducciones en torno al impacto de la IA en la sociedad.

Ortega (2023) menciona que la ética actúa como guía para el comportamiento correcto, aunque no establece deberes ni castigos por no seguirlos; por lo tanto, no es sorprendente que haya un aumento en el financiamiento de organizaciones dedicadas a la ética de la IA como *The Berkman Klein Center for Internet & Society* (Martinovs, 2019), *Global AI Ethics and Governance Observatory* (Ethics of Artificial Intelligence | Unesco, s. f.) o la Alianza Nacional de Inteligencia Artificial en México (Sobre Ania, s. f.), en comparación con la escasez de legislación para su control. Así pues, los principios éticos deben de funcionar para generar materias regulatorias concretas en torno a “deberes y obligaciones para los creadores y proveedores de los algoritmos” (Ortega, 2023).

En resumen, este capítulo analizará si los instrumentos jurídicos internacionales de derechos humanos incorporan los principios éticos propuestos al inicio de esta tesis para regular o enfrentar la IA. Además, la contribución relevante será examinar la evolución de los derechos humanos a partir de los documentos o actores que los consagran. Y finalmente, cómo su contenido prevé o contempla

las herramientas para contrarrestar el impacto de una tecnología tan influyente y presente en la sociedad como lo es la IA.

1. Problemas éticos en torno a la IA

El desarrollo y la implementación de la inteligencia artificial (IA) y el aprendizaje profundo han traído consigo oportunidades sin precedentes en diversas áreas, desde la educación y la salud hasta la seguridad y el medio ambiente. Sin embargo, su uso también ha generado preocupaciones éticas y legales que requieren una respuesta inmediata para evitar vulneraciones a los derechos humanos. Estos problemas no responden a escenarios de ciencia ficción, como la rebelión de las máquinas, sino a formas concretas en las que la IA afecta la privacidad, la equidad, el acceso a la información y la no discriminación.

Diferentes posturas han surgido en torno al impacto de la IA. De acuerdo con De Asís (2020), existen dos corrientes principales: los bioconservadores, que enfatizan los riesgos que el avance tecnológico puede traer para la humanidad y el medio ambiente, y los transhumanistas, que ven la tecnología como una herramienta para mejorar la vida humana y abordar problemas sociales. En este contexto, es crucial analizar los dilemas éticos que plantea la IA desde una perspectiva basada en derechos humanos.

Ahora bien, uno de los problemas más urgentes es la amenaza a la privacidad debido al uso de sistemas de IA para la vigilancia masiva y la recolección de datos personales. Los algoritmos de reconocimiento facial y las tecnologías de análisis de comportamiento han sido empleados sin el consentimiento explícito de las personas, lo que plantea serios riesgos para la protección de la información personal (Amnistía Internacional, 2022). Un caso emblemático es el uso de la IA por parte de Clearview AI, cuya base de datos de rostros fue recopilada sin autorización y vendida a entidades gubernamentales y privadas, lo que generó múltiples demandas por violaciones al derecho a la privacidad (Hill, 2020).

La IA no es neutral. Los sesgos en los datos de entrenamiento pueden generar sistemas discriminatorios que perpetúan desigualdades sociales, o generar

inscripciones que las reflejan. Por ejemplo, estudios han demostrado que los algoritmos de reconocimiento facial tienen tasas de error significativamente más altas al identificar a personas racializadas en comparación con individuos de piel clara (Buolamwini & Gebru, 2018). Este tipo de discriminación algorítmica ha llevado a arrestos erróneos en Estados Unidos y a la exclusión de ciertos grupos en procesos de contratación automatizados (O'Neil, 2016).

Los sistemas de IA empleados en plataformas digitales pueden restringir el acceso a información de manera arbitraria. Los algoritmos de recomendación, diseñados para maximizar la retención del usuario, pueden crear burbujas de filtro que limitan la diversidad informativa y refuerzan sesgos ideológicos (Pariser, 2011) o convertir a las plataformas social en cajas de resonancia en torno las preferencias ideológicas (Legends, 2024). Además, la IA ha sido utilizada para la generación y difusión de *deepfakes*, lo que plantea nuevos desafíos en la lucha contra la desinformación y la manipulación mediática (Chesney & Citron, 2019).

Los sistemas de IA empleados en plataformas digitales pueden restringir el acceso a información de manera arbitraria. Por ejemplo, los algoritmos de recomendación, y posteriormente los sistemas de IA diseñados para maximizar la retención del usuario, pueden crear burbujas de filtro que limitan la diversidad informativa y refuerzan sesgos ideológicos (Pariser, 2011; Legends, 2024). Además, la IA ha sido utilizada para la generación y difusión de *deepfakes*, lo que plantea nuevos desafíos en la lucha contra la desinformación y la manipulación mediática (Chesney & Citron, 2019).

Las decisiones automatizadas basadas en IA pueden afectar derechos fundamentales sin mecanismos claros de apelación. En sectores como la salud o el crédito financiero, los sistemas de IA han negado tratamientos médicos o préstamos sin que los afectados comprendan las razones detrás de estas decisiones (Eubanks, 2018). La falta de transparencia en estos procesos impide que las personas impugnen decisiones que podrían afectar gravemente sus vidas.

El impacto de la IA en los usuarios ha generado preocupaciones sobre la transgresión de los derechos humanos. Ante ello, diversas organizaciones no

gubernamentales y supranacionales han propuesto medidas y recomendaciones para abordar estos problemas. Sin embargo, es fundamental analizar estas iniciativas considerando tanto sus fortalezas como sus limitaciones.

A. Recomendación sobre la ética de la IA de la UNESCO

Como resultado de la 41ª reunión de la Conferencia General de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), celebrada en París del 9 al 24 de noviembre de 2021, se aprobó la Recomendación sobre la ética de la inteligencia artificial. Este documento se considera uno de los principales instrumentos a nivel global que define valores y principios éticos para fortalecer la cooperación y la solidaridad mundial, con el objetivo de enfrentar los desafíos que plantea la IA (UNESCO, 2021).

La guía ética y normativa, que pretende ser el marco regulatorio de ética de la IA, se da en documentos internacionales fundamentales, como la Declaración sobre el Derecho al Desarrollo (1986); la Declaración sobre las Responsabilidades de las Generaciones Actuales para con las Generaciones Futuras (1997); la Declaración Universal sobre Bioética y Derechos Humanos (2005); la resolución de la Asamblea General de las Naciones Unidas sobre el examen de la Sociedad de la Información (A/RES/70/125) (2015).

La Recomendación establece cuestiones éticas relacionadas con la IA a través de una reflexión sistemática basada en un marco integral, global, multicultural y evolutivo de valores, principios y acciones interdependientes. Este marco tiene como objetivo guiar a los Estados miembros en la responsabilidad de enfrentar los efectos conocidos y desconocidos de las tecnologías de la IA en los seres humanos, las sociedades, el medio ambiente y los ecosistemas (UNESCO, 2021).

Entre los objetivos de la Recomendación se encuentran proporcionar un marco universal de valores, principios y acciones para orientar a los Estados miembros en la formulación de leyes, políticas y otros documentos relacionados con la IA; guiar las acciones de individuos, grupos y empresas para asegurar la incorporación de la ética en las etapas del ciclo de vida de los sistemas de IA;

proteger, promover y respetar los derechos humanos y las libertades fundamentales, especialmente la dignidad humana y la igualdad; fomentar el diálogo multidisciplinario y pluralista entre las partes interesadas sobre cuestiones éticas relacionadas con la IA.

El documento subraya que los valores y principios éticos propuestos deben ser respetados por todos los actores durante el ciclo de vida de los sistemas de IA. Estos valores deben orientar las políticas y regulaciones nacionales y ajustarse en todo momento al derecho internacional, en particular a la Carta de las Naciones Unidas en materia de derechos humanos. Además, deben alinearse con los Objetivos de Desarrollo Sostenible de las Naciones Unidas (Agenda 2030).

La Recomendación de la UNESCO reconoce que los principios y valores éticos aplicados a la IA pueden generar tensiones entre sí. Por ello, en cada caso se deberá realizar una evaluación del contexto, teniendo en cuenta el principio de proporcionalidad y considerando siempre el respeto de los derechos humanos y las libertades fundamentales.

En este sentido, los valores desempeñan una función importante como ideales que deben orientar las políticas públicas y las normas jurídicas de todos los Estados miembros. Los valores que propone la Recomendación de la UNESCO son:

1. Respeto, protección y promoción de los derechos humanos, las libertades fundamentales y la dignidad humana: “el respeto, la protección y la promoción de la dignidad humana y de los derechos establecidos por el derecho internacional, en particular el derecho internacional de los derechos humanos, son esenciales a lo largo del ciclo de vida de los sistemas de IA” (UNESCO, 2022).
2. Prosperidad del medio ambiente y los ecosistemas: “La prosperidad del medio ambiente y los ecosistemas debería ser reconocida, protegida y promovida a lo largo del ciclo de vida de los sistemas de IA” (idem).
3. Garantizar la diversidad y la inclusión: La diversidad en las elecciones de estilo de vida, creencias, opiniones, expresiones o experiencias personales,

incluyendo el uso opcional de sistemas de IA y la co-creación de estas arquitecturas, no debería ser restringida en ninguna etapa del ciclo de vida de dichos sistemas (idem).

4. Vivir en sociedades pacíficas, justas e interconectadas: “Los actores de la IA deberían propiciar sociedades pacíficas y justas, sobre la base de un futuro interconectado en beneficio de todos, compatibles con los derechos humanos y las libertades fundamentales, y participar en su construcción”.

Por otro lado, la Recomendación sobre la ética de la inteligencia artificial indica que los principios éticos definen de manera más concreta los valores subyacentes, facilitando su aplicación en declaraciones políticas y acciones relacionadas con la IA. Estos principios éticos deben guiar la integración de la IA al servicio del bienestar de la humanidad, las personas, las sociedades y el medio ambiente y los ecosistemas. Es por ello, que la Recomendación establece principios éticos que se deben considerar en el desarrollo de la IA, entre los que destacan:

1. Proporcionalidad e inocuidad: “reconocerse que las tecnologías de la IA no garantizan necesariamente, por sí mismas, la prosperidad de los seres humanos ni del medio ambiente y los ecosistemas” (idem).
2. Seguridad y protección: Se deben evitar y mitigar los daños no deseados y las vulnerabilidades a ataques a lo largo del ciclo de vida de los sistemas de IA para garantizar la seguridad y protección de los seres humanos, el medio ambiente y los ecosistemas (idem).
3. Equidad y no discriminación: “Los actores de la IA deberían promover la justicia social, salvaguardar la equidad y luchar contra todo tipo de discriminación, de conformidad con el derecho internacional” (idem).
4. Sostenibilidad: La evaluación continua de los efectos de las tecnologías de IA en los ámbitos humano, social, cultural, económico y ambiental debe realizarse con pleno conocimiento de sus repercusiones en la sostenibilidad, alineándose con los Objetivos de Desarrollo Sostenible (ODS) de la ONU (idem).

5. Derecho a la intimidad y protección de datos: Los datos para los sistemas de IA deben recopilarse, utilizarse, compartirse, archivarse y eliminarse de acuerdo con el derecho internacional y los valores de la Recomendación, respetando los marcos jurídicos nacionales, regionales e internacionales pertinentes
6. Supervisión y decisión humanas: Los marcos de protección de datos y los mecanismos relacionados deben basarse en principios y normas internacionales de protección de datos, asegurando la recopilación, uso y divulgación de datos personales y el ejercicio de los derechos de los interesados, garantizando un objetivo legítimo y una base jurídica válida para el tratamiento de los datos personales, incluido el consentimiento informado.
7. Transparencia y explicabilidad: “La transparencia es necesaria para que los regímenes nacionales e internacionales pertinentes en materia de responsabilidad funcionen eficazmente” (idem). Y la explicabilidad implica hacer comprensibles los resultados de los sistemas de IA y proporcionar información sobre ellos.
8. Responsabilidad y rendición de cuentas: Los actores de la IA y los Estados Miembros deben respetar, proteger y promover los derechos humanos y las libertades fundamentales, así como la protección del medio ambiente y los ecosistemas. Deben asumir su responsabilidad ética y jurídica según el derecho nacional e internacional, especialmente en derechos humanos, y seguir las directrices éticas durante todo el ciclo de vida de los sistemas de IA.
9. Sensibilización y educación: La sensibilización y comprensión pública sobre las tecnologías de IA y el valor de los datos deben promoverse a través de educación accesible, participación cívica, competencias digitales, ética de la IA, alfabetización mediática e informacional, y capacitación conjunta de gobiernos, organizaciones, sociedad civil, universidades, medios, líderes comunitarios y sector privado, considerando la diversidad lingüística, social

y cultural, para garantizar una participación pública efectiva y decisiones informadas sobre el uso de sistemas de IA.

10. Gobernanza y colaboración adaptativas de múltiples partes interesadas:

“En la utilización de datos deben respetarse el derecho internacional y la soberanía nacional” (idem).

Si bien el documento reconoce que la IA tiene el potencial de transformar múltiples ámbitos de la vida humana, también plantea riesgos que deben abordarse desde una perspectiva ética. Para ello, la Recomendación establece principios fundamentales, como el respeto a los derechos humanos, la equidad, la transparencia y la supervisión humana en el desarrollo y uso de sistemas de IA.

No obstante, la efectividad de la Recomendación enfrenta varias limitaciones. En primer lugar, su carácter no vinculante implica que los Estados no están obligados a implementarla, lo que genera diferencias significativas en su adopción. Además, aunque el documento enfatiza la necesidad de responsabilidad y rendición de cuentas, no establece mecanismos concretos para garantizar su cumplimiento ni para sancionar a quienes no respeten sus directrices lo cual es uno de los grandes problemas y discusiones en torno a leyes, propuestas regulatorias y recomendaciones sobre la investigación, desarrollo y aplicación de los sistemas de la IA.

Otro desafío radica en la aplicación práctica de estos principios en contextos nacionales con diferentes marcos regulatorios y niveles de desarrollo tecnológico. La implementación de políticas éticas en IA depende en gran medida de la voluntad política, los recursos disponibles y la capacidad institucional de cada país. En algunos casos, el exceso de regulación puede ralentizar la innovación, mientras que una regulación insuficiente podría permitir el desarrollo de sistemas de IA con sesgos discriminatorios o riesgos para la privacidad. Además, la reciente tendencia hacia la desregulación, impulsada por la competencia internacional en el desarrollo de IA, puede llevar a que la protección de los derechos humanos quede relegada en favor de avances tecnológicos rápidos. Por ejemplo, durante la Cumbre de Acción sobre Inteligencia Artificial celebrada en París en febrero de

2025, se observó un enfoque predominante en la promoción de inversiones y una adopción acrítica del potencial de la IA, con escasa atención a los riesgos asociados y a la necesidad de una gobernanza basada en los derechos humanos (Article 19, 2025).

A pesar de estas limitaciones, la Recomendación ha servido como referencia para diversos organismos y gobiernos en la formulación de estrategias nacionales sobre IA. Sin embargo, aún se requieren estudios que evalúen su impacto real y la forma en que ha influido en la toma de decisiones en distintos países.

B. Estudio particular de principios éticos de la IA

En esta investigación importa analizar los siguientes principios éticos que deberán formar parte de los marcos normativos que los Estado miembro de la UNESCO aceptaron al aprobar la Recomendación sobre la ética de la inteligencia artificial:

- La privacidad y la protección de datos: La IA analiza grandes cantidades de datos personales, que pueden recopilarse, compartirse y usarse sin el consentimiento de los afectados. Esto puede vulnerar el derecho a la intimidad y exponer a las personas a riesgos de vigilancia, manipulación o discriminación de sus datos personales (UNESCO, 2021). Asimismo, este afecta en esencia el derecho a la privacidad, que de acuerdo con el Diccionario de protección de datos personal del Instituto Nacional de Protección de Datos Personal (INAI, México) “se puede entender como el ámbito de la vida privada que se tiene derecho a proteger de cualquier intromisión” (Ibarra, 2020: pp 272).
- Transparencia y explicabilidad: La IA puede tomar decisiones que afectan a los derechos y libertades de las personas, como el acceso a servicios públicos, oportunidades laborales o justicia. Sin embargo, estos procesos de decisión pueden ser opacos, complejos o incomprensibles para los humanos, lo que dificulta el control, la supervisión y la impugnación de los resultados (UNESCO, 2021). Además, la responsabilidad por los daños o

perjuicios causados por la IA puede ser difícil de determinar o atribuir (Parlamento Europeo, 2020). Asimismo, el INAI plantea que el derecho de acceso a la información:

- Se logra a través de la transparencia de la información pública. En este mismo sentido, la Suprema Corte de Justicia de la Nación (SCJN) ha señalado que la dimensión colectiva del derecho de acceso a la información constituye el pilar de los Estados democráticos y que permite a la sociedad el intercambio libre y abierto de información incluyendo aquella que critique al Estado, fomentando con ello el ejercicio de la tolerancia y permitiendo el pluralismo y la participación ciudadana (Cejudo, 2019).
- Equidad y no discriminación: La IA puede reproducir o amplificar los sesgos y prejuicios existentes en los datos, los algoritmos o los usuarios, lo que puede conducir a resultados injustos o discriminatorios para determinados grupos o individuos. Por ejemplo, la IA puede afectar negativamente a las personas en función de su género, raza, edad, discapacidad u orientación sexual (UNESCO, 2021; IDHC, 2024). La IA también puede generar o agravar las brechas digitales entre países o regiones con diferentes niveles de desarrollo o acceso a la tecnología (UNESCO, 2021).
- Rendición de cuentas y responsabilidad: Es importante que los actores responsables por la implementación, investigación, comercialización de la inteligencia artificial tengan responsabilidad sobre las consecuencias negativas que esta tenga. En este sentido, la UNESCO (2022) plantea que: “Los actores de la IA deben asumir la responsabilidad de la concepción y la aplicación de los sistemas de IA de manera que se garantice la protección de la información personal durante todo el ciclo de vida del sistema de IA”.
- La seguridad y la paz: La IA puede tener aplicaciones militares o de seguridad que plantean graves amenazas para los derechos humanos y el derecho internacional humanitario. Por ejemplo, el uso de armas autónomas letales que pueden seleccionar y atacar objetivos sin

intervención humana significativa puede socavar el principio de humanidad y el control humano sobre el uso de la fuerza (IDHC, 2021). La IA también puede usarse con fines maliciosos o criminales, como el ciberataque, la desinformación o la incitación al odio (UNESCO, 2021).

I. El problema de la discriminación

Uno de los mayores desafíos de la IA es la discriminación y los sesgos que puede generar. Los sesgos se pueden dividir en dos categorías: los creados por la IA y los aprendidos por la IA. En el primer caso, la IA puede promover sesgos entre los actores humanos al actuar como un objeto que tiene influencia sobre otros actores. En el segundo caso, la IA puede aprender sesgos a través de la etapa de entrenamiento de datos cuando se trata de aprendizaje supervisado o de la base de datos cuando se trata de modalidades de aprendizaje no supervisado.

Un ejemplo concreto de cómo la IA puede generar sesgos es el caso de Facebook. La plataforma social ha utilizado herramientas de IA y aprendizaje de máquina para crear sesgos entre los usuarios. Esto se hizo evidente en 2021 cuando una antigua colaboradora de la compañía hizo declaraciones que causaron escándalo:

“El algoritmo de la plataforma está diseñado para fomentar una mayor participación de los usuarios de cualquier forma posible, incluso sembrando discordia y recompensando la indignación. Este problema fue planteado por el equipo de integridad de Facebook, que también propuso cambios en el algoritmo que suprimiría, en lugar de acelerar, tal animadversión entre los usuarios” (Sadowski, 2021).

Además, es importante mencionar que el derecho a la no discriminación es fundamental en el uso de la IA. Es esencial asegurar que la IA no discrimina a personas en base a su género, raza, orientación sexual, discapacidad, entre otros aspectos, ya que esto podría tener consecuencias graves en la sociedad. Por lo tanto, es crucial desarrollar medidas para prevenir y detectar la discriminación en el uso de la IA.

En octubre de 2021, Frances Haugen reveló cómo la plataforma social y su ecosistema de plataformas, especialmente Instagram, fomentan contenido que hace que los usuarios se enganchen. Aunque se descubrieron los sesgos que los algoritmos de la plataforma causan y se propusieron soluciones, el fundador de la compañía prefiere ignorarlos para continuar con la dinámica de enganche-datos-publicidad-perfección de algoritmos de IA-ganancia. El problema de los sesgos promovidos por Facebook (ahora Meta) y sus plataformas es evidente, pero puede ocurrir en cualquier otra plataforma donde la IA promueva sesgos para generar enganche por parte de las personas.

La discusión sobre los sesgos generados por la IA ha llegado incluso a la Organización de las Naciones Unidas (UN/ONU) que en 2021 expresó las consecuencias que han tenido las herramientas de la IA para promover contenido que enganche a sus usuarios:

“Esto ha llevado a la popularidad de publicaciones extremistas y llenas de odio, difundidas por grupos que de otra manera serían conjuntos marginales poco conocidos. Durante la pandemia de COVID-19, esa práctica también ha llevado a la diseminación de información erróneamente peligrosa sobre el virus, lo que podría llevar a que más personas se infecten” (Sesgos, racismo y mentiras: afrontar las consecuencias no deseadas de, 2021).

El segundo tipo de sesgo, el que aprende la IA, puede resultar igual o incluso más perjudicial que el primero. Esto se debe a que puede perpetuar problemas como la discriminación o racismos sistémicos, abriendo la puerta a otros fenómenos adversos como la discriminación automatizada en la solicitud de trabajo, fallas en el sistema de justicia a partir de decisiones tomadas por los sistemas de vigilancia, y una gran duda sobre la fiabilidad de los resultados que arroja la IA.

Es importante mencionar que el peligro de que la IA cree sesgos no es algo nuevo. Muchas personas han alertado sobre los riesgos potenciales de las prácticas de Facebook. Por ejemplo, como se menciona en un artículo de O'Neill de 2016, "en cuestión de horas, Facebook podría recopilar información de decenas de millones de personas, o más, midiendo el impacto que sus palabras y enlaces compartidos tenían entre sí, y podría utilizar ese conocimiento para influir en las acciones de las personas".

Para influir intencionalmente o no en el comportamiento de las personas, los algoritmos de IA utilizan proxies. Estos proxies son categorías específicas que sirven como punto de referencia al tomar decisiones o diseñar modelos de aprendizaje automáticos. Por ejemplo, la edad, la preferencia musical, el número de seguidores en Twitter, entre otros.

En este sentido, O'Neill (2016) explica que una de las maneras en que las personas y los algoritmos toman decisiones es a través de proxies o banderas de características. Pueden ser útiles para distinguir entre el ruido de información al momento de tomar una decisión. Sin embargo, la autora advierte que "cuando se crea un modelo a partir de proxies, es mucho más fácil que la gente lo explote. Esto se debe a que los proxies son más fáciles de manipular que la complicada realidad que representan" (O'Neill, 2016).

Como ejemplo, imaginamos una campaña para contratar a un cuidador de gatos cuyo proxy son las fotos de gatos rescatados en el perfil de redes sociales del candidato. A primera vista, este proxy es beneficioso ya que existe evidencia empírica de que la persona que busca ser cuidador de gatos tiene experiencia y amor por estos animales. Sin embargo, como bien advierte O'Neill, si la gente se entera de este proxy, todos los que busquen ese trabajo intentarán tomarse fotos con gatos para simular que son buenos samaritanos o que tienen afecto a estos animales, a pesar de que solamente busquen un fin económico.

Otro ejemplo negativo de los proxies son las clasificaciones de las mejores universidades. En los Estados Unidos, el sistema de calificaciones más importante fue establecido por un periódico, y ese modelo —junto con sus proxies— se convirtió en estándar al que muchas universidades aspiran. Sin embargo, O'Neill (quien sugirió ese ejemplo en su libro *Weapons of math destruction*) se dio cuenta que el modelo de periódico *US Today* forzó de manera inadvertida: "...a apuntar exactamente a los mismos objetivos, lo que crea una carrera de ratas y muchas consecuencias dañinas no deseadas" (O'Neill, 2016).

Si se conocen los proxies, se conoce lo que la IA valora en cuanto a contenido. De esta forma, una universidad puede ajustar sus programas para subir en las

calificaciones de las mejores universidades, y una persona puede estructurar su currículum en base a lo que pide una empresa o un área laboral específica. De acuerdo con O'Neill (2016): "La clave es saber qué buscan las máquinas. Pero aquí también, en un universo digital que se promociona como justo, científico y democrático, los iniciados encuentran la manera de obtener una ventaja crucial".

En el caso de las admisiones a las universidades en los Estados Unidos, O'Neill observa que los algoritmos y las herramientas de IA aumentan la desigualdad. Esto es importante en los estudiantes que envían sus solicitudes. La autora argumenta que "quienes no toman estos pasos quizás nunca sepan que envían su currículum a un agujero negro". Es un ejemplo más en el que los ricos e informados tienen la ventaja y los pobres tienen más probabilidades de salir perdiendo" (O'Neill, 2016).

En este sentido, Andrew Ng (s.f en Coursera) menciona que "la IA puede estar sesgada y discriminar injustamente a las minorías u otros grupos". Se observa que el fenómeno de la imparcialidad es uno de los mayores problemas en torno a la IA. Esto se debe en primer lugar a los sesgos promovidos por las mismas herramientas de IA, así como a los sesgos aprendidos de los seres humanos, lo cual genera problemas éticos mucho más serios y complejos, como la discriminación.

La IA puede aprender a discriminar. De acuerdo con la ONU, una de las consecuencias de que la IA es su capacidad adquirir patrones de discriminación a partir de los datos con los que aprende o se entrena. Por ello, en su artículo sobre Sesgos, racismo y mentiras: afrontar las consecuencias no deseadas de la IA (2021) expuso el siguiente ejemplo:

Se cree que este es el primer arresto injusto de este tipo: los oficiales de policía involucrados confiaron en una IA de reconocimiento facial para atrapar al hombre, pero la herramienta no conocía las diferencias entre los rostros de los ciudadanos negros porque las imágenes usadas para entrenarlo eran casi blancas. Por suerte, rápidamente quedó claro que no se parecía en nada al sospechoso que se ve en una foto tomada de las cámaras de seguridad de la tienda, y fue liberado, aunque pasó varias horas en la cárcel.

De acuerdo con Raso et al., (2018) “La IA puede perpetuar fácilmente los patrones existentes de prejuicio y discriminación, ya que la forma más común de implementar estos sistemas es entrenarlos para que reproduzcan los resultados logrados por los tomadores de decisiones humanos”. Existe entonces un nivel de complejidad en aras de mitigar los patrones de discriminación. Dado que, como se mencionó anteriormente, la IA no solamente puede aprender con datos que reciba de los seres humanos, sino también puede aprender por sí misma de datos que percibe de su entorno: datos y contenido generados por humanos; y por consiguiente, susceptibles a sesgos.

Por tanto, la IA solo requiere de datos sesgados que no hayan pasado por un proceso de supervisión por parte de los humanos para aprender a discriminar. En este sentido, Andre Ng (en Courser, s.f), explica:

“Por ejemplo, si una niña de ocho años va a un motor de búsqueda de imágenes y busca a un director ejecutivo, si solo ven fotos de hombres o si no ven a nadie que se parezca a ellos ni por género ni por etnia, no lo hacemos. No quiero que se desanime de seguir una carrera que podría” (Ng, en Coursera, s.f).

Los problemas de los sesgos que aprende la IA no solamente se limitan a temas penales. Como menciona O'Neill, si las personas tienden a confiar en la IA sin preguntarse por qué sus herramientas toman algunas decisiones, entonces se pierde el incentivo de vigilar la ya de por sí compleja serie de herramientas ligadas a esta tecnología. Esto puede provocar que pasen desapercibidos fenómenos como la discriminación laboral, en la salud, y discriminación de género en distintos espacios.

En lo que se refiere al problema de los sesgos, Schuchmann (2019) explica que esto también tiene que ver con la propia naturaleza de la inteligencia. La tendencia que tienen las herramientas del aprendizaje de máquina a generalizar. En este sentido “Los modelos de aprendizaje profundo tienen problemas para generalizar más allá de los datos de entrenamiento proporcionados. Kansky et al, demostraron que un modelo entrenado en el juego *Atari Breakout* fallaba cuando se realizaban pequeños cambios en el entorno” (Schuchmann, 2019).

En conclusión, la IA genera sesgos. Esto tiene consecuencias enormes como dividir a la sociedad y discriminar a las personas a partir de los resultados que proponen las herramientas del aprendizaje de máquina. Estos problemas se acentúan por las mismas características de la IA, como el hecho de que se desconozca o se proteja la forma en qué se tomen decisiones lo que deriva en una caja negra que dificulta tomar medidas de contingencia.

II. Privacidad

La privacidad es otro aspecto fundamental relacionado con la IA. Aunque los problemas relacionados con la privacidad no son productos directos de la actividad de la IA y del aprendizaje automático, sí son consecuencia de los mecanismos que hacen que ésta opere. Es decir, a partir de los datos: los datos que muchas plataformas recolectan etiquetan y clasifican a partir de la información y actividad de los usuarios y proveedores de servicios de estas.

Este tipo de procesos donde una plataforma recolecta información se llama datificación. Según van Dijck et al., (2018): "La datificación del proceso de aprendizaje, monitoreando minuciosamente el comportamiento interactivo y cognitivo de los alumnos, produce una abundancia de datos más allá de los resultados de las pruebas académicas". Esto significa que se busca extraer información de los usuarios a partir de procesos y herramientas que permitan que éste se involucre con la plataforma.

Van Dijck et al., (2018) elaboran lo anterior al comentar que la datificación y la personalización difícilmente pueden analizarse separadas del mecanismo de mercantilización o "la monetización del contenido y los flujos de datos..." (Dijck et al., 2018). Y más adelante agregan que "los algoritmos de personalización se basan en inferencias de las necesidades o intereses de los usuarios. Por tanto, un subproducto de los algoritmos de aprendizaje personalizados" (Dijck et al., 2018).

Entre los mayores problemas de la datificación se encuentra en que muchos de los usuarios desconocen para qué se están usando sus datos o si la plataforma que usan los recolecta. En este sentido "...el aprendizaje automático desafía nuestra comprensión de conceptos clave como la privacidad y el consentimiento,

arroja nueva luz sobre riesgos como estereotipos estadísticos y plantea cuestiones novedosas en torno a la interpretabilidad, la verificación y la solidez” (Royal Society, 2017). Así, uno de los elementos que requiere la IA se convierte en un problema ético, en plataformas usadas por muchas personas. Por ello, la Organización de las Naciones Unidas advirtió que:

“Desafortunadamente, muchos de los datos que se utilizan para entrenar la IA se obtienen actualmente de consumidores de todo el mundo, a menudo sin su consentimiento explícito: los países más pobres a menudo carecen de la capacidad para garantizar que los datos personales estén protegidos o para proteger a sus sociedades de los daños ciberataques y de la desinformación que ha aumentado desde la pandemia de COVID-19” (ONU/UN, 2021).

La Real Sociedad (2017) aúna en lo anterior al comentar: “El aprendizaje automático desestabiliza todavía con la distinción actual entre datos sensibles o personales y no sensibles: permite que los conjuntos de datos que al principio parecen inocuos se empleen de manera que permitan inferir lo sensible de lo mundano”. El problema de la privacidad se vuelve aún más delicado cuando surge la pregunta de cómo se usan los datos.

Uno de los ciclos más complejos de recolección de datos tiene que ver con el uso que se le da a esto para generar publicidad. Es decir, una plataforma recolecta información para alimentar su aprendizaje automático y los algoritmos ligados a éste, pero también para mercantilizarlos en aras de que algunas compañías generen publicidad dirigida específicamente a los perfiles de cada usuario. Ahora bien, para que esto pueda funcionar, se requiere que el usuario proporcione datos constantemente a partir de su interacción con la plataforma; y para ello, el usuario debe estar involucrado.

Un ejemplo en el que la privacidad se pone en constante riesgo es con el perfil de usuario de Facebook. Este perfil de usuario sirve para conocer qué le gusta a la persona, cuáles son sus acciones políticas, qué tipo de contenido le gustaría consumir más, en qué productos está interesado y hacen que compre. Así, Facebook genera una experiencia personalizada, un pequeño universo, una habitación donde las personas usuarias se sientan cómodas al consumir su

contenido y recibir noticias de aquellas personas con las que se sienten cómodas interactuar, a cambio de usar y recolectar información personal.

De esta manera, las herramientas de IA de las plataformas tratarán de promover contenido que mantenga la atención de las personas. De modo que los usuarios pasen más tiempo en alguna plataforma social, de compras, entretenimiento, entre otras; y así, generar mayor contenido debido a los datos que proporcionan directamente o a los datos que se derivan de su actividad y sus interacciones con el contenido, otros usuarios o con la misma plataforma, aunque esta actividad encierra el riesgo de generar y promover sesgos.

III. Responsabilidad

Además de los problemas relacionados con la privacidad, el acceso a la información y la discriminación, la IA plantea un problema de responsabilidad. Esto plantea preguntas como: ¿quién debe regular la IA?, ¿cuáles son las implicaciones económicas y sociales de la regulación? y ¿quién es responsable si la IA y las aplicaciones del aprendizaje profundo fallan o causan daño social o individual? Estas preguntas son sumamente complejas desde un punto de vista jurídico, social y ético, debido a problemas propios de la IA como la caja negra, la falta de imparcialidad, la privacidad, los algoritmos adversarios y los *deep fake*.

Muchos de estos problemas ya se han manifestado en casos empíricos. Sin embargo, es importante enfocarse en el problema de la responsabilidad. Es difícil determinar quién es responsable si una aplicación de IA genera una situación adversa. Por lo tanto, surge la pregunta: ¿quién debería ser responsable, el diseñador de los algoritmos de aprendizaje automático, la empresa que lo comercializa, el gobierno que no implementa la regulación adecuada o las propias herramientas de IA que actuaron de forma desconocida debido al fenómeno de la caja negra?

Con base en la anterior van Dijck et, al (2018), articulan las siguientes preguntas:

¿Quién es responsable de la distribución de las ideas de los extremistas a través de las redes sociales (Herrman 2016)? ¿Quién es responsable de la circulación de información errónea? ¿Quién es responsable cuando los ciudadanos comienzan a

pelear después de que un usuario ha distribuido un llamado a las armas? ¿Quién tiene la culpa cuando una aplicación médica que se basa en datos generados por el usuario resulta ser inexacta? ¿Quién puede ser responsable, si no responsable, cuando los niños pequeños se ven inundados con anuncios en línea de dulces recubiertos de azúcar?

Las preguntas de van Dijck et, al (2018) están dirigidas a las plataformas sociales, en especial el ecosistema controlado por Facebook (Facebook, Instagram y Whatsapp). Precisamente porque estas plataformas utilizan algoritmos, muchos con sistemas de IA, lo cual genera un ambiente que favorece el tribalismo entre sus usuarios, noticias falsas, y hasta ansiedad. Así pues, las preguntas que plantean los investigadores están ligadas a fenómenos que nacen de manera constante en plataformas que operan con instrumentos del aprendizaje de máquina profundo.

Ahora bien, en el tema de la ansiedad, en especial la ansiedad que genera Instagram en la adolescencia, el testimonio de Haugen, antigua gerente de producto en el departamento de integridad cívica de Facebook:

“Entre los documentos publicados se encuentran estudios que muestran que Instagram, que Facebook compró en 2012, afecta negativamente la salud mental de los usuarios jóvenes, particularmente las niñas. Según un estudio de Facebook filtrado por Haugen, el 13,5% de las adolescentes del Reino Unido dijeron que Instagram empeora los pensamientos suicidas. Otro estudio filtrado encontró que el 17% de las adolescentes dicen que sus trastornos alimentarios empeoraron después del uso de Instagram” (Romo, 2021).

El problema de las plataformas de Meta (Facebook) también se puede ampliar. Es decir, las preguntas se pueden hacer en torno a quién es responsable si un vehículo no tripulado decide atropellar a una persona para evitar a cinco - algo similar al problema del tranvía planteado por Philippa Foot (Renda, 2018). Así, surge la pregunta de quién es responsable si una compañía de seguros niega el seguro a una adulta mayor con base en los resultados automáticos dictados por su plataforma de aprendizaje de máquina, y así subsecuentemente.

En resumen, el problema de la responsabilidad es sumamente complejo, ya que intervienen varios actores humanos a partir de relaciones complejas que forman una red. Por ello es necesario introducir este problema para comprender y

sintetizar los conceptos y repercusiones éticas y sociales en torno a la IA, especialmente en relación con el aprendizaje de máquina profundo. Así pues, hay que examinar algunos casos empíricos en torno a la implementación y aplicación de esta tecnología en escenarios reales como Facebook (Meta) y la prevención del suicidio.

2. La Taxonomía de riesgos de la IA: por una IA confiable

Un elemento importante, más allá del aspecto ético y de los derechos humanos, es tener una taxonomía de riesgos en torno a la IA. Esto es de suma importancia, especialmente si se busca conocer el impacto más concreto y el riesgo de plataformas como Alphabet-Google, Uber y los sistemas de video vigilancia metropolitanos con IA. Además, una taxonomía de riesgos permite tanto hacer un diagnóstico más importante de los problemas con respecto a la IA, así como propuestas alrededor de una regulación prospectiva (hacia adelante) con respecto a esta tecnología.

Desde la perspectiva de este trabajo, una IA confiable o *Trustworthy AI* significa generar mecanismos regulatorios, de diseño, y de autogestión en torno a una serie de riesgos con respecto a la IA. Esto genera una pauta para abordar elementos en torno a derechos humanos como privacidad, acceso a la información y transparencia, y no discriminación; pero en torno a soluciones estructurales alrededor de elementos técnicos, socio técnicos y principios clave que sirvan como bases auditoras y reguladores. Ahora bien, no existe una definición universal de IA confiable (Newman, 2023), y tampoco una categoría constante de taxonomías de riesgos para buscar esta IA confiable y señalar los problemas de esta tecnología; no obstante, se buscará generar un marco de análisis que permita identificar y evaluar los riesgos de la IA en diferentes contextos y situaciones, así como proponer medidas para mitigarlos o prevenirlos.

De acuerdo con el *The National Institute of Standards and Technology (NIST)* (2021) existen tres grandes categorías de riesgo en torno a la IA: riesgos

asociados a atributos técnicos y de diseño, riesgos asociados alrededor de cómo se perciben los sistemas de IA, y riesgos políticos regulatorios y principios. Asimismo, estas categorías de riesgo se relacionan con los posibles impactos negativos que la IA puede tener en la sociedad, la economía, el medio ambiente, la seguridad y la ética; por ejemplo, los riesgos asociados a atributos técnicos y de diseño pueden incluir problemas de confiabilidad, robustez, transparencia, explicabilidad y sesgo.

Además, los riesgos asociados alrededor de cómo se perciben los sistemas de IA pueden involucrar cuestiones de confianza, aceptación, responsabilidad y control. Y los riesgos políticos regulatorios y principios pueden abarcar aspectos de gobernabilidad, legalidad, equidad y derechos humanos. Por lo tanto, es importante que los desarrolladores, usuarios y reguladores de la IA estén conscientes de estos riesgos y los mitiguen adecuadamente. Así pues, se podrá aprovechar el potencial de la IA para mejorar la calidad de vida de las personas, sin comprometer sus derechos, valores y principios

Con base en los riesgos asociados, hay que desglosar las características que hacen que un sistema genere confianza. La primera de estas tiene que ver con los atributos técnicos o que los riesgos técnicos se pueden gestionar mejor mediante técnicas para asegurar la validez de los métodos de aprendizaje automático (idem). Y asimismo estos riesgos pueden medirse usando estadísticas de desviación estándar (idem).

NIST plantea cuatro elementos para medir los riesgos, o alrededor de estos: Precisión, Fiabilidad, Robustez, Resiliencia o seguridad. Así pues, estos elementos se refieren a las características que deben tener los sistemas de IA para funcionar correctamente y evitar errores, fallas o ataques. La precisión se refiere a la capacidad de los sistemas de IA para producir resultados correctos y consistentes, la fiabilidad de los sistemas de IA para mantener su funcionamiento bajo condiciones normales o esperadas, la robustez de la capacidad de los sistemas de IA para resistir o adaptarse a condiciones adversas o inesperadas, y

la resiliencia o seguridad de los sistemas de IA para recuperarse o restaurarse tras una interrupción o daño.

Otra serie de elementos importantes son los atributos socio técnicos. En este sentido, los atributos tienen en común es que el grado de los riesgos correspondientes puede ser medido directamente usando algún tipo de proceso automatizable que no requiere mucha entrada humana (idem). No obstante, es necesario crear mecanismos para guiar la creación e implementación de los sistemas alrededor de los problemas técnicos y sociales que genera la IA.

De acuerdo con Blackman (2022) hay dos tipos de aproximaciones técnicas y sociotécnicas con respecto a la IA. La primera es la de crear una IA para bien, y la segunda es la IA para no mal. Según el autor, en el primer campo están quienes buscan crear herramientas con impacto social positivo; en el segundo, quienes buscan mitigar los impactos o consecuencias negativos de la IA.

A lo largo de la tesis se ha hablado con mucha extensión de los problemas de la IA con respecto a los derechos humanos; en especial, privacidad y el derecho a lo discriminación. No obstante, es necesario rescatar como estos elementos también se ajustan alrededor de una taxonomía de riesgos socio técnicos. Blackman (2022) que en torno a la IA existen tres grandes riesgos: privacidad, no discriminación o sesgos, y explicabilidad.

Los primeros dos problemas se han abordado ampliamente en la tesis. Por ello, hay que detenerse un momento para explicar, valga la redundancia, la situación en torno al problema de la explicabilidad. De acuerdo a Blackman (2022), el problema con respecto a la IA, en especial el aprendizaje de máquina y sus herramientas, es que los patrones son a menudo tan complejos o abarcan una variedad de variables tan distintas a las que normalmente prestamos atención, que muchas veces no podemos explicar por qué la IA dio la salida que dio.

Asimismo, y a nivel general, Blackman (2022) plantea que en su opinión el interés casi total en los tres grandes desafíos por parte de la comunidad de ética de la IA (salvo por las preocupaciones sobre casos de uso específicos, como el software de reconocimiento facial) ha creado un peligro en sí mismo. En este sentido, todo

ello genera una forma compleja y poco balanceada de afrontar los riesgos y esto naturalmente conduce a fallar en reconocer esos riesgos.

Otro elemento importante antes de abordar los problemas socio técnicos en torno a la IA, es la forma en que las empresas y organizaciones los perciben. La primera, de acuerdo con Blackman (2022) se da a nivel estructural en la organización o la forma en la que se buscan mitigar los riesgos y los daños. La segunda, es el contenido, o cómo entiendes esos riesgos éticos (idem).

Con base en esto, el NIST (2021) plantea como los juicios humanos en torno a los riesgos socio técnicos se dan con base en los atributos ya mencionados como explicabilidad, interpretación, privacidad, así como otros en torno a seguridad, además de formas de mitigación de los sesgos. Y estos juicios humanos se refieren a cómo las personas evalúan, comprenden, confían y usan los sistemas de IA, así como cómo se responsabilizan de sus acciones y resultados. Estos juicios humanos pueden variar según el contexto, el propósito, la experiencia y los valores de cada individuo o grupo.

El primer de estos atributos que se debe de mencionar, para no ahondar más en los tres grandes problemas de privacidad, explicabilidad y sesgos que menciona Blackman, es el de la interpretación o *Interpretability*. NIST (2021) expone que la interpretación o los intentos de aumentar la interpretabilidad buscan llenar un déficit de significado; y que la suposición subyacente es que las percepciones de riesgo provienen de una falta de capacidad para dar sentido, o contextualizar, la salida del modelo adecuadamente. Por lo tanto, la interpretación se refiere a la capacidad de los usuarios o las partes interesadas de comprender el funcionamiento, el propósito, los límites y los resultados de los sistemas de IA.

Con respecto a la interpretabilidad, Newman (2023) plantea: ¿Cómo lograremos que la incertidumbre del modelo sea más comprensible al añadir características como salidas de intervalos de confianza, predicciones probabilísticas condicionales codificadas mediante oraciones y calibración? Esta pregunta se responde con base en los planteamientos de Blackman donde los problemas de

contenido deben ser previstos en la estructura de la organización que irá planteando guías sobre la mitigación de riesgos.

Finalmente hay que distinguir entre interpretabilidad y transparencia, y que si bien estos elementos buscan darle luz a la caja negra que genera la inteligencia, ambos mecanismos lo hacen de maneras diferente forma. Esto es porque muchas veces —la transparencia no implica la interpretabilidad ya que a menudo, la interpretabilidad se relaciona con representaciones simples, mientras que la transparencia puede generar un exceso de información (NIST, 2021). Así pues, como lo plantea NIST, la interpretabilidad básicamente busca unir los puntos, de manera intuitiva, de los elementos que se desarrollan en un sistema transparente.

En términos de la teoría de actor-red, la interpretabilidad se refiere a la capacidad de los actores humanos y no humanos de comprender y comunicar el sentido y el propósito de las acciones y las decisiones de la IA. Por ende, la noción de interpretabilidad implica que los actores puedan establecer relaciones de confianza, cooperación y responsabilidad entre ellos, así como identificar y resolver posibles conflictos o contradicciones. Esto es porque la interpretabilidad actúa como un parámetro al actor humano identificar el programa que busca empujar la IA, o la intención de este programa, para decidir; por tanto, la interpretabilidad es clave para la integración social y la gobernabilidad de la IA.

El siguiente elemento digno de mencionar en los aspectos sociotécnicos es el de seguridad. El concepto de seguridad en torno a la IA responde a las preguntas de ¿Cómo garantizamos que el sistema de IA no provoque daños físicos o psicológicos o conduzca a una situación en la que la vida, la salud, la propiedad o el medio ambiente humanos estén en peligro? ¿Cómo propondremos los posibles modos de fallo o condiciones inseguras? (Newman, 2023). Además, esta noción también tiene que ver con la resiliencia que tiene un sistema a hacer atacado, y con ello, poner en peligro la integridad de los usuarios que usan estos sistemas.

Asimismo, con respecto a la noción de seguridad, la UNESCO (2023) menciona que los daños no deseados (riesgos de seguridad) y las vulnerabilidades a los ataques (riesgos de seguridad) deben evitarse y tratarse, prevenirse y eliminarse

durante todo el ciclo de vida de los sistemas de IA para garantizar la seguridad humana, ambiental y del ecosistema. Una forma de alcanzar esto es mediante el acceso a datos sostenibles y pocos tendenciosos. Pero también desde mecanismos que van desde la estructura en la que se diseña la IA, la creación de herramientas para crear sistemas de IA que no genere mal, y a través de estrategias de diseño que fomenten la IA para bien; todo esto, en torno a mecanismos de autogestión y de regulación tanto perspectiva como prospectiva.

Finalmente, como propuesta sociotécnica, está generar mecanismos para *manejar los sesgos* que se generen o se encuentren dentro de los sistemas. El problema general es que una IA puede producir resultados que tienen impactos diferenciales éticamente (y, en algunos casos, legalmente) inaceptables en diversas subpoblaciones, y esto puede ocurrir sin que los ingenieros tengan la intención de discriminar (Blackman, 2022). Una forma de abordar este problema es mediante la auditoría de la IA, que consiste en evaluar el desempeño de un sistema de IA con respecto a criterios éticos, legales y sociales, y proponer medidas correctivas en caso de detectar problemas (IA para no mal) en torno a lo que se conoce como *algorithm management* o gestión de los algoritmos desde la estructura de asociaciones que los crean; otra forma es mediante el diseño inclusivo, que busca incorporar la diversidad y la participación de los usuarios y las partes interesadas en todas las etapas del desarrollo de la IA, desde la definición del problema hasta la evaluación del impacto que entra dentro de las nociones de IA para el bien desde el diseño planteadas por investigadores como Floridi et al (2020) plantean.

Una vez establecidos los elementos técnicos, sociotécnicos y las categorías de riesgos en torno a la IA, es importante rescatar principios generales para crear una IA confiable. Estos principios se basan en las categorías de equidad, rendición de cuentas y transparencia. Y sirven como un esquema general sobre cómo se pueden regular, auditar, evaluar y diseñar los sistemas de IA, así como los datos que los alimentan.

La noción de equidad, o *fairness* en inglés, en la IA se refiere a que los actores de la IA deben fomentar la justicia social y proteger la equidad y la no discriminación

de cualquier tipo en conformidad con el derecho internacional (UNESCO, 2023). Asimismo, la categoría de equidad sirve como guía y principio para acotar problemas con respecto al diseño de los sistemas. En este sentido, NIST (2021) explica que muchos ingenieros creen que los algoritmos de aprendizaje automático son justos por naturaleza porque se aplica el mismo procedimiento sin importar el usuario; sin embargo, esta percepción se ha debilitado recientemente al aumentar la conciencia sobre los algoritmos y los conjuntos de datos sesgados.

La categoría de rendición de cuentas es una que se ha discutido a lo largo de la tesis. Especialmente, para determinar si un algoritmo o un sistema puede ser responsable del daño que este haga. No obstante, el principio de rendición de cuentas también abarca conceptos de riesgo y “culpa” - es decir, la parte responsable en caso de que se produzca un resultado riesgoso (NIST, 2021). Asimismo, el principio de rendición de cuentas se extiende a la pregunta de quién es responsable de los problemas o daños que provoque un sistema de IA.

Más allá de la caja negra en torno a los datos y el cómo los sistemas generan un resultado, también se puede hablar de una caja negra con respecto a la responsabilidad. Esto se puede explicar con el concepto de *Irresponsabilidad organizada* de Ulrich Beck (2002). De acuerdo con el autor, el riesgo del mercado global es una nueva forma de “irresponsabilidad organizada” porque es una forma institucional sin rostro que no tiene obligaciones, ni siquiera consigo misma. En el caso de las estructuras y organizaciones que crean los sistemas de la IA la irresponsabilidad organizada tiene que ver con el cada vez mayor número de actores humanos y no humanos que diseñan, comercializan, gestionan, prueban, alimentan de datos, y comercializan la IA.

La situación de irresponsabilidad organizada plantea el desafío de determinar quién es el responsable cuando un sistema de IA causa un daño o un perjuicio a una persona o a la sociedad. ¿Es el desarrollador, el proveedor, el usuario, el regulador, el sistema mismo o una combinación de todos ellos? ¿Qué criterios se deben aplicar para asignar la responsabilidad y qué mecanismos se deben establecer para garantizar la reparación? Estas son algunas de las preguntas que

se deben abordar desde una perspectiva ética, legal y social, para evitar que la caja negra de la responsabilidad se convierta en una fuente de impunidad y desconfianza.

El siguiente principio con respecto a la categoría de riesgos para fomentar una IA justa es el de la transparencia. A lo largo de esta tesis se le ha dado un lugar privilegiado a la transparencia en torno al derecho humano de acceso a la información; no obstante, es importante retomar una noción más técnica de este concepto para después aplicarlo como mecanismos de evaluación de riesgos con respecto a Alphabet Google, Uber, y los sistemas de reconocimiento facial metropolitano. Según NIST (2021), la transparencia se refiere a que un tomador de decisiones accede a la información sobre un sistema de IA cuando hace un juicio sobre él, y puede abarcar desde los datos utilizados para entrenar el modelo, la estructura del modelo, su caso de uso previsto, hasta cómo y por quién se tomaron las decisiones, cuándo, entre otras.

La UNESCO, por otra parte, plantea que la transparencia con respecto a los sistemas de IA es necesaria para que los regímenes de responsabilidad nacionales e internacionales funcionen de manera efectiva; además, una falta de transparencia podría también socavar la posibilidad de impugnar eficazmente las decisiones basadas en los resultados producidos por los sistemas de IA y podría violar el derecho a un juicio justo y a un recurso efectivo, y limitar las áreas en las que estos sistemas pueden ser usados legalmente. Eso significa, que transparencia, desde la perspectiva de una IA, es un mecanismo de contra peso. Y esto complementa la noción de la transgresión del derecho al acceso a la información, al situar a la transparencia como un problema si existe falta de ésta, pero también como una herramienta de gestión de riesgos.

Así pues, la taxonomía de riesgos de la IA aborda aspectos generales, técnicos, cómo se perciben los sistemas de IA se perciben, y guías regulatorias y de principios en torno a estos sistemas; dentro de los aspectos técnicos se encuentran la Precisión, Fiabilidad, Robustez y Resiliencia o Seguridad; dentro de los aspectos sociotécnicos resaltan la Explicabilidad, interpretabilidad, privacidad,

seguridad y gestión del sesgo; finalmente dentro de los elementos relacionados con Principios rectores que contribuyen a la confiabilidad de la IA sobresalen Equidad, rendición de cuentas y transparencia. Y con base a esta taxonomía de riesgos, se puede examinar los problemas que existen en los casos empíricos propuestos en esta tesis desde una perspectiva ética con aspectos técnicos en torno a los derechos humanos. Y, además, generar propuestas regulatorias y de gestión y de autogestión de riesgos más puntuales.

¿Cómo se instrumentaliza esta taxonomía con el propósito de esta tesis que responder en qué medida las plataformas digitales de Alphabet-Google, Uber y los sistemas de videovigilancia (C5) que utilizan IA vulneran los derechos humanos relacionados con la privacidad, transparencia e imparcialidad, y cómo estos problemas afectan la responsabilidad y aplicabilidad de las decisiones tomadas por la IA? Primero en la medida en que responden a cómo la IA genera transgresiones a nivel de derechos humanos, a nivel de asociaciones humanas en torno a la teoría de actor-red, a nivel técnico, y a nivel sociotécnico. Segundo, en la forma en que genera pautas para subsanar esas transgresiones (Imagen 9).

ELEMENTO	DESCRIPCIÓN
Taxonomía de riesgos de la IA	Clasificación de los riesgos de la IA según cuatro dimensiones: generales, técnicos, sociotécnicos y principios rectores.
Aspectos generales	Aspectos que abordan cómo se perciben los sistemas de IA, y qué guías regulatorias y de principios existen en torno a ellos.
Aspectos técnicos	Aspectos que se refieren a la calidad, confiabilidad, robustez y seguridad de los sistemas de IA.

Aspectos sociotécnicos	Aspectos que se relacionan con la explicabilidad, interpretabilidad, privacidad, seguridad y gestión del sesgo de los sistemas de IA.
Principios rectores	Aspectos que contribuyen a la confiabilidad de la IA, como la equidad, la rendición de cuentas y la transparencia.
Propósito	Responder en qué medida las plataformas digitales de Alphabet-Google, Uber y los sistemas de videovigilancia (C5) que utilizan IA vulneran los derechos humanos relacionados con la privacidad, transparencia e imparcialidad, y cómo estos problemas afectan la responsabilidad y aplicabilidad de las decisiones tomadas por la IA.
Instrumentalización	Aplicar la taxonomía de riesgos para analizar cómo la IA genera transgresiones a nivel de cada dimensión y cómo se pueden generar pautas para subsanar esas transgresiones.

(Imagen 9, elaboración propia)

Los aspectos generales de los riesgos en IA abarcan la percepción pública de estos sistemas y los marcos regulatorios que los rigen. Uno de los principales problemas en esta dimensión es la falta de normativas claras y armonizadas a nivel internacional, lo que genera un vacío legal que permite la implementación de tecnologías sin una supervisión adecuada en términos de derechos humanos (Crawford, 2021). Un caso representativo es el uso de IA en sistemas de vigilancia masiva, como los implementados en China, donde algoritmos de reconocimiento facial y análisis de comportamiento han sido utilizados para monitorear y restringir la movilidad de la población uigur, vulnerando el derecho a la privacidad y facilitando la discriminación sistemática y la represión política (Mozur, 2019).

A diferencia de Clearview AI, cuyo uso indebido afectó a ciudadanos en múltiples países sin su consentimiento, el caso chino demuestra cómo los gobiernos pueden desplegar la IA como herramienta de control social. En ciudades como Shenzhen, estos sistemas operan con escasa transparencia y sin mecanismos efectivos de rendición de cuentas, consolidando un modelo de vigilancia masivo. Básicamente en Shenzhen el modelo de seguridad se impone a costa de la privacidad y otras libertades fundamentales (TechAltar, 2025).

Los aspectos técnicos hacen referencia a la calidad, confiabilidad y seguridad de los sistemas de IA. Un problema recurrente en esta dimensión es la falta de explicabilidad de los algoritmos, lo que impide que las personas comprendan cómo se toman decisiones automatizadas que pueden afectar sus vidas. Esto ha sido especialmente problemático en el ámbito financiero y judicial. Por ejemplo, en Estados Unidos, el sistema PredPol, un sistema que analizó registros históricos de delitos y estimó, con precisión horaria, las ubicaciones con mayor probabilidad de incidencia delictiva; no obstante, la propia vigilancia genera nuevos datos que refuerzan la necesidad de mayor control policial, lo que resulta en prisiones abarrotadas de cientos de miles de personas condenadas por delitos sin víctimas. (O'Neil, 2016).

Los aspectos sociotécnicos están vinculados a cómo los sistemas de IA interactúan con la sociedad, considerando elementos como la explicabilidad, la privacidad y la seguridad. En este sentido, uno de los desafíos más grandes es la discriminación algorítmica, donde los sesgos presentes en los datos de entrenamiento se reflejan en las decisiones de los sistemas automatizados. Un ejemplo alarmante ocurrió con el sistema de reclutamiento de Amazon, que penalizaba las solicitudes de empleo de mujeres porque el algoritmo había sido entrenado con datos históricos que reflejaban una predominancia masculina en la industria tecnológica (Dastin, 2018; Trejo, 2020). Este tipo de fallos compromete el derecho a la no discriminación y a la igualdad de oportunidades en el ámbito laboral.

Finalmente, los principios rectores incluyen elementos fundamentales para la confiabilidad de la IA, como la equidad, la rendición de cuentas y la transparencia. Cuando estos principios no se cumplen, las personas afectadas por decisiones algorítmicas tienen pocas o nulas opciones para impugnarlas, lo que genera un problema de responsabilidad. Un caso emblemático es el escándalo del sistema de bienestar social en los Países Bajos, donde un algoritmo gubernamental acusó erróneamente a miles de familias de fraude en subsidios infantiles, llevando a muchas de ellas a la ruina económica sin que pudieran entender o apelar eficazmente la decisión automatizada (Eubanks, 2018). Este caso refleja cómo la falta de mecanismos de rendición de cuentas puede traducirse en graves violaciones de derechos humanos.

Para concluir, la taxonomía de riesgos de la IA permite analizar de manera estructurada cómo estos sistemas afectan los derechos humanos al considerar dimensiones clave como los aspectos generales, técnicos, sociotécnicos y los principios rectores. Al clasificar los riesgos en estas categorías, se facilita la identificación de fallas sistémicas, desde la opacidad en los algoritmos y la falta de explicabilidad hasta el uso de la IA en prácticas discriminatorias y de vigilancia masiva. Esta taxonomía no solo ayuda a diagnosticar los problemas, sino que también proporciona un marco para exigir mayor rendición de cuentas y transparencia en el desarrollo e implementación de estas tecnologías, reforzando la necesidad de marcos regulatorios que prioricen la equidad y la supervisión efectiva.

A. Más allá de los valores y regulación occidental de la IA

En este sentido, y es algo que se tiene que subrayar, Jecker y Nakazawa (2022) contrastan lo que llaman los valores de prevención de riesgos gestados en el occidente: Transparencia, Justicia, No maleficencia, Responsabilidad y Privacidad; con sistemas de valores respecto a oportunidades o armonía propuestos por Jobin et al. (2019) y que se consideran minoritarios: Benevolencia, Libertad y Autonomía, Sustentabilidad, Dignidad y Solidaridad. De acuerdo con las autoras,

estos seis valores se encuentran insertos en las políticas de implementación y rutas de IA japonesas. Además, estas perspectivas de la implementación de la IA en Japón, más allá de la prevención de riesgos, están estrechamente relacionadas con la perspectiva histórica, cultural, económica y religiosa del país.

Jecker y Nakazawa (2022), en su texto, plantean una perspectiva y de regulación mixta para hacer que la implementación y la vigilancia de los sistemas de IA sea verdaderamente internacional. Esta propuesta busca integrar los valores y principios éticos de las culturas orientales y occidentales, respetando la diversidad y la autonomía de cada sociedad. Y explican de manera puntual el por qué es importante hacer un comparativo entre las perspectivas regulatorias, pero también entre los valores de la IA entre latitudes diferentes.

En este sentido, el objetivo también se basa en crear mecanismos que no sólo prevengan la transgresión de los sistemas de la IA, sino también armonicen estos desde una perspectiva realmente universal, y también local como es la de México. Así pues, se puede apreciar cómo Japón ofrece un ejemplo de cómo incorporar la IA en la sociedad, no como una amenaza, sino como un actor que puede aportar beneficios y armonía, siempre que se respeten los valores y la dignidad humana.

Esta visión refleja la influencia de sus tradiciones religiosas y filosóficas, especialmente el shintoísmo y el budismo, que enfatizan la interdependencia y la animación de todos los seres (Honda, 2019). Ahora bien, con base en las metas de Saltos Lunares o Moonshot Goals, Jecker y Kanazawa (2022) plantean que hay una relación entre las intenciones de la implementación de la IA en Japón y sus valores positivos o de armonía en torno a estos. Así pues, las investigadoras plantean una relación entre estos saltos lunares y la forma en la que se pueden obtener beneficios.

El valor de beneficencia se asocia con el florecimiento humano y el bienestar, que son los beneficios que se esperan obtener con el objetivo del salto lunar número 2, que consiste en “Eliminar la mortalidad prematura prevenible” mediante el uso de la IA y la robótica, así como los avatares cibernéticos. Mientras que el valor de libertad/autonomía se asocia con la libertad de elegir las actividades que las

personas desean realizar, lo cual se relaciona con el objetivo del salto lunar número 3, que consiste en “Crear una órbita autónoma” mediante el uso de la IA y la robótica.

Por otro lado, el valor de confianza se asocia con las relaciones humanas positivas e inclusivas, lo que se vincula con el objetivo del salto lunar número 4, que consiste en “Crear una armonía entre la humanidad y la sociedad de las máquinas” mediante el uso de la IA y la robótica. Este objetivo refleja la visión de Japón de crear una sociedad que fomente la confianza, la cooperación y la armonía entre las personas y las máquinas. El valor de sostenibilidad se asocia con la protección ambiental y el desarrollo sostenible, lo que se conecta con el objetivo del salto lunar número 6, que consiste en “Capturar el CO₂, reforestar con biodiversidad, incorporar plásticos libres de basura y crear un sistema hiperenergético para los países altamente desarrollados” mediante el uso de la IA y la robótica.

Asimismo, el valor de dignidad se asocia con el apoyo a las capacidades humanas básicas, lo que se corresponde con el objetivo del salto lunar número 7, que consiste en “Superar las enfermedades principales y lograr una buena salud y longevidad” mediante el uso de la IA y la robótica. Finalmente, el valor de solidaridad se asocia con el considerar a los robots como parte de las relaciones valoradas y de una buena sociedad, lo que se relaciona con el objetivo del salto lunar número 8, que consiste en “Crear robots en las familias y los entornos naturales” mediante el uso de la IA y la robótica. Este objetivo refleja la visión de Japón de crear una sociedad que promueva la solidaridad, la inclusión y la convivencia entre las personas y los robots.

Finalmente, Jecker y Kanazawa (2022) proponen implementar los valores de prevención junto con los valores de beneficios de la IA. Esto es, crear una aproximación más holística entre lo que se llama IA para bien e IA para no mal. Así pues, se plantea una perspectiva donde se encuentren elementos para prevenir los riesgos o las transgresiones de los sistemas con IA, como (1) transparencia, (2) justicia/equidad, (3) no maleficencia, (4) responsabilidad y (5)

privacidad; y para lo que en este trabajo se llamarán valores de armonización: (6) beneficencia, que se refiere al florecimiento humano y el bienestar; (7) libertad/autonomía, que se refiere a la libertad de elegir las actividades que las personas desean realizar; (8) confianza, que se refiere a las relaciones positivas entre las personas y los robots; (9) sostenibilidad, que se refiere a la protección ambiental y el desarrollo sostenible; (10) dignidad, que se refiere al apoyo a las capacidades humanas básicas; y (11) solidaridad, que se refiere a la incorporación de los robots en las relaciones valoradas y en una buena sociedad

B. La Declaración Universal de los Derechos Humanos y la IA

La Declaración Universal de los Derechos Humanos es un instrumento jurídico internacional. Este instrumento contiene un conjunto de normas que establece las condiciones mínimas para que las normativas nacionales de los Estados miembro respeten y garanticen la dignidad y la libertad de los seres humanos. La Declaración Universal de los Derechos Humanos ha enfrentado obstáculos y desafíos que han afectado su aplicación, además de la necesidad de adaptar su contenido a contextos y situaciones tan complejas y diversas como los derechos humanos universales.

Con base en la metodología y el marco teórico, se estudiará el concepto de los derechos humanos establecido en la Declaración Universal de los Derechos Humanos de 1948 como objeto de contorno y los derechos humanos como programa. Asimismo, con base en las herramientas de la teoría actor-red, se analizará el origen de la Declaración y cómo esta surgió a partir de la disputa de diferentes actores que intentan colocar sus propios programas y antiprogramas en el documento.

Ahora bien, antes de comenzar el análisis, es importante responder tres preguntas. La primera tiene que ver con qué es la Declaración universal de los derechos humanos. Mientras que la segunda pregunta intentará responder cómo funciona esa Declaración, es decir, la manera en la que se aplica a modo de programa en distintas situaciones; y la tercera pregunta tiene que ver con las implicaciones éticas que contiene la Declaración en torno los derechos

transparencia y *explicabilidad*, rendición de cuentas y responsabilidad, y derecho a la intimidad y protección de datos.

La Declaración de los derechos humanos es un documento con principios inalienables para los seres humanos. Asimismo, es un actor que busca la universalidad de estos principios elevados al grado de derechos, a la figura que indica que todas las personas deben de tenerlos, respetarlos y protegerlos. Sin embargo, la Declaración también tiene un estatus de universalidad; es decir, contempla que estos derechos deben de ser respetados en todas partes del mundo, en todas las culturas, y en todas las latitudes, lo cual encierra un problema de traducción.

En cuanto a los principios éticos, Mark considera que: “Los derechos humanos constituyen un conjunto de normas que rigen el trato de individuos y grupos por parte de actores estatales y no estatales sobre la base de principios éticos en lo que la sociedad considera fundamental para una vida digna” (Marks, 2014). Esto quiere decir que los derechos humanos son normas, en palabras de Latour programas, que están ligadas a principios éticos. Y estas normas buscan tener un carácter elemental.

Hay que resaltar que, para nuestro análisis, se usará la noción de norma a modo de código en aras de generar una traducción entre los actores no humanos y los actores humanos. De acuerdo con Thevenot, un código es un término empleado en una serie de textos para hacer valer la ley; por ello, es normal escuchar en disertaciones jurídicas que “de acuerdo con el código civil esto, de acuerdo al código penal aquello, entre otros ejemplos”. Sin embargo, Thevenot amplía el concepto de código al mencionar que: “...el término *código* también se utiliza para referirse a cualquier conjunto de reglas de principio en materia de moralidad o de honor, y esto demuestra que, junto al código civil o penal, existen códigos menos formalizados y, a veces, ni siquiera por escrito, cuyos efectos no están tan cuidadosamente controlados”.

Así pues, la Declaración universal de los derechos humanos de 1948 es un código que busca reconocer los derechos más elementales que se deben respetar en las

personas bajo una serie de principios. Así pues, la Declaración como actante encierra un código que después funciona como programa para proteger ciertos elementos de la vida, así como condiciones, de la especie humana. Por lo tanto, la declaración de derechos humanos se puede comparar con el concepto de Boltanski y Thevenot de *equivalencia*, es decir, un objeto que contiene una serie de códigos dentro de sí mismo, como los hilos de un paraguas que, juntos, protegen de la lluvia, aunque por sí solos sean insuficientes.

Hay que resaltar que el grado de universalidad de la Declaración de los derechos humanos se puede observar al inicio de esta:

“Proclama la presente Declaración Universal de Derechos Humanos como ideal común por el que todos los pueblos y naciones deben esforzarse, a fin de que tanto los individuos como las instituciones, inspirándose constantemente en ella, promuevan, mediante la enseñanza y la educación, el respeto a estos derechos y libertades, y aseguren, por medidas progresivas de carácter nacional e internacional, su reconocimiento y aplicación universales y efectivos, tanto entre los pueblos de los Estados Miembros como entre los de los territorios colocados bajo su jurisdicción” (Comisión de los derechos humanos, 1948).

La característica universal de los humanos hace que este actor requiera puntuales trayectorias. Dado que se desprenden las preguntas: ¿qué son los derechos humanos o que significan para naciones bajo el corte de occidente y oriente? Es decir, cómo se puede conciliar un concepto, un grupo de instrucciones que garanticen o velen las condiciones básicas de dignidad de las personas bajo distintas latitudes, sociedades y primicias éticas.

Este tipo de observaciones son retomados por autores como Papacchini (1997), quien plantea: “Los derechos humanos responden a exigencias humanas universales. Sin embargo, en su configuración específica son un producto de la modernidad”. La postura de Papacchini está ligada a que la Declaración fue concebida en un momento puntual histórico de la humanidad, bajo una serie de valores, de principios, que se articulan bajo los ideales de un tipo de sociedad en específico.

Santiago Nino considera que los derechos humanos son producto de la modernidad: “Es indudable que los derechos humanos son uno de los más grandes inventos de nuestra civilización” (Santiago-Nino, 1989). Asimismo, en

cuanto discusión sobre por qué los derechos humanos son considerados universales, el autor alude a que pueden considerarse como parte del orden *iusnaturalista*. Dado que, “que hay principios que determinan la justicia de las instituciones sociales y establecen parámetros de virtud personal que son universalmente válidos independientemente de su reconocimiento efectivo por ciertos órganos o individuos” (Santiago-Nino, 1989).

Ahora bien, existen autores que plantean que los derechos humanos mantienen un resquicio de la perspectiva *iusnaturalista*. Por ejemplo, Schabas expone que: “Desde tal perspectiva, los derechos fundamentales y la protección de la dignidad humana son y han sido siempre nuestro derecho. Siguiendo esta perspectiva de derecho natural, la Declaración Universal de los Derechos Humanos es más una codificación de los principios existentes que un ejercicio de legislación creativa” (Schabas, 2013). Ahora bien, como lo explica el propio Schabas, “...ya sea que la inclinación filosófica de uno sea hacia la ley natural o sea más positivista, el proceso mediante el cual se identificaron las normas fundamentales durante 1947 y 1948 es de profundo interés”; por ende, es importante examinar los procesos de traducción con base en la Declaración de los derechos humanos de 1948 para comprender no solamente su impacto, sino su contenido y su rol dentro de los principios de la IA.

En resumen, la Declaración de los derechos humanos de 1948 es un conjunto de principios éticos o códigos. Y estos principios en parte se pueden argumentar como naturales o inherentes a la vida de cualquier persona desde el momento en que nace. Aun así, estos principios se codificaron con la esperanza de que su alcance sea universal; y con ello, puedan insertarse en la vida de los humanos bajo su premisa de universalidad.

Al examinar directamente la Declaración, existe un artículo que reconoce el derecho a la privacidad. Y este es el artículo doce que plantea: “Nadie será objeto de injerencias arbitrarias en su vida privada, su familia, su domicilio o su correspondencia, ni de ataques a su honra o a su reputación. Toda persona tiene

derecho a la protección de la ley contra tales injerencias o ataques” (Asamblea de las Naciones Unidas, 1948).

De esta manera la Declaración Universal de Derechos Humanos de 1948 establece que todas las personas tienen derecho a la privacidad y al respeto de su vida privada. Esto adquiere importancia por la cantidad de datos personales que se recopilan y procesan con esta tecnología. Por ende, es importante que se respete el derecho a la privacidad de las personas y se garantice que se proteja adecuadamente la información personal y se utilice de manera responsable. Esto incluye medidas de seguridad para evitar el acceso no autorizado o el uso indebido de estos datos, datos que cobijan la esencia de los usuarios, su alma, su código hecho información; por ejemplo, se alienta a las empresas a encontrar soluciones técnicas para asegurar y proteger la confidencialidad de las comunicaciones digitales, como la codificación y el anonimato. Asimismo, se insta a los Estados a no interferir en el uso de estas soluciones y a que cualquier restricción se ajuste a sus obligaciones según el derecho internacional de los derechos humanos (A/HRC/RES/42/15) (2019).

El siguiente principio de derecho que es importante examinar es el de acceso a la información y su vínculo con el valor de la transparencia. Según Van Dijck et al. (2018), "la transparencia es otro valor público urgente que se aplica no solo a los flujos de datos y algoritmos, sino también a los modelos comerciales y el estado de propiedad de las plataformas". Esto tiene que ver con la forma en que operan las plataformas digitales de trabajo, que requieren una gran cantidad de datos y herramientas de IA. Sin embargo, muchas empresas que crean estas plataformas y se autorregulan obstaculizan el acceso para responder preguntas sobre cómo se utilizan estos datos o cómo funcionan sus propias herramientas que utilizan tecnologías de IA.

La transparencia es esencial para garantizar que las empresas utilicen los datos y las herramientas de IA de manera responsable y ética. Por ende, sin acceso a información clara y detallada sobre cómo se recopilan y utilizan estos datos, es difícil para los usuarios entender cómo sus decisiones y acciones pueden ser

influenciadas por estas tecnologías. Además, la falta de transparencia puede dar lugar a abusos de poder y a violaciones de derechos, como la privacidad y la protección de datos personales con lo que el acceso a la información se vuelve fundamental para proteger otros derechos derivados de los problemas sociales de la IA.

El derecho de acceso a la información se deriva del artículo 19 de la Declaración Universal de los Derechos Humanos. En este se establece que: “Todo individuo tiene derecho a la libertad de opinión y de expresión; este derecho incluye el de no ser molestado a causa de sus opiniones, el de investigar y recibir informaciones y opiniones, y el de difundirlas, sin limitación de fronteras, por cualquier medio de expresión” (Declaración Universal de Derechos Humanos, 1948). Esto implica que todas las personas tienen derecho a buscar, recibir y difundir información y opiniones libremente, sin ser obstaculizadas o sancionadas por ello; y el derecho a buscar información es especialmente relevante ante los problemas de opacidad en torno a la IA.

Además, en el contexto de la IA, el acceso a la información es esencial para el desarrollo y el uso responsable de esta tecnología. Y dado que IA se basa en el procesamiento y el análisis de grandes cantidades de datos, el acceso a información diversa y de alta calidad es fundamental para entrenar y mejorar los modelos de IA. De acuerdo con el reporte de *Privacy and Freedom of Expression In the Age of Artificial Intelligence* de Privacy International y ARTICLE 19 (2016 en Del Pozo, Claudia May, Ferreira, Rodrigo, Martín del Campo, Ana Victoria, et al. 2020): “la transparencia y el acceso a la información, tanto de organismos públicos como privados, son cruciales para garantizar la gobernanza democrática. Cada vez más, las decisiones sobre el acceso a la información, tradicionalmente tomadas por humanos, ahora se basan en aplicaciones de IA”.

El acceso a la información es un componente importante para abordar el problema de la responsabilidad, la falta de transparencia y el problema de la caja negra en la IA. Particularmente, porque IA se basa en el procesamiento y el análisis de grandes cantidades de datos, y el acceso a información diversa y de alta calidad

es fundamental para entrenar y mejorar los modelos de IA. No obstante, si la información utilizada para entrenar y evaluar los sistemas de IA es parcial o sesgada, esto puede tener consecuencias negativas en la toma de decisiones y en el impacto de la IA en la sociedad; en este sentido: “La transparencia es una condición para la rendición de cuentas, pero el mundo en línea se está volviendo más opaco día a día” (Dijck et al., 2018).

Ahora bien, esto no quiere decir que el principio de la transparencia en la IA no haya entrado en la agenda de las Naciones Unidas. Con base en los foros conocidos como *AI for good*⁷, Garcia (2021) explica que “...el secretario general de Naciones Unidas recuerda la necesidad de colaborar estrechamente para superar los desafíos y conciliar los intereses, especialmente en las esferas de la privacidad y los derechos humanos, la ética, la igualdad y la equidad, la soberanía y la responsabilidad, la transparencia y la rendición de cuentas”.

Así pues, es necesario señalar que el derecho del acceso a la información se ha comenzado a reconocer dentro de distintos circuitos como el académico, el de las organizaciones civiles, e incluso en organismos internacionales como lo es la Organización de las Naciones Unidas. Este derecho también se discute en estos circuitos. Y dados los problemas éticos que se desprenden de la IA, poco a poco se ha ido reconociendo su relevancia.

El siguiente tema en torno a la IA que vamos a examinar es el derecho a la no discriminación. Van Dijck et al. (2018) argumentan: “Una sociedad que se aferra a los principios de igualdad, inclusión y trato justo sin favoritismo ni discriminación debe poder evaluar las decisiones automatizadas tomadas por algoritmos que ahora están ocultos al escrutinio público”. De esta manera, se puede ver que el principio de imparcialidad está relacionado con otros términos y conceptos importantes como la equidad y, especialmente, la no discriminación.

⁷ “El objetivo de AI for Good es identificar aplicaciones prácticas de IA para promover los Objetivos de Desarrollo Sostenible de las Naciones Unidas y escalar esas soluciones para lograr un impacto global. Es la principal plataforma global e inclusiva de las Naciones Unidas sobre IA orientada a la acción” (International Telecommunication Union, 2022).

De hecho, el derecho a la no discriminación o el concepto de este se encuentra en la Declaración de Derechos Humanos de 1948 en los artículos 7 y 10. El artículo 7 se centra en la no discriminación, mientras que el artículo 10 se centra en la equidad. El artículo 10 menciona: "Toda persona tiene derecho, en condiciones de plena igualdad, a ser oída públicamente y con justicia por un tribunal independiente e imparcial, para la determinación de sus derechos y obligaciones o para el examen de cualquier acusación contra ella en materia penal" (Asamblea de las Naciones Unidas, 1948)

Todos son iguales ante la ley y tienen, sin distinción, derecho a igual protección de la ley. Todos tienen derecho a igual protección contra toda discriminación que infrinja esta Declaración y contra toda provocación a tal discriminación" (Asamblea de las Naciones Unidas, 1948). Esto significa que, desde que se publicó la Declaración, es un principio como la imparcialidad que busca llamar la atención sobre los problemas de discriminación relacionados con la IA, especialmente en los casos aprendidos. Por tanto, se puede partir de una base importante al discutir la noción de imparcialidad y las consecuencias de la no discriminación, ya que no debe agregarse a las consideraciones éticas y de derechos humanos planteadas por las Naciones Unidas.

Finalmente, hay que hablar del problema de la responsabilidad en torno a la IA. La responsabilidad es un valor o principio que debe de regir la IA. Esto es, porque, así como el acceso a la información, se debe exigir un grado de obligación a los actores dueños de las plataformas que operan con las herramientas de la inteligencia, así como a los creadores y diseñadoras de dichas herramientas; además de; "Idealmente, la creación de valor público para el bien común debería ser responsabilidad compartida de todos los actores sociales..." (Dijck et al., 2018).

El derecho de responsabilidad no existe en la Declaración de los derechos humanos de 1948. Pero la Declaración de Derechos Humanos no menciona explícitamente los problemas de responsabilidad en el contexto de la IA; no obstante, sí establece principios fundamentales que son relevantes para discutir la responsabilidad en este ámbito. Por ejemplo, el derecho a un juicio justo y a una

protección adecuada contra la discriminación pueden ser aplicables a decisiones automatizadas tomadas por algoritmos y pueden servir como guía para considerar cómo deben abordarse los problemas de responsabilidad en este contexto.

Aun así, el tema de la responsabilidad ha sido objeto de intensos debates en diferentes foros, en la sociedad civil y en la investigación sobre IA. Por ejemplo, debido a que la IA "perpetúa fácilmente los patrones existentes de prejuicio y discriminación, ya que la forma más común de implementar estos sistemas es entrenarlos para que reproduzcan los resultados obtenidos por los tomadores de decisiones humanos" (Raso et al., 2018). Así pues, se han generado intensos debates sobre la necesidad de justicia, responsabilidad y transparencia en la IA, especialmente en el aprendizaje automático (idem).

Basándonos en estos debates, podemos plantear las siguientes preguntas: ¿quién es responsable de lo que haga la IA: la IA en sí misma. ¿Las personas que la diseñaron, las personas que la entrenaron, las plataformas que la utilizan o los usuarios? Estas preguntas también están relacionadas con un sesgo de confianza en la tecnología (que no puede equivocarse). Y en relación con los derechos humanos, Raso et al. (2018) han planteado lo siguiente: "Sin duda, los Principios Rectores dejan en claro que la diligencia debida en materia de derechos humanos es una responsabilidad continua precisamente por esta razón: no todos los impactos pueden predecirse, incluso con una diligencia razonable".

Para concluir, no todos los problemas propuestos en esta tesis en torno a la IA se encuentran en la Declaración de los derechos humanos. No obstante, sí se encuentran los derechos fundamentales en torno a las controversias de esta tecnología: no discriminación, acceso a la información y privacidad. Además, la discusión con respecto a estos derechos abre un debate importante entre distintos actores para que estos se incorporen a marcos regulatorios incluidos anclados a organizaciones como las Naciones Unidas.

3. Marco regulatorio de la Unión Europea en torno a la IA

Es importante examinar la propuesta del Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de IA (Ley de IA) en relación con los derechos humanos y cómo esa propuesta se compara con la regulación de la IA en Japón y México. Esto permite comparar y analizar cómo distintos países y regiones abordan la regulación de la IA respecto a los derechos humanos, identificar similitudes y diferencias entre las regulaciones y cómo afectan a su implementación, encontrar posibles mejoras en cada regulación y promover una mayor armonización internacional en cuanto a los derechos humanos y la IA.

De acuerdo con el Parlamento Europeo y Consejo (2021), el objetivo del reglamento es “preservar el liderazgo tecnológico de la Unión y garantizar que los europeos puedan aprovechar nuevas tecnologías que se desarrollen y funcionen de acuerdo con los valores, los derechos fundamentales y los principios de la UE”. Muchos de estos valores giran en torno a los derechos humanos. Además de los problemas sociales causados por la IA.

Desde 2017, la Comisión Europea ha realizado varias investigaciones para determinar el impacto de la IA. En 2020, las conclusiones de estas investigaciones instaron a abordar "la opacidad, la complejidad, el sesgo, cierto grado de imprevisibilidad y un comportamiento parcialmente autónomo de ciertos sistemas de IA, para garantizar su compatibilidad con los derechos fundamentales y facilitar la aplicación de las normas jurídicas" (Parlamento Europeo y Consejo, 2020). Esto fue acompañado por una serie de resoluciones en 2019 sobre ética, responsabilidad civil y propiedad intelectual.

Ahora bien, más allá del contexto en el que surgió la propuesta del Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de IA (Ley de IA o el *AI act*), es importante conocer sus objetivos. El Parlamento Europeo y Consejo (2021) los resumen en cinco puntos:

- i. Garantizar que los sistemas de IA introducidos y usados en el mercado de la UE sean seguros y respeten la legislación vigente en materia de derechos fundamentales y valores de la Unión;
- ii. Garantizar la seguridad jurídica para facilitar la inversión e innovación en IA;
- iii. Mejorar la gobernanza y la aplicación efectiva de la legislación vigente en materia de derechos fundamentales y los requisitos de seguridad aplicables a los sistemas de IA;
- iv. Facilitar el desarrollo de un mercado único para hacer un uso legal, seguro y fiable de las aplicaciones de IA y evitar la fragmentación del mercado.
- v. Crear Salvaguardas en torno la IA de uso general (*general purpose AI*)

En este sentido hay que resaltar como La Comisión Europea (2023) propone una regulación de la Inteligencia Artificial en torno a niveles de riesgos de estos sistemas. Esto facilita los aspectos regulatorios en un esquema holístico donde no se considera un sólo tipo de plataformas de inteligencia artificial; por ende, un sólo tipo de riesgos que estas pueden tener. Se pueden resumir las categorías de riesgo de la siguiente manera (idem):

- Riesgos mínimos: “Las aplicaciones de riesgo mínimo, tales como los sistemas de recomendación o los filtros de correo no deseado basados en la inteligencia artificial, disfrutarán de libertad y ninguna obligación al respecto”. Esto es, porque estos sistemas representan un riesgo mínimo.
- Alto riesgo: “Algunos ejemplos de estos sistemas de inteligencia artificial de alto riesgo son determinadas infraestructuras críticas, por ejemplo, en los ámbitos del agua, el gas y la electricidad; productos sanitarios...”.
- Riesgo inadmisibles: “Esto abarca los sistemas o las aplicaciones de inteligencia artificial que manipulen el comportamiento humano para eludir la voluntad de los usuarios...”.
- Riesgo específico de transparencia: “Al utilizar sistemas de inteligencia artificial como robots conversacionales, los usuarios deben ser conscientes de que están interactuando con una máquina” Dentro de estos modelos se encuentran también los *deep fakes*, falsos profundos o ultra falsificaciones.

Además, el Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de IA (Ley de IA) (2020/2021) también propone una definición única de la IA. En concreto, se propone una definición de IA en relación con los sistemas de IA. Esta definición se puede leer de la siguiente manera:

“«Sistema de IA (sistema de IA)»: el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa” (ídem).

La Unión Europea ya cuenta con marcos regulatorios que protegen a las personas en torno a problemas relacionados con la privacidad, el acceso a la información y el derecho a la no discriminación. Estos marcos normativos también se reflejan en corpus regulatorios que han inspirado propuestas jurídicas como la japonesa, como el Libro Blanco. Y dentro de las protecciones en torno a la privacidad, el acceso a la información y el derecho a la no discriminación en la Unión Europea, se pueden subrayar las siguientes (aunque hay que tener en cuenta que estas son algunas de las consideraciones que tiene la Unión Europea en torno a estos aspectos):

- i. El Tratado de Funcionamiento de la Unión Europea (TFUE), que establece los derechos fundamentales de los ciudadanos de la UE, incluyendo el derecho a la protección de datos personales (artículo 8).
- ii. El Reglamento General de Protección de Datos (RGPD), que establece normas específicas sobre la protección de datos personales en el contexto de la UE.
- iii. La Directiva de Servicios de la Sociedad de la Información (DSSI), que establece normas sobre el acceso a servicios de la sociedad de la información, incluyendo el derecho a la información.
- iv. La Directiva sobre el Acceso a la Información Pública, que establece normas sobre el acceso a la información pública, incluyendo el derecho a la información.

- v. La Directiva sobre el Acceso a la Justicia en Asuntos de Discriminación, que establece normas sobre el acceso a la justicia en casos de discriminación, incluyendo el derecho a la no discriminación (AI Act, 2024).

Un documento importante en torno al tema más específicos de los derechos a la privacidad, acceso a la información, y el derecho a la discriminación es *El Libro Blanco de la IA de la Comisión Europea* (2020). Particularmente, porque este documento es una de las bases de la propuesta regulatoria japonesa, así como de otras propuestas regulatorias en torno a la IA. El Libro Blanco resalta la importancia de proteger la privacidad de los ciudadanos de la UE al implementar y desarrollar tecnologías de IA.

Entre los aspectos más importantes del *El Libro Blanco de la IA de la Comisión Europea* (2020), se puede resaltar:

- Garantizar que se cumplan los principios de privacidad por diseño y privacidad por defecto, así como establecer medidas de seguridad adecuadas para proteger los datos personales de los usuarios.
- Garantizar el acceso a la información y la transparencia en el uso de la IA.
- Garantizar el derecho a la no discriminación en el uso de la IA. Esto incluye asegurar que la IA no se utilice de manera discriminatoria o injusta hacia ciertos grupos de personas, y garantizar que se tomen medidas para evitar cualquier forma de discriminación en el uso de la IA (Comisión Europea, 2020).

Con base en estos planteamientos, es importante examinar cómo el Reglamento responde a los problemas relacionados con la privacidad, acceso a la información, y no discriminación, además de rendición de cuentas y responsabilidad en torno a la IA. Esto es, para averiguar si estos elementos son mencionados y discutidos en el documento, y para tener un marco comparativo con las legislaciones y propuestas regulatorias que se inspiran en el marco europeo, como la japonesa.

En el caso de la privacidad, el Parlamento Europeo y Consejo (2021) en su artículo 10, sección 5 plantea:

“En la medida en que sea estrictamente necesario para garantizar la vigilancia, la detección y la corrección de los sesgos asociados a los sistemas de IA de alto riesgo, los proveedores de dichos sistemas podrán tratar las categorías especiales de datos personales que se mencionan en el artículo 9, apartado 1, del Reglamento (UE)2016/679; el artículo 10 de la Directiva (UE)2016/680, y el artículo 10, apartado 1, del Reglamento (UE) 2018/1725, ofreciendo siempre las salvaguardias adecuadas para los derechos y las libertades fundamentales de las personas físicas, lo que incluye establecer limitaciones técnicas a la reutilización y la utilización de las medidas de seguridad y protección de la privacidad más recientes, tales como la seudonimización o el cifrado, cuando la anonimización pueda afectar significativamente al objetivo perseguido.”. X

Ahora bien, la versión actualizada del Artificial Intelligence Act (2024) plantea: “Por «gestión de la privacidad y de los datos» se entiende que los sistemas de IA se desarrollan y utilizan de conformidad con normas en materia de protección de la intimidad y de los datos, al tiempo que tratan datos que cumplen normas estrictas en términos de calidad e integridad” (AI Act, 2024). El Reglamento también se enfoca en garantizar el acceso sin sesgos a los datos y en proteger la información confidencial en el ámbito de la salud. Esto significa que no solo aborda los sesgos en la IA, sino también el acceso a los datos para desarrollar esta tecnología de manera responsable; además, el Reglamento respeta el principio de privacidad en los derechos humanos en su artículo 10.

Con respecto al acceso a la información, Parlamento Europeo y Consejo (2021) se encuentra en los siguientes artículos:

- Artículo 13, sección 2: Parlamento Europeo y Consejo (2021) establecen que los usuarios tienen derecho a recibir información clara y completa sobre el funcionamiento de los sistemas de IA que se consideren de alto riesgo (Artículo 13, sección 2).
- Artículo 14, sección 4: Este artículo establece que los usuarios tienen derecho a acceder a la información que se utiliza para el análisis de datos masivos que afecta a sus derechos y libertades (Artículo 14, sección 3).
- Artículo 17 contempla el derecho al olvido, en donde establece que los titulares de datos tienen derecho a solicitar la supresión de sus datos personales cuando los mismos ya no son necesarios para los fines para los

que fueron recogidos, cuando el titular se retira del consentimiento que dio para su tratamiento, o cuando se opone al tratamiento de sus datos. X

Ahora bien, de manera más concreta, el derecho a la no discriminación aparece en el artículo 15, sección 4 que menciona:

“Los sistemas de IA de alto riesgo que continúan aprendiendo tras su introducción en el mercado o puesta en servicio se desarrollarán de tal modo que los posibles sesgos en la información de salida debidos al uso de esta como datos de entrada en futuras operaciones («bucle de retroalimentación») se subsanen debidamente con las medidas de mitigación oportunas” (ídem).

Además, el reglamento ya aprobado por el parlamento de la Unión Europea contempla los riesgos discriminatorios con respecto a los procesos de contratación laboral debido a los sesgos aprendidos por los sistemas. En este sentido, la propuesta explica que esto se refleja en “las mujeres, ciertos grupos de edad, personas con discapacidad o personas de orígenes raciales o étnicos concretos o con una orientación sexual determinada” (ídem). Y por ende hace hincapié en que los sistemas de IA de alto riesgo deben ser transparentes en torno a su funcionamiento, y contener herramientas para evitar prácticas o predecir procesos discriminatorios.

Es importante resaltar que la versión final del AI Act (2024) establece siete principios éticos no vinculantes para garantizar el funcionamiento ético de los sistemas y plataformas de inteligencia artificial. Estos siete principios son “acción y supervisión humanas; solidez técnica y seguridad; gestión de la privacidad y de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental; y rendición de cuentas” (AI Act, 2024). Cabe destacar que, como se ha visto a lo largo de la tesis, estos principios coinciden con los problemas de derechos humanos en los que se centra este trabajo, así como con las preocupaciones más importantes respecto a los sistemas de inteligencia artificial planteadas por Blackman; es decir, los problemas del derecho a la privacidad, el acceso a la privacidad o transparencia, y el derecho a la no discriminación y diversidad.

Ahora bien, comparar los marcos regulatorios en torno a la IA de México y Japón con base en la Unión Europea puede ser útil para identificar similitudes y diferencias en cómo cada una de estas regiones o países aborda el desarrollo y aplicación de la IA, así como para evaluar las fortalezas y debilidades de cada enfoque. Esto puede ayudar a entender cómo estos marcos regulatorios se ajustan a las necesidades y contextos específicos de cada región o país. Y también para identificar posibles lecciones que se puedan aprender de uno a otro a modo de espejo.

Comparar estos marcos regulatorios con base en la Unión Europea puede ser útil para evaluar cómo cada uno de ellos aborda los derechos humanos y los problemas sociales relacionados con la IA; particularmente Japón ha armonizado sus propuestas regulatorias con la EU. La Unión Europea es conocida por tener una de las estrategias de IA más ambiciosas y avanzadas en términos de protección de los derechos humanos y la privacidad. Finalmente, comparar los marcos regulatorios de México y Japón, es especialmente útil para identificar áreas en las que es necesario mejorar o en las que se pueden adoptar enfoques similares a los de la Unión Europea para garantizar un desarrollo y aplicación responsable y beneficiosa de la IA para toda la sociedad.

Ahora bien, el *AI Act* europeo adopta un enfoque híbrido que combina elementos del derecho de protección de datos y del derecho regulatorio, pero sin establecer derechos subjetivos claros para las personas afectadas por sistemas de inteligencia artificial (Sousa e Silva, 2024). En particular, la falta de disposiciones específicas para la impugnación de decisiones automatizadas y la limitada transparencia en la lógica de los sistemas sugiere que el *AI Act* no logra un equilibrio efectivo entre innovación y protección de derechos fundamentales. De este modo, se mantiene la ambigüedad en áreas clave, dejando la implementación sujeta a interpretaciones que podrían debilitar su aplicación efectiva.

Asimismo, la clasificación de riesgos basada en el uso previsto de los sistemas de IA introduce excepciones que generan inconsistencias regulatorias. Por ejemplo, los sistemas de identificación biométrica a distancia se consideran de alto riesgo,

pero se permite su uso para la verificación de identidad, lo que abre la puerta a aplicaciones masivas que podrían vulnerar la privacidad (Sousa e Silva, 2024). A esto se suma la dificultad de evaluar el impacto real de ciertos sistemas, ya que la regulación permite que algunos eviten la etiqueta de alto riesgo si no influyen “materialmente” en la toma de decisiones, un criterio vago y susceptible de explotación.

Además, la falta de una definición clara de inteligencia artificial en el *AI Act* refleja una tensión entre enfoques técnicos y jurídicos, lo que dificulta su aplicación efectiva. La definición adoptada sigue la línea de la OCDE, centrándose en sistemas basados en máquinas con distintos niveles de autonomía y adaptabilidad (Sousa e Silva, 2024). Sin embargo, esta definición deja sin abordar problemas clave relacionados con la opacidad de los modelos de propósito general, el uso masivo de datos personales y los riesgos sistémicos que estas tecnologías pueden generar.

Otro problema en torno a la regulación del *AI Act* es que esta podría desalentar la innovación y llevar a la adopción de modelos modificados en comparación con otras regiones. Esto generaría un ecosistema fragmentado, donde las empresas desarrollen versiones menos avanzadas de sus sistemas para cumplir con la normativa europea, mientras que en otros lugares se fomenten enfoques más flexibles y competitivos. A largo plazo, esta rigidez podría aumentar la dependencia de actores externos, limitar el liderazgo de la Unión Europea en inteligencia artificial y favorecer a grandes corporaciones en detrimento de emprendimientos innovadores más pequeños.

Así pues, si la tecnología artificial se implementa sin regulación, puede ocasionar problemas sociales de manera indiscriminada; por otro lado, como señaló la doctora Moranchel Pocatererra, directora de este trabajo, en una disertación (2022): si se implementan candados regulatorios demasiado rígidos, entonces las promesas económicas y sociales de la IA nunca podrán despegar. Por ende, es importante destacar el compromiso de la Comisión Europea con el equilibrio entre regulación e innovación a través de una propuesta regulatoria basada en el riesgo.

Con respecto a la teoría social en esta tesis, los marcos regulatorios en torno a la IA de México, Japón y la Unión Europea pueden verse reflejados en el pensamiento y la teoría de Bruno Latour sobre la tecnociencia y la política. Dado que Latour propone la idea de que la tecnociencia y la política están interrelacionadas y que la tecnología y la ciencia tienen un papel importante en la toma de decisiones políticas y en la construcción de la realidad social. Esto también se ve aplicado con respecto al modelo de programa y antiprograma que propuso Latour.

Desde la perspectiva de la teoría del actor-red, los marcos regulatorios en torno a la IA de México, Japón y la Unión Europea pueden entenderse como el resultado de las interacciones y negociaciones entre diversos actores humanos y no humanos que conforman una red sociotécnica. Estos actores tienen intereses, objetivos y visiones distintas sobre la IA y su impacto en la sociedad, y buscan traducir e inscribir sus intereses en los instrumentos jurídicos que regulan la IA. Así, los marcos regulatorios son objetos de contorno que contienen programas y antiprogramas que reflejan las alianzas y conflictos entre los actores de la red. Estos objetos de contorno son dinámicos y cambiantes, ya que están sujetos a las transformaciones que ocurren en la red a medida que se incorporan nuevos actores, se generan nuevas controversias o se producen nuevos avances científicos y tecnológicos; por ejemplo, en 2024 El Consejo Europeo adoptó el primer tratado internacional de inteligencia artificial firmado por países como Canadá, Estados Unidos, México y Japón, en donde su lineamiento jurídico menciona que los Estados signatarios deben proporcionar recursos legales efectivos para las víctimas de violaciones de derechos humanos resultantes del uso de Inteligencia Artificial, conforme a los requisitos del derecho internacional y en armonía con su legislación local, para asegurar su accesibilidad real (Steta, 2024).

En este sentido, los marcos regulatorios en torno a la IA de México, Japón y la Unión Europea pueden ser entendidos como ejemplos de cómo la tecnología y la

ciencia están influyendo en la política y en la construcción de la realidad social en cada una de estas regiones o países. Estos marcos regulatorios establecen principios éticos y medidas para guiar el desarrollo y aplicación de la IA y proteger los derechos humanos y la privacidad, y pueden regular y controlar el impacto de la IA en la sociedad.

De esta manera, comparar los marcos regulatorios de la IA de México y Japón con los de la Unión Europea permite tanto entender cómo cada país enfrenta la implementación de esta tecnología en su contexto social, como también contrastar los enfoques de cada uno en relación con los derechos humanos y los problemas sociales asociados con la IA. En el caso de México, la Agenda Nacional Mexicana de la IA establece principios éticos y medidas para regular la IA en el país. En el caso de Japón, se ha establecido una estrategia de IA que promueve la investigación y el desarrollo de la IA, así como su aplicación en diversos sectores, y que incluye medidas para proteger los derechos humanos y la privacidad. Desde la perspectiva de la teoría de Latour.

Así pues, desde la perspectiva de la teoría de Latour, estos marcos regulatorios muestran cómo la tecnología y la ciencia están influyendo en la política y en la construcción de la realidad social en México y Japón. Además, se puede observar cómo estos países han establecido medidas y principios éticos para regular la IA y proteger los derechos humanos y la privacidad de sus ciudadanos, lo que refleja el concepto de antiprograma de Latour en la respuesta a los problemas sociales generados por la IA. De este modo, la teoría de Latour puede ser utilizada para comprender cómo la tecnología y la ciencia están influyendo en la política y en la sociedad y cómo se están utilizando para abordar problemas sociales en el contexto de los marcos regulatorios de la IA en México y Japón.

4. La regulación de México en torno a la IA

Es importante considerar la situación de la sociedad mexicana respecto a la IA y su regulación en cuanto a derechos humanos. Esto permitirá examinar de manera más detallada la legislación y las propuestas regulatorias respecto a esta tecnología en el país. Se evaluarán el marco normativo relacionado con la IA para

determinar cómo el contexto social está impactando este tema. Más importante aún, se busca determinar si la academia, el sector privado o el gobierno están considerando los derechos humanos, la no discriminación, la privacidad, el acceso a la información y la responsabilidad como pilares esenciales del marco regulatorio, y si estos están siendo influenciados tanto por el impacto de la tecnología en la sociedad como por la influencia de la sociedad mexicana en la tecnología.

La constitución mexicana formula su legislación en torno a tratados internacionales, especialmente los relacionados con los derechos humanos. En este sentido, Alvarez Ledesma explica: “Se reconoce la doble fuente de derechos humanos: la Constitución y los tratados internacionales” (2023). Y esta característica vinculante se expresa de la siguiente manera:

En los Estados Unidos Mexicanos todas las personas gozarán de los derechos humanos reconocidos en esta Constitución y en los tratados internacionales de los que el Estado Mexicano sea parte, así como de las garantías para su protección, cuyo ejercicio no podrá restringirse ni suspenderse, salvo en los casos y bajo las condiciones que esta Constitución establece (Constitución Política de los Estados Unidos Mexicanos [Artículo primero], 2011).

Es importante señalar que en México la regulación de la IA no es un tema prioritario para los legisladores, por lo que aún no existe un marco normativo específico para regular la IA en el país. De momento, se utilizan leyes cuyo contenido puede aplicarse en materia de IA como la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados, la Ley Federal de Protección de Datos Personales en Posesión de los Particulares, inclusive, la Ley Federal de Transparencia y Acceso a la Información Pública, para regular aspectos específicos de la IA. Aunque esto proporciona alguna protección, se requiere una regulación más completa para abordar todos los aspectos legales relacionados con la IA. Por tanto, se analizarán los derechos humanos, la IA y su historia como elementos importantes, considerando el contexto histórico, social y tecnológico como eje de desarrollo del análisis.

A. El contexto de la IA en México

Es necesario comprender la perspectiva de la sociedad mexicana sobre la IA y la regulación basada en los derechos humanos para poder analizar de manera adecuada la legislación y propuestas regulatorias relacionadas con esta tecnología en el país. En particular, es importante enfocarse en tres aspectos: primeramente, entender el contexto de inversión en investigación, desarrollo e implementación de la IA en México; en segundo lugar, cómo la sociedad mexicana interactúa con esta tecnología (incluyendo el acceso a ella); y finalmente, identificar las áreas de oportunidad para México en relación con la IA.

En relación con el contexto de investigación, desarrollo e innovación de la IA en México, Microsoft y otros investigadores (2018) proponen tres escenarios en cuanto al impacto económico de esta tecnología en el país: negativo, neutral y positivo. Con base en estos escenarios, los investigadores sugieren el camino que México debería tomar para aprovechar o ignorar las oportunidades de la IA. El escenario negativo plantea que México siguiendo el mismo padrón que con relación a las tecnologías de la tercera revolución industrial, es decir, no desarrollando su potencial de innovación, desarrollo e implementación de la IA y más bien continúan consumiendo estas tecnologías del exterior:

“Este escenario es de continuidad y asume que la economía mexicana absorbe las tecnologías asociadas a IA al mismo ritmo que lo hizo con las TIC durante la tercera revolución industrial. En el modelo elaborado a propósito de este ejercicio, el escenario se representa con un Efecto Automatización suave que acelera el crecimiento económico, pero sin Efecto Innovación” (ídem).

El informe de Microsoft (2018) plantea dos escenarios posibles para el impacto de la IA en el crecimiento económico de México. El escenario positivo supone que México adopta la IA y la usa para aumentar su productividad y su innovación, lo que le permite generar más valor agregado en sus productos y servicios y volverse más competitivo. El escenario positivo implica también que México invierte en tecnología, capital e innovación, y que re asigna sus recursos entre los sectores de acuerdo con las demandas del mercado. El escenario negativo, en cambio, supone que México no adopta la IA y se queda rezagado frente a otros países. En

este escenario, México no se beneficia de la automatización ni de la innovación, y su crecimiento económico se estanca o disminuye. En este escenario, tampoco hay reasignación de recursos entre los sectores, lo que implica una pérdida de eficiencia y competitividad.

Actualmente, la situación de México con relación con la innovación, desarrollo e implementación de la IA no es alentadora. En primer lugar, el país presenta una gran brecha en la adopción de Internet y de Internet de alta velocidad por parte de su población. Y esto dificulta la implementación de herramientas de tecnología de la IA, especialmente aquellas que dependen de redes de mayor capacidad para funcionar correctamente al ser incorporadas a Internet de las Cosas.

En segundo lugar, no existen políticas públicas que promuevan la inversión en el desarrollo e investigación de la IA tanto en el sector público como en el privado. Sin embargo, hay una tendencia hacia la inversión y desarrollo por parte del capital privado, especialmente en empresas denominadas *unicornios mexicanos*. Según Palafox (2021), esta tendencia ya existe en Estados Unidos y es probable que se replique en México en los próximos años.

Hay que tomar en cuenta que, según Palafox, los siete unicornios mexicanos están buscando trabajadores especializados en IA. Muchas de las herramientas y plataformas de esas empresas utilizan algoritmos e instrumentos de aprendizaje de máquina que forman parte de la IA. Por eso, se está incentivando a que muchas personas se especializan en campos relacionados con esta tecnología.

Aunque algunas empresas mexicanas están en la vanguardia de la innovación y el desarrollo de la IA, el país presenta algunos retrasos. Según el informe *La IA en las compañías latinoamericanas (2020)* realizado por el MIT y Everis, en México “la comprensión sobre qué es la IA y su posible impacto en la transformación de las distintas industrias sigue siendo un tema pendiente en muchas empresas” (Opinno, 2020).

Asimismo, el informe menciona que las empresas: “Aunque tienen claro que buena parte de la robotización inteligente de procesos repetitivos puede ser un primer paso de implementación de IA, se trata de un conocimiento superficial, que

no permite diferenciar la robotización, de la IA o el machine learning” (Opinno, 2020). En este sentido el informe cita la opinión de analistas que plantean que las empresas mexicanas todavía se enfocan en la automatización de tareas sumamente repetitivas donde no determina el papel de una IA que también puede tomar algunas decisiones. Es decir, la mayoría de las empresas mexicanas aún carecen de un conocimiento profundo en torno a la IA; y ello, les puede impedir desarrollar todo su potencial.

En cuanto al problema de adopción de tecnología, México todavía tiene una enorme brecha digital. De acuerdo con el Instituto Nacional de Geografía y Estadística (INEGI, 2021), en México “78.3% de la población urbana es usuaria de internet. En la zona rural la población usuaria se ubica en 50.4 por ciento”. Buena parte de esta población - el 67.4% - se conecta a través de dispositivos móviles.

Ahora bien, estos porcentajes aumentaron en 2023, según la más reciente cifra de la Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH). De acuerdo con la ENDUTIH, en 2023, 97.0 millones de personas en México usaban internet (81.2 % de la población de 6 años o más), mientras que 97.2 millones utilizaban un teléfono celular (81.4 %), el 43.8 % de los hogares contaba con una computadora (16.9 millones de hogares) y el 90.4 % de los hogares, equivalentes a 34.9 millones, disponían de al menos un televisor (INEGI, 2024).

Precisamente, es a través de estos teléfonos donde se implementan las mayores plataformas con herramientas y tecnología de IA. Esto se debe al cada vez mayor poder de procesamiento de los chips que se encuentra en estos dispositivos a partir de arquitecturas como ARM⁸ y arquitecturas múltiples que incluyen la

⁸ “La arquitectura de la CPU Arm se basó originalmente en los principios de la computadora de conjunto de instrucciones reducido (RISC) e incorporó:

Un archivo de registro uniforme, donde las instrucciones no se limitaban a actuar sobre registros específicos.

Una arquitectura de carga / almacenamiento, donde el procesamiento de datos operaba solo sobre el contenido

del registro y no directamente sobre el contenido de la memoria.

Modos de direccionamiento simple, donde todas las direcciones de carga / almacenamiento solo se determinaron a partir del contenido del registro y los campos de instrucción” (Arm Ltd., s. f.).

incorporación de núcleos especializados en el aprendizaje de máquina como lo es la famosa serie A de Apple Inc. Además, las plataformas sacan ventaja del hardware (los elementos físicos) de los dispositivos como cámaras, sistemas de geolocalización, interfaces diegéticas como la pantalla táctil, giroscopios, bocinas y audífonos; lo cual, les permite adquirir una mayor cantidad de datos, y con ellos, tener el material suficiente para el funcionamiento eficaz e intuitivo de sus herramientas de IA (Apple Inc., n.d.).

En contraste, en Japón, la tasa de adopción de internet alcanzó el 86.2 % en 2024, según datos de Statista (s.f.). Este alto nivel de conectividad refleja no solo la infraestructura tecnológica avanzada del país, sino también su integración cultural de herramientas digitales en la vida cotidiana. Sin embargo, las preferencias de uso varían significativamente respecto a México, destacando plataformas como LINE y YouTube, que se adaptan a las necesidades locales. Al igual que en México, el acceso a dispositivos móviles y la capacidad de procesamiento de estos han permitido la implementación de tecnologías de IA, aunque con enfoques distintos derivados de las diferencias culturales y regulatorias entre ambos países.

La población mexicana tiene menor acceso a las herramientas de IA debido a que el país no desarrolló una adecuada infraestructura de tecnologías de información y comunicación durante la tercera revolución industrial. Esto ha generado obstáculos para la innovación y el desarrollo en torno a la IA. Sin embargo, esto no significa que no se puedan implementar mecanismos de inversión, desarrollo e implementación de la IA en segmentos específicos de la población; y la brecha digital existente puede llevar a una brecha inteligente, donde la investigación e innovación solamente beneficia a unos pocos y hace más rentable comprar tecnología en lugar de desarrollarla; aun así, hay que señalar que el “ mercado mexicano de Inteligencia Artificial alcanzará un valor de 3,700 millones de dólares en 2024, un incremento de 30% con respecto al año previo, cuando su valor fue de 2,820 millones de dólares, según datos de Statista Market Insights” (Riquelme, 2024).

En suma, México es un país que se enfrenta frente a un dilema en torno a la IA. Por un lado, el país ha logrado gestar siete unicornios que emplean ampliamente este tipo de tecnología y que buscan especialistas para su desarrollo e implementación. Por otro lado, hay una enorme brecha digital que impacta en acceso a las herramientas con IA, una falta de interés en torno a políticas públicas que la impulse, y un enorme desconocimiento del sector empresarial lo cual puede generar un escenario negativo para el país; por ejemplo, el país tiene que superar los problemas de Falta de recursos humanos competentes, Infraestructura tecnológica deficiente, Falta de inversión en I+D, Escasez de datos, y Políticas públicas insuficientes (El panorama de la inteligencia artificial en México, s.f). Por ende, es importante examinar en detalle el contexto social y político en torno a la IA en México, y esto incluye el estado de inversión en investigación, desarrollo e implementación de esta herramienta, con el objetivo de comprender las propuestas regulatorias relacionadas con ella. Por ello, es necesario dedicar un apartado para conocer brevemente las políticas públicas o ausencias de ellas, en relación con la IA, así como la relación de la sociedad mexicana con este tipo de tecnologías.

Aunque buena parte de los mexicanos tienen acceso a internet, el porcentaje de acceso a la IA es mucho menor que en países como Japón. Sin embargo, es importante tener en cuenta que el hecho de que muchas personas en México tengan acceso a plataformas que emplean IA representa una relación cercana con esta tecnología. Por ejemplo, un estudio de Kaspersky Lab reveló que más de la mitad de los mexicanos confían en la IA para encontrar una pareja (Ramos, 2021).

En algunos casos extremos, la relación entre herramientas de IA y las relaciones afectivas ha llevado a que la relación afectiva se dé directamente con la IA, como el caso de una persona que se enamoró de un chatbot llamado Replika (Phillips, 2022) una aplicación de la llamada de inteligencia artificial afectiva que genera lo que provoca que una relación con la IA es que funcione como una cámara de eco (Ferus, 2023). Sin embargo, el hecho de que las personas mexicanas confíen en la IA para encontrar el amor representa un cambio en el paradigma, en el que la

brecha entre actores humanos y no humanos relacionados con la IA en México ya no es tan amplia como podría asumirse.

Asimismo, las personas mexicanas emplean herramientas y plataformas de la IA. Por ejemplo, las personas que trabajan en las plataformas digitales de trabajo de transporte bajo demanda como Uber, aceptan que esta herramienta les sugiera desde que pasaje recoger, hasta la ruta para llegar a su destino. Además, plataformas mexicanas como Kavak (Expansión, 2021) crean un puente entre compradores y vendedores de vehículos usados a través de proceso de análisis que hace la IA del historial y potencial crediticio de los usuarios; ahora bien, para aprovechar mejor las oportunidades en torno a esta tecnología, el panorama de la IA en México debe centrarse en reentrenar a trabajadores y personas en el uso y desarrollo de la IA, entrenar y capacitar a nuevos egresados y técnicos, y crear un entorno de regulación y gobernanza (Torres, 2024).

Aunque los japoneses no consideran que las cosas tienen alma, pueden confiar en la IA por el sesgo de confianza o la tendencia a antropomorfizar a las máquinas. Según Ryan (2020), los seres humanos a menudo tendemos a dar cualidades humanas a la IA. Por ello, es posible que muchas personas confíen más en los robots que en los humanos cuando se trata de dinero (Fernández, 2021), y esta tendencia puede haber sido acelerada por el cambio en las relaciones sociales causado por la pandemia de 2020.

Hay que resaltar que Bryan sugiere una explicación no animista con respecto a por qué las personas confían en la IA. De acuerdo con el autor:

“Uno puede confiar en otro basado en hábitos confiables, pero depositar una confianza en alguien requiere que actúe de buena voluntad hacia el fideicomitente. Esta es la razón principal por la que los objetos hechos por humanos, como la IA, pueden ser confiables, pero no dignos de confianza, según el relato afectivo.

En la cuenta normativa, el fideicomisario debe ser considerado responsable de sus acciones, lo que AI no puede hacer. Considerando que, la IA confiable coloca la carga de la responsabilidad en quienes desarrollan, implementan y usan estas tecnologías”.

De esta forma, muchas personas en México tienen acceso a la IA, aunque la penetración de esta tecnología en la sociedad mexicana no es tan amplia como en

la sociedad japonesa debido a la brecha digital que es se generan desigualdades que afectan desproporcionadamente a los grupos menos favorecidos y a las poblaciones más pequeñas mientras que en las grandes ciudades se tiene acceso al servicio, en las localidades con pocos habitantes hay menos interés económico para proporcionarlo («La brecha digital», 2022). Esto ha llevado a que las personas mexicanas que utilizan la IA desarrollen un alto grado de confianza en ella, lo cual puede explicarse desde una perspectiva animista o desde la noción de confianza en los desarrolladores de estas herramientas.

Al examinar el estado de la IA en México, se puede comprender de manera más completa las propuestas regulatorias en torno a esta tecnología. Por lo tanto, el siguiente apartado examinará los esbozos regulatorios en relación con los aspectos éticos promovidos para contener los efectos negativos de la IA en México. Esto permitirá tener una visión más robusta al evaluar el impacto de las plataformas de movilidad en México y Japón y cómo los elementos regulatorios basados en principios como la responsabilidad, transparencia, imparcialidad y privacidad pueden hacerles frente.

B. La regulación de la IA en México

México ha mostrado un interés creciente en el desarrollo e implementación de la inteligencia artificial (IA), reconociendo su impacto en sectores como la economía, la educación y la administración pública. Sin embargo, el país aún carece de una ley específica que regule el uso de esta tecnología, aunque ha dado pasos iniciales mediante estrategias y documentos orientadores que, lamentablemente, han quedado relegados. Un ejemplo claro es la Agenda Nacional de Inteligencia Artificial (Coalición IA2030Mx), que se presentó como uno de los primeros esfuerzos por establecer un marco de referencia para la adopción de la IA en México, promoviendo principios éticos y recomendaciones para su desarrollo sostenible.

En términos regulatorios, el país se apoya en normativas preexistentes para abordar algunos de los desafíos que plantea la IA, como la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (DOF, 2010). Sin

embargo, estos marcos legales no abordan cuestiones clave como la transparencia. Además, organismos descentralizados que históricamente han servido como contrapeso frente a posibles transgresiones de derechos humanos, como el Instituto Nacional de Acceso a la Información (INAI), están enfrentando un proceso de extinción (Jiménez, 2024) para ser reemplazados por entidades centralizadas que no garantizan la misma transparencia ni autonomía. Ante este panorama, el debate sobre la necesidad de una regulación específica se ha intensificado, con llamados a establecer directrices claras que equilibren el avance tecnológico con la protección de los derechos humanos.

Así pues, es necesario examinar a fondo la regulación o propuestas regulatorias de la IA en México para comprender el impacto de las plataformas de movilidad en el país. Especialmente, porque comprender estas regulaciones o propuestas regulatorias permitiría entender cómo se enfrentan problemas sociales como la discriminación, la invasión a la privacidad, el acceso a la información, la responsabilidad y la explicabilidad. Además, analizar estos marcos normativos facilita la identificación de vacíos legales que podrían exacerbar desigualdades o riesgos éticos en el uso de tecnologías impulsadas por inteligencia artificial.

Si bien en México existe la necesidad de impulsar la investigación en Inteligencia Artificial (IA), hasta el momento no se han aprobado leyes específicas que la regulen. En 2021, la Cámara de Senadores impulsó un dictamen que promueve la investigación y desarrollo de la IA, reconociendo que "en México, aproximadamente entre el 19 y el 53 por ciento de los empleos serán impactados por la IA" (Senado de la República, 2021b). Este dictamen propone fomentar la investigación en IA con un enfoque social, con el objetivo de evitar el rezago en comparación con otras naciones. No obstante, a diferencia de los casos mencionados a lo largo de este trabajo, el gobierno mexicano no cuenta con un plan concreto para lograr este proceso de investigación, desarrollo e inversión.

En mayo de 2024, el diputado Ignacio Loyola Vera del Grupo Parlamentario del Partido Acción Nacional presentó una iniciativa con proyecto de decreto para expedir la Ley de Regulación Ética de la Inteligencia Artificial y la Robótica. Otra

propuesta es la del senador Ricardo Monreal que se ancla más a una perspectiva regulatoria como la Unión Europea y donde destaca el artículo 6 del proyecto que propone crear una “Comisión Nacional de Inteligencia Artificial” como órgano consultivo del Instituto Federal de Telecomunicaciones (Salgado Borge, 2024). Hay que resaltar que esta comisión estará integrada por cinco científicos de reconocido prestigio en desarrollo tecnológico y tecnologías de la información, designados de forma honorífica (idem).

A pesar de la falta de una normativa específica en materia de IA, en el Nuevo Tratado de Libre Comercio de América del Norte (T-MEC), incluye normas regulatorias comunes en torno a esta nueva tecnología. En ese sentido, en el informe sobre la implementación del T-MEC "se destaca la importancia de abordar la privacidad de los datos y la ciberseguridad. A medida que México busca beneficiarse de las tecnologías digitales, existe una oportunidad y de hecho una necesidad de colaborar abiertamente en cómo se elaborarán las disposiciones relacionadas con los datos" (Senado de la República, 2021b).

Por lo tanto, el informe destaca que las normas dentro del tratado pueden utilizarse para garantizar que el flujo de datos respete los códigos regulatorios de cada uno de los países que participan en el tratado. Esto significa que la regulación de los datos en el T-MEC se articula con la legislación o propuestas regulatorias locales. Para ello, se redactó el artículo 19.16 del T-MEC, que establece:

“Reconociendo la naturaleza global del comercio digital, las Partes procurarán: (a) intercambiar información y compartir experiencias sobre regulaciones, políticas, observancia y cumplimiento en relación con el comercio digital, incluidos: (i) la protección de la información personal, particularmente con el fin de fortalecer los mecanismos internacionales existentes para la cooperación en el cumplimiento de leyes que protegen la privacidad, (ii) la seguridad en las comunicaciones electrónicas, (iii) la autenticación, y (iv) el uso gubernamental de herramientas y tecnologías digitales para lograr un mejor desempeño gubernamental; ... Las Partes considerarán establecer un foro para abordar cualquiera de los asuntos enumerados anteriormente, o cualquier otro asunto relacionado con el funcionamiento de este Capítulo” (Senado de la República, 2021b).

Que exista un apartado en el T-MEC relacionado con la privacidad de los datos es muy importante. Esto se debe a que los datos son un elemento fundamental para el funcionamiento de los algoritmos de aprendizaje automático, que a su vez son esenciales para muchas herramientas de IA actuales. De esta manera, al hacer énfasis en la privacidad, la propuesta en el T-MEC, se alinea con los principios éticos regulatorios universales en los derechos humanos.

El siguiente documento elegido para su examen es la Iniciativa que reforma y agrega diversas disposiciones de la Ley de Ciencia y Tecnología, presentada por el diputado Mario Alberto Rodríguez Carrillo (LXIV Legislatura 07.01.2021). Este documento es relevante ya que reconoce los problemas éticos relacionados con la IA y la necesidad de establecer un marco regulatorio. Al igual que en el caso de la propuesta regulatoria británica, el diputado del partido Movimiento Ciudadano sostiene que es mejor incluir una regulación temprana en el marco de códigos existentes, modificando algunos artículos de la Ley de Ciencia y Tecnología.

En el documento resalta la importancia del derecho a la privacidad. En especial, la propuesta resalta elementos muy técnicos sobre el funcionamiento técnico de la IA como los datos de entrenamiento y cómo el mal uso de estos puede erosionar la privacidad de las personas. Por ello, el diputado argumentó que hay:

“exigir que los datos de entrenamiento tengan que ser almacenados por un cierto tiempo, que su metodología de recopilación y elaboración esté documentada y fundamentada, que los conjuntos de datos sean completos, y en el caso de datos representativos de poblaciones humanas, que éstos representen fielmente las características de género y etnia de la población y sean capaces de evitar que los sistemas de IA ejerzan actos de discriminación” (LXIV Legislatura 07.01.2021).

Asimismo, el documento también reconoce el derecho del acceso a la información en torno a las herramientas ancladas a la IA. En este sentido, el diputado Rodríguez Carillo (2021) argumenta:

“...debido al alto nivel de sofisticación técnica que pueden tener los sistemas de IA, así como la necesidad que tienen las empresas de mantener secretos industriales, los sistemas basados en IA con frecuencia son opacos, lo que abre posibilidades para que puedan ser usados con fines cuestionables. Una vía para reducir dichas posibilidades es exigir requisitos de transparencia en dichos sistemas, tales como exigir a los operadores de sistemas de IA que informen claramente al público que

están interactuando con una IA y no con un humano, o avisar de forma clara sobre las limitaciones y capacidades del sistema de IA.

Los problemas de no discriminación y responsabilidad también son mencionados en la Iniciativa. No se abordan como temas importantes, sino como mecanismos de apoyo a los derechos a la información y a la privacidad. Esto se refleja en la argumentación del diputado de que “es necesario lograr un delicado equilibrio entre exigir un mínimo de transparencia y responsabilidad, y evitar infringir la propiedad intelectual y el secreto industrial detrás de dichos algoritmos” (ídem), y en la mención de la necesidad de que los datos reflejen adecuadamente las características sociales de toda la población para evitar fenómenos como el sesgo y la discriminación aprendida, así como el sesgo de confirmación, que “está impulsado por la recopilación de datos al azar y las correlaciones espurias, reforzado por las desigualdades institucionales y contaminado por el sesgo de confirmación” (O'Neill, 2016).

El último documento que se examinará es la Agenda Nacional Mexicana de la IA. En este documento se presentan las rutas para la investigación, desarrollo e innovación en el área de la IA, pautas para la gobernanza basada en esta tecnología y sugerencias regulatorias ante los retos que plantea. Según Del Pozo, Claudia May, Gómez Mont, Constanza y Martínez Pinto, Cristina y coordinación (2020):

“En 2018, México fue uno de los primeros 10 países del mundo en presentar una estrategia de IA que buscaba sentar bases para el aprovechamiento de ésta desde la esfera pública. Fue entonces que se crea la coalición ciudadana IA2030Mx fundada por nueve instituciones de todos los sectores. Desde IA2030Mx nos planteamos la necesidad de contar con un ejercicio nacional, multidisciplinario, multisectorial y colaborativo para desarrollar un plan de acción que le sirviera al gobierno, a la academia, a la sociedad civil y a la industria”.

El documento se divide en varios capítulos que brindan recomendaciones sobre un tema en particular. Sin embargo, en el apartado correspondiente a la ética, que es el centro de atención de esta investigación, los investigadores mencionan que es “urgente que México evalúe los beneficios de la automatización, como la eficiencia, la precisión y la personalización, en contraposición a valores como la justicia, la privacidad y la libertad” (Del Pozo et. al, 2020). Más adelante, el

documento señala que estos valores deben apoyarse en los derechos humanos para tomar decisiones que permitan que la IA tenga un impacto positivo en las personas (ídem).

Es importante destacar que la idea de relacionar los derechos humanos con los principios éticos universales planteados en este trabajo surgió a partir de los planteamientos de la Agenda Nacional Mexicana de la IA. Esto es, para crear un hilo conductor y argumentativo que se refuerce a lo largo del apartado. Cabe mencionar que en esta tesis se abordaron problemas sociales relacionados con los derechos humanos, en parte debido a la justificación mencionada, específicamente cuando se menciona que:

“Uno de los principales beneficios de utilizar los derechos humanos como elemento central para el desarrollo de IA es que estos establecen claramente las responsabilidades que los gobiernos deben cumplir para proteger estos derechos y proponen mecanismos para prevenir y remediar posibles violaciones. Otra ventaja significativa de utilizar los derechos humanos como base para la creación y la gobernanza de la IA es que gozan de una aceptación internacional, construida a través de décadas de colaboración y cooperación” (ídem).

Esto significa que los derechos humanos ya incluyen actores no humanos, actantes que sirven como guía regulatoria para la IA. Además, estos códigos contienen mecanismos de inscripción que nos permiten enfrentar los códigos emergentes del programa (o algoritmo) de las plataformas de IA. Al contar con un grado de aceptación a nivel global, estos códigos otorgan un alto grado de legitimidad a una propuesta regulatoria que justifique los principios éticos que se quieren defender frente a los problemas de IA basándose en los derechos humanos.

Ahora bien, a partir de establecer la importancia de los derechos humanos como eje regulatorio de la IA, los autores identifican varios problemas éticos relevantes relacionados con esta tecnología. Según los autores del apartado ético de la Agenda, los principales desafíos de la IA son: “la libertad de expresión, la igualdad y no discriminación, la privacidad” (ídem). En la agenda hay que subrayar la importancia del derecho a la libertad de expresión, ya que este elemento ausente

en las propuestas es muy importante en torno a los problemas sociales de las TIC y de la IA.

Es importante mencionar las consideraciones de la agenda en torno al problema de la discriminación. De acuerdo con las autoras:

“Las prácticas discriminatorias que son producto de la implementación de sistemas autónomos inteligentes tienen serias repercusiones a nivel social, como lo son: el reforzar estereotipos (como los de género) [Noble 2018], la falta de representatividad poblacional [Shankar 2017], el perpetuar desigualdades sociales [Fry 2018] y el acceso desigual a oportunidades (como las laborales) [Dastin 2018 y Ali, 2019]” (ídem).

Esto quiere decir que los datos con los que se entrena la IA producen discriminación. Un ejemplo de ello es la discriminación financiera en la IA. En este sentido, el doctor Gilles dentro del curso sobre el aprendizaje de máquina explica:

“Tendemos a pensar en las computadoras como objetivas e imparciales, pero investigadores como Kate Crawford han demostrado que los modelos de aprendizaje automático pueden mostrar el mismo tipo de sesgo que tienen los humanos. Por ejemplo, los sistemas de aprendizaje automático utilizados por los bancos pueden rechazar préstamos de manera desproporcionada a ciertos grupos étnicos” (Gilles, en Coursera, s.f).

Ahora bien, en relación con la libertad de expresión, este es uno de los principios fundamentales consagrados en la Declaración de Derechos y es, incluso, la piedra angular de varios marcos legislativos, como el de Estados Unidos, establecido en la Segunda Enmienda de su Constitución. Sin embargo, los autores de la Agenda Nacional Mexicana de la IA sostienen que la libertad de expresión se encuentra bajo ataque por parte de la IA en dos sentidos: por un lado, en la forma en que la IA controla la distribución y el contenido de la información; y por otro, en el enorme poder de vigilancia relacionado con esta tecnología, a través de herramientas de reconocimiento espacial, geolocalización y análisis de comportamiento.

En primer lugar, los autores de la Agenda señalan que la IA está “transformando la forma en que accedemos y compartimos información, como la administración de contenido en redes sociales, la función de asistentes personales, los filtros de spam y la creación de motores de búsqueda” (ídem), entre otras cosas. En segundo lugar, afirman que “la vigilancia constante y la limitación de las libertades

(a través de técnicas como el reconocimiento facial o el análisis de comportamiento permitidos por la IA) de una sociedad pueden restringir o incluso eliminar por completo la libertad de expresión” (ídem).

Bajo la perspectiva del autor de esta tesis y siguiendo el modelo del programa y antiprograma propuesto en el diagrama anterior, el problema de la equidad puede abordarse a través de herramientas ligadas al derecho al acceso a la información, pero también bajo propuestas como el derecho al entendimiento articulado por la EU y Japón. Por otra parte, el problema de la vigilancia relacionado con la libertad de prensa puede enlazarse con el derecho a la privacidad. Sin embargo, es importante considerar que la libertad de prensa es un problema que requiere soluciones más específicas y que deben ser abordadas y analizadas con mayor profundidad, especialmente en el contexto mexicano, que se ha vuelto mucho más complejo para los periodistas.

Así pues, México cuenta con tres documentos importantes relacionados con la propuesta regulatoria de la IA. Uno de ellos es el nuevo Tratado de Libre Comercio de América del Norte, que contiene códigos vinculantes para el cumplimiento del tratado, incluyendo un código relacionado con el derecho a la privacidad en el contexto de la ciencia de datos, lo cual es relevante para la IA. Los otros dos documentos son una iniciativa y una propuesta regulatorias que incluyen en su arsenal de códigos regulatorios los derechos a la privacidad, no discriminación e igualdad, acceso a la información y libertad de expresión, y que pueden y deben ser tratados en futuros trabajos en torno a los problemas sociales de la IA.

En síntesis, a partir del análisis social se pueden comprender las diferencias de las propuestas regulatorias entre México y Japón. En el caso japonés, se ha incorporado a la regulación de la IA no sólo como algo que se deba de prevenir a partir de un antiprograma sino como una herramienta para solventar los problemas sociales que enfrenta la nación. Mientras que en el caso mexicano existen Iniciativas y propuestas regulatorias en etapa muy inmadura a pesar de estar muy elaboradas que están las cuales están enfocadas en generar códigos que

contrarresten los códigos anti normativos que surgen a partir de la implementación de herramientas y plataformas con IA.

Además, en el caso legislativo de ambas propuestas regulatorias en torno a la IA, se encontró que, a pesar de que la propuesta japonesa considera a las herramientas de la IA como un medio para abordar sus problemas sociales, ambas ramas y propuestas regulatorias en ambos países protegen los derechos a la privacidad, no discriminación e igualdad y acceso a la información. Y es importante señalar que, en Japón, el problema de la responsabilidad es el que se encuentra en un nivel más contingente. En cambio, en la propuesta mexicana, especialmente en la Agenda Nacional Mexicana de la IA, se menciona que es necesario determinar quiénes son los actores responsables y establecer sanciones en caso de que la IA incumpla uno de los principios éticos universales establecidos al inicio de la tesis, así como el derecho a la libertad de expresión.

Finalmente, con base en estos elementos, se tienen las herramientas necesarias para analizar el impacto de las plataformas y sistemas de IA en la sociedad mexicana y japonesa, utilizando una constelación que incluye marcos de conocimiento técnico en relación a la IA, aspectos sociales en torno al uso y estado de la IA en ambos países, y aspectos legislativos derivados de la regulación o propuestas regulatorias de México y Japón en relación a los derechos humanos y principios éticos establecidos al inicio del trabajo. Esto permitirá, a través de un análisis de constelaciones, determinar si las plataformas de movilidad en México y Japón han violado alguno de los derechos a la privacidad, no discriminación e igualdad, acceso a la información, o si generan problemas en torno a la entendibilidad y responsabilidad. De esta manera, se podrá responder de manera holística a la pregunta de investigación.

5. Japón en torno a la IA

La siguiente sección se enfocará en la Sociedad 5.0, los derechos humanos y el Plan de gobernanza de la IA en Japón. Esto es, porque es importante conocer primero el objetivo de la implementación de tecnologías de la información y comunicación en Japón, para luego examinar cómo se articulan en torno a las

regulaciones de la IA en el Plan de gobernanza de esta tecnología y los marcos normativos relacionados con los derechos humanos. De esta manera, se podrá conocer la perspectiva regulatoria japonesa desde los problemas sociales vinculados a la implementación de esta tecnología, y comprender mejor sus propuestas regulatorias y su relación con los derechos humanos.

Japón se ha posicionado como un líder global en la integración de tecnologías avanzadas, incluida la inteligencia artificial (IA), para abordar desafíos sociales y económicos específicos. Su enfoque único combina innovación tecnológica con principios éticos y sociales profundamente arraigados en su cultura, lo que se refleja en iniciativas como la Sociedad 5.0 , un modelo centrado en el ser humano que busca utilizar herramientas digitales para mejorar la calidad de vida y resolver problemas estructurales, como el envejecimiento de la población y la disminución de la fuerza laboral.

Para comprender el papel de la IA en Japón, es fundamental examinar tanto los objetivos detrás de la implementación de estas tecnologías como las regulaciones y marcos normativos que las guían. En este contexto, el Plan de Gobernanza de la IA de Japón emerge como un documento clave que articula principios éticos y sociales, alineándose con los derechos humanos y las necesidades de la sociedad japonesa. Esto es porque la estrategia también resalta un enfoque ágil, basado en riesgos y con la participación de múltiples actores para la regulación de la IA, buscando equilibrar la protección de los derechos humanos y los valores con la promoción de la innovación y la adopción de esta tecnología (Habuka, 2023).

Además, Japón busca aprovechar sus fortalezas en manufactura, robótica e infraestructura social, así como sus perspectivas culturales y éticas únicas, para crear una sociedad centrada en el ser humano que contribuya al bien común global (Society 5.0 – ASCA Tech Lab, s.f.; Consejo para los Principios Sociales de la IA Centrada en el Ser Humano, 2016). Lo cual coloca sus propuestas regulatorias entre las más llamativas debido al balance entre innovación y derechos humanos.

Este análisis no solo permitirá comprender cómo Japón aborda los desafíos éticos y regulatorios de la inteligencia artificial (IA), sino también comparar sus propuestas con estándares internacionales y perspectivas globales, como las de la Unión Europea o México. Además, destacará cómo Japón se mantiene como un referente clave en el ámbito de la IA, a pesar de ocupar el octavo lugar en índices y estudios como el Artificial Intelligence Index Report . Cabe señalar que, según este informe, Japón se ubicó en décimo lugar en 2022 respecto al número de nuevas empresas de IA financiadas por área geográfica (Clark et al., 2023).

A. La sociedad 5.0

La regulación de la IA está relacionada con valores universales en torno a los derechos humanos y las necesidades de su sociedad. Por lo tanto, es necesario examinar la regulación de la IA en el contexto del análisis social y ético construido en ese país. En particular, según las observaciones de reportes como los de Gal y Kim, y la propuesta regulatoria japonesa encarnada en la Gobernanza Japonesa de la IA 1.0 (AI Governance in Japan Ver. 1.0, 2021).

Japón tiene una relación armónica con la tecnología, especialmente en torno a procesos y fenómenos como el tecno-animismo. El término tecno-animismo fue acuñado en 2006 por Anne Allison en su libro *Millennial Monsters: Japanese toys and global imagination (tecno-animism)* y, en este libro, la autora sostiene: “La mitología no desapareció en esta era de la tecnología, sino que se arraigó dentro de la tecnología misma” (Allison, 2006, p. 28). Esto significa que las prácticas y condiciones de las religiones y ética tradicionales simplemente emigraron hacia la sociedad moderna.

Además de este aspecto social y espiritual en torno a la tecnología, Japón también ha visto a esta como un medio de escape ante distintos periodos de crisis, y la IA no es diferente en este momento. En especial, Japón está considerando a la IA como herramienta para compensar su cada vez más decreciente fuerza laboral ligada a una baja tasa de natalidad y una población más envejecida. Además, estas mismas herramientas tecnológicas se están desarrollando con un enfoque social, especialmente en relación con el cuidado de los adultos mayores.

A lo largo de su historia, Japón ha utilizado la tecnología como una forma de resolver crisis. Esto ha llevado a que la tecnología sea parte de su identidad nacional y elemento esencial de su política pública. En este sentido, Kitano (2007) comenta: "La búsqueda japonesa de tecnología avanzada, incluyendo robots, siempre ha estado relacionada con el crecimiento de la economía nacional, que se ha narrado tanto en Japón como en Occidente como una singularidad de Japón" (Kitano, 2007).

El contexto sociohistórico de este apego a la tecnología de nueva cuenta surge ante la caída del Bakufu de Tokugawa bajo la política de *bunmei-kaika*. En palabras de Kitano (2007):

"Por la política bunmei-kaika, se introdujeron muchas ideas occidentales en Japón. Varios nunca se habían conocido en Japón, debido a la política de cierre de la nación adoptada por el shogunato Tokugawa durante doscientos años. Se inventaron muchos términos novedosos para definir los conceptos occidentales, como Shakai (sociedad), Tetsugaku (filosofía), Risei (la razón), Kagaku (ciencia), etc. En los demás casos, los términos indígenas tuvieron que agregar el nuevo concepto occidental y cambiar su significado original, como Gijyutsu (tecnología), Shizen (naturaleza) y Rinri (ética)".

Kitano también explica que los enormes avances tecnológicos de Japón se pueden explicar con la perspectiva ética que premia la armonía entre los individuos y el bienestar colectivo. Asimismo, esto se traduce en la política económica, donde la tecnología siempre ha estado relacionada con el crecimiento económico. Y quizás no existe mayor ejemplo de la persecución de armonía y la incorporación de la IA y la robótica en torno al crecimiento tecnológico, y en especial a resolver los problemas sociales, que el programa de la Sociedad 5.0.

De acuerdo con Fukuyama (2018): "La sociedad 5.0 propuesta por el gobierno japonés es un concepto claro. Fue redactado en el V Plan Básico de Ciencia y Tecnología por el Consejo de Ciencia, Tecnología e Innovación, y aprobado por decisión del Gabinete en enero de 2016". El objetivo de esta propuesta es resolver los problemas socioeconómicos que enfrenta el país a partir del uso de las tecnologías de la información donde se incluye a la IA y la robótica. Entre esos problemas sociales destacan el envejecimiento de la población, la necesidad de

suplementar o compensar la pérdida de la fuerza de trabajo, y la reparación de la infraestructura creada durante el periodo de crecimiento de 1950-1970.

La sociedad 5.0, según Fukuyama (2018), es la propuesta del gobierno japonés para resolver los problemas socioeconómicos que enfrenta el país a través del uso de tecnologías de la información, incluyendo la IA y la robótica. Estos problemas incluyen el envejecimiento de la población, la necesidad de complementar o compensar la pérdida de la fuerza de trabajo y la reparación de la infraestructura creada durante el período de crecimiento de 1950-1970. Otros países que enfrentan problemas similares recurren a la migración para solucionar la escasez de mano de obra y la población productiva, pero Japón ha elegido abordar estos problemas a través de la tecnología.

Es necesario conocer cuáles fueron las cuatro sociedades anteriores. Esto es, en aras de comprender la noción de la Sociedad 5.0. De acuerdo con Fukuyama (2018):

“Mirando hacia atrás en la historia humana, podemos definir diferentes etapas de las sociedades. La sociedad 1.0 se define como grupos de personas que cazan y recolectan en convivencia armoniosa con la naturaleza; La sociedad 2.0 formó grupos basados en el cultivo agrícola, el aumento de la organización y la construcción de naciones; La sociedad 3.0 es una sociedad que promueve la industrialización a través de la revolución industrial, haciendo posible la producción en masa; y la Sociedad 4.0 es una sociedad de la información que obtiene un mayor valor agregado al conectar activos intangibles como redes de información. En esta evolución, la Sociedad 5.0 es una sociedad de la información”.

La sociedad 5.0 es una sociedad con datos importantes y usan herramientas tecnológicas como algoritmos y redes neuronales de IA. El objetivo de la sociedad 5.0 es crear una sociedad centrada en el ser humano que tenga un desarrollo económico y social sostenible y que las personas puedan disfrutar de una alta calidad de vida. Y la estrategia de la sociedad 5.0, desarrollada durante la administración del ministro Abe en 2016, tiene como objetivo lograr un equilibrio entre el desarrollo social, la calidad de vida de las personas y el crecimiento económico.

Gal (2020) señala que, aunque los actores no humanos no son parte de la sociedad como iguales de los seres humanos, pueden ser considerados más que simples herramientas. Esto es basado en entrevistas realizadas a personas que trabajan en la investigación y desarrollo de herramientas robóticas y de IA. Por ejemplo, Yasuo Kuniyoshi, director del Centro de Investigación de IA de Próxima Generación de la Universidad de Tokio, sostiene que los japoneses son conocidos por su admiración por el hardware, a veces por el software. Además, Arisa Ema, asistente profesor de la Universidad de Tokio, comenta que los japoneses, en ocasiones, han llegado a ver a los robots como socios debido a una larga historia de cultura popular amigable con los robots.

El plan de la Sociedad 5.0 impulsó el desarrollo de la IA en Japón, lo que llevó a la creación de iniciativas de investigación, desarrollo e implementación de esta tecnología. Según Kim (2019), "como parte de la realización de las visiones de la Sociedad 5.0, el Gabinete creó el Consejo Estratégico para la Tecnología de la IA para abordar específicamente la I+D y la implementación de la IA".

La idea de la Sociedad 5.0 en Japón es utilizar tecnologías, como la IA, al servicio de las personas y no como parte de un nuevo entramado social. Esto cuestiona la idea de que la IA forma parte de un entramado ético afectivo que ve a la IA como una herramienta para resolver los grandes problemas sociales del país. Sin embargo, según el investigador Ohya (citado en Gal, 2020), "los japoneses parecen diferenciar entre seres humanos e IA o robots en menor grado, lo que podría hacerlos menos centrados en los humanos, al menos en comparación con las culturas occidentales".

Además, la sociedad 5.0 busca mejorar las condiciones de vida de todos, incluyendo a personas con diferentes capacidades y enfermedades crónicas degenerativas. Una forma de lograr este objetivo es mediante la transformación de las cadenas de valor y el uso de tecnologías financieras o Fintech. Esto integra diferentes componentes tecnológicos, incluyendo la IA, en el corpus de la sociedad japonesa. Y proporciona una nueva perspectiva para analizar la regulación japonesa en torno a la IA y examinar el Plan de gobernanza de la IA en

Japón versión 1.0, que establece los principios que guiarán la investigación y desarrollo de esta tecnología en línea con las metas del plan de la Sociedad 5.0.

B. Plan de gobernanza de la IA en Japón versión 1.0 y 1.1

El Plan de gobernanza de la IA en Japón versión 1.0 y 1.1 es un documento que establece los principios que deben guiar el rumbo del desarrollo e investigación de la IA en Japón, siempre en línea con las metas de la sociedad 5.0. El objetivo del plan es lograr un equilibrio entre el desarrollo económico, el desarrollo social, la calidad de vida de las personas y los derechos humanos, la democracia y el Estado (en la versión 1.1).

En cuanto a la relación del Plan de gobernanza de la IA en Japón con los derechos humanos, el plan establece ciertos principios que deben tenerse en cuenta al desarrollar y utilizar la IA. Estos principios incluyen la protección de los derechos humanos, el acceso a la información, la resolución del problema de la responsabilidad en el uso de la IA y la promoción de la confianza pública en la tecnología. Según el documento, es necesario tomar estos principios en cuenta para garantizar un desarrollo sostenible y responsable de la IA en Japón.

“...que Japón debería romper con los modelos de gobernanza convencionales en los que el gobierno desempeña un papel de liderazgo en el proceso desde el diseño hasta la supervisión y el cumplimiento de las normas y que, en cambio, las empresas también deberían tomar la iniciativa en el diseño, el seguimiento y el cumplimiento de las normas, y propuso que en cada proceso de gobernanza, es decir, elaboración de normas, seguimiento y aplicación, los gobiernos deben garantizar la participación activa de las empresas que diseñan e implementan arquitecturas ciber físicas, así como de las comunidades y las personas que las utilizan” (METI, 2021)⁹.

Esta aproximación regulatoria se enfoca más en la reputación que en las reglas, lo que significa que las empresas e instituciones que investiguen desarrollen e implementen herramientas de IA se basarán en un modelo donde el control

⁹ METI significa ministro de economía comercio e industria en sus siglas en inglés. El METI ya tiene un papel activo en torno a la inteligencia artificial en Japón. Por ejemplo: “MEXT y METI han creado juntos programas universitarios relacionados con la IA, convocando a un organismo consultivo nacional con instituciones post-secundarias, la Federación Empresarial de Japón y partes interesadas de la industria para abordar la escasez de talento y las debilidades en la investigación” (Kim, 2019).

regulatorio no impida la innovación. El Plan de gobernanza de la IA en Japón versión 1.0 se centra en siete principios socialmente céntricos de la IA: “1. Centrado en el ser humano, 2. Educación/alfabetización, 3. Protección de la privacidad, 4. Garantizar la seguridad, 5. Competencia justa, 6. Justicia, responsabilidad y transparencia, y 7. Innovación” (idem).

Aunque es difícil ignorar el enfoque positivo demostrado por la cultura y las prácticas locales hacia las asociaciones entre humanos, IA y robots (Gal, 2020), la IA y los robots asociados pueden resultar herramientas útiles para una población que envejece. Especialmente porque Japón ve la automatización como una bendición económica, no como una amenaza (Gal, 2020). Sin embargo, esta tensión plantea otras cuestiones más complicadas (Gal, 2020).

Hay que resaltar que buena parte de la propuesta regulatoria de la IA en Japón se basa en las propuestas regulatorias de la Unión Europea. De acuerdo con el METI (2021):

“Dado que el tema de la gobernanza de la IA se ha discutido a nivel nacional e internacional, nuestra discusión debe ubicarse en el contexto de discusiones anteriores en todo el mundo. Por ejemplo, además de las opiniones que escuchamos del Grupo de Expertos y en otras ocasiones, las opiniones de empresas, grupos industriales y otras entidades de Japón y otros países sobre el Libro Blanco de IA y la Evaluación de Impacto Inicial para la regulación de IA de la Comisión Europea deben ser tenido en cuenta. También debemos analizar los debates basados en las opiniones de los consumidores en Japón en el Grupo de Trabajo de AI del Comité sobre Desafíos para la Digitalización del Consumidor de la Agencia de Asuntos del Consumidor”.

Precisamente, el Libro Blanco de la IA de la Comisión Europea es la propuesta regulatoria y de inversión y desarrollo de la Unión Europea. En el documento se plantean los grandes problemas éticos derivados de la implementación de la IA, en especial la privacidad, y se ofrecen soluciones para balancear la implementación y desarrollo de esta tecnología sin violar los principios éticos de privacidad, acceso a la información, responsabilidad y no discriminación. Así pues, el libro blanco "presenta opciones políticas para permitir un desarrollo fiable y seguro de la IA en Europa, con pleno respeto de los valores y derechos de los ciudadanos de la UE" (Comisión Europea, 2020).

Ahora bien, basándonos en los antecedentes sociales, éticos y legislativos de la propuesta regulatoria de la IA japonesa, es momento de analizar el documento. Para ello, examinaremos los principios éticos de esta propuesta más allá de los elementos regulatorios y las herramientas punitivas que contiene. Es decir, compararemos si el caso japonés incluye los principios de privacidad, acceso a la información y no discriminación planteados al inicio de esta tesis.

Se puede dividir la propuesta regulatoria y la ruta de la IA japonesa en tres categorías: social, ética, y comercial. En la categoría social, destacan el punto uno y dos, relacionados con desarrollar procesos de innovación centrados en el ser humano y generar un entorno que facilite la educación y alfabetización de la población respecto a esta tecnología. En cuanto a la parte económica y de desarrollo, se encuentra el punto siete, que se centra en la innovación.

Finalmente, los puntos tres, cuatro, cinco y seis se centran en los aspectos éticos de la privacidad, la seguridad, la no discriminación y la responsabilidad y el acceso a la información (punto 6) respectivamente. Estos valores y principios éticos coinciden con las preocupaciones éticas en torno a la IA del Libro Blanco de la Comisión Europea. Pero más importante aun, los principios éticos de la Ruta de la IA de Japón coinciden con los derechos humanos de no discriminación, privacidad y acceso a la información; además, de que busca resolver el problema responsabilidad.

En cuanto a la privacidad, el Plan de gobernanza de la IA en Japón versión 1.0 destaca que este principio es uno de los que más se mencionan en las propuestas regulatorias internacionales. Además, junto con los demás principios mencionados, la noción de privacidad debe ser implementada mediante una normativa intermedia que permita sanciones que no inhiban necesariamente la innovación en el sector, más allá de la autorregulación en las empresas. Es decir:

“...los miembros de un grupo de la industria para la promoción comercial de IA intentaron implementar principios de IA en reglas internas para operaciones y luego concluyeron que necesitan pautas intermedias y prácticas para incorporar los principios de IA en reglas internas...” y para “...abordar este problema, es ideal establecer una regla intermedia como una guía con múltiples partes interesadas. A la vez, debemos tomar nota del daño potencial a la innovación si la regla

intermedia se opere legalmente vinculante, como las regulaciones. Las sugerencias y notas anteriores apuntan a pautas intermedias no vinculantes basadas en objetivos para promover la innovación y el despliegue de IA” (METI, 2021).

El Plan de gobernanza de la IA en Japón versión 1.0 reconoce la diversidad de la IA y sus aplicaciones. Por ello, propone un marco regulatorio no estrictamente punitivo ni vinculante, pero que incluya excepciones en casos en los que sí sea necesario imponer restricciones en la implementación de ciertas herramientas de esta tecnología, como por ejemplo las cámaras de vigilancia con reconocimiento facial, cuyo impacto en la privacidad puede ser mayor que el de una plataforma de movilidad como Google Maps.

La regulación de la IA en Japón busca armonizar con la legislación internacional y con los principios éticos. Además, teniendo en cuenta su cultura empresarial y social relacionada con la innovación tecnológica y su capacidad para resolver problemas sociales (como se menciona en la Sociedad 5.0), promueve un marco regulatorio con reglas flexibles que beneficien a las empresas que las cumplan. No obstante, es necesario continuar debatiendo la gobernanza de la IA en Japón, ya que las discusiones concretas sobre el tema acaban de empezar y es probable que se intensifiquen en el futuro (METI, 2021).

Hay que resaltar que en la versión 1.1 ya se hace una mención a los derechos humanos cuando se plantea la gobernanza en torno a esta tecnología de la siguiente manera:

“La construcción de la gobernanza de la IA está en progreso a nivel mundial en paralelo en múltiples capas: discusión de alto nivel en GPAI, OCDE y UNESCO; consideraciones de un marco legal, por CAHAI (Comité ad hoc sobre IA) establecido dentro del Consejo de Europa, para desarrollar, diseñar y aplicar IA que pueda satisfacer los estándares del consejo relacionados con los derechos humanos, la democracia y el estado de derecho” (METI, 2021).

Es importante que este plan de gobernanza también menciona el problema de la explicabilidad en la IA. En este sentido, el Capítulo 3, se menciona que “es importante que los sistemas de IA sean fáciles de entender para los usuarios y destinatarios de sus resultados”, particularmente en casos de alto riesgo como los que se utilizan en la salud, el transporte o la justicia penal. Ahora bien, dentro de

los principios éticos-normativos de este documento, resalta en sí el principio de entendimiento que menciona que “es importante que los sistemas de IA sean fáciles de entender para los usuarios y destinatarios de sus resultados” (idem); lo cual, tiene afinidad con las propuestas regulatorias de la Unión Europea en materia de la IA.

En torno a los datos, hay que subrayar un elemento importante. Durante 2023 Japón declaró que los datos usados para entrenar IA generativa como Mid Journey o Dall-e (de la misma empresa que creó ChatGPT), no estarán sujetos a la protección de derechos de autor. De acuerdo con Lanz (2023) Japón ha declarado que el uso de conjuntos de datos para entrenar modelos de IA no viola la ley de derechos de autor. Este elemento es confirmado por el Centro para estudios estratégicos internacionales (Habuka en CSIS, 2023) quien plantea: “La Ley de derechos de autor se modificó en 2017 para promover el uso de datos en el aprendizaje automático. La enmienda aclaró que descargar o procesar datos a través de Internet u otros medios para desarrollar modelos de IA no es una infracción de los derechos de autor”.

Asimismo, Habuka plantea en esto *Japan’s Approach to AI Regulation and Its Impact on the 2023 G7 Presidency*, un informe sobre regular la IA y su impacto en la presidencia del G7 en 2023. En este informe se explica que Japón ha desarrollado y revisado sus regulaciones relacionadas con la IA con el objetivo de maximizar su impacto positivo en la sociedad, en lugar de suprimirla por riesgos sobreestimados. Asimismo, se destaca que Japón enfatiza un proceso basado en el riesgo, ágil y multilateral, en lugar de una obligación o prohibición única para todos; y el informe también discute qué argumentos podría hacer Japón en la cumbre del G7 sobre el enfoque centrado en el ser humano para la IA.

Finalmente, la versión 2021 del Plan de gobernanza de la IA en Japón hace énfasis en la armonización internacional. En particular, señala que es necesario alinear los principios de la IA, las reglas intermedias y los estándares de la IA porque están estrechamente relacionados. Esto muestra que la regulación de la IA está muy vinculada a los marcos éticos universales que se basan en las

propuestas regulatorias internacionales sobre derechos humanos. Al mismo tiempo, la Sociedad 5.0 busca implementar la IA en Japón para abordar sus problemas de la sociedad japonesa (como la disminución de población), lo que sirve como marco importante para contrastar con otras propuestas regulatorias como la mexicana.

C. Los derechos humanos desde el sistema jurídico japonés

Los derechos humanos son un componente importante del derecho japonés. Esto se debe a que el derecho moderno japonés ha sido influenciado por disciplinas francesas y alemanas. Además, en Japón, los derechos humanos están protegidos por la Constitución de Japón y por tratados internacionales de derechos humanos que han sido ratificados por el país, lo que afecta las propuestas regulatorias de Japón en torno a la IA. De acuerdo con Mori (2019):

“En Japón, existió la disciplina social del sentido mencionado desde la época antigua... Japón adoptó los derechos occidentales en un corto tiempo relativamente a partir de 1880, y con la serie de reformas realizadas en la posguerra, actualmente cuenta con un derecho moderno-contemporáneo de forma completa en el ámbito de la formalidad de las leyes escritas”.

La constitución moderna de Japón sufrió importantes transformaciones durante el siglo XIX y, especialmente, durante las reformas de 1947, durante el periodo de ocupación de la Segunda Guerra Mundial. En ese momento, además de cambiar el papel del emperador a una figura simbólica, se articularon tres principios elementales que, según Sato (2019), destacan para muchos investigadores como parte de la constitución japonesa:

- i. Principio básico 1: Soberanía popular.
- ii. Principio básico 2: Respeto de los derechos humanos fundamentales.
- iii. Principio básico 3: pacifismo.

Es importante destacar que, durante la reforma constitucional de 1947, la constitución japonesa también hizo una separación entre el estado y la religión. Este cambio fue especialmente relevante debido a que durante el periodo de la

constitución Meiji, la religión sintoísta había sido utilizada como un instrumento del estado para legitimar la figura del emperador en el gobierno. Algunos de los artículos de la constitución que reflejan esta separación son:

“El artículo 20 (3) de la Constitución dice —El Estado y sus organismos se abstendrán de intervenir en la educación religiosa y en cualquier otra actividad religiosa—, y el artículo 89 dice —Los fondos u otros bienes públicos no serán utilizados o aplicados para uso, beneficio o mantenimiento de instituciones o asociaciones religiosas, o para empresas de caridad, educacionales o humanitarias, que no estén bajo el control de la autoridad pública—” (idem).

Ahora bien, la constitución japonesa ha evolucionado en torno a las discusiones de los derechos humanos, agregando artículos que giran alrededor de los derechos sociales, como el derecho a mantener estándares mínimos para una vida sana y culta (art. 25), el derecho a la educación (art. 26), el derecho al trabajo (art. 27) y los derechos laborales (sindicato, negociación colectiva, acción colectiva) (art. 28) (Yuta, 2019). Y desde la Restauración del Emperador Meiji en 1868, Japón comenzó un proceso de modernización que incluyó su sistema de derecho, influenciado por pensadores franceses y alemanes. Aunque también adoptó conceptos jurídicos de América, Francia y Alemania para resolver adecuadamente las disputas que estaban teniendo lugar (Mori, 2019).

Más tarde, el carácter jurídico de Japón siguió cambiando, incluyendo las enmiendas surgidas de la ocupación estadounidense. Y elementos particulares para resolver controversias considerando rasgos particulares de la sociedad japonesa que se examinarán en el siguiente capítulo. Con base en esto, ya se puede examinar la propuesta regulatoria japonesa para la IA: Plan de gobernanza de la IA en Japón versión 1.0 y 1.1.

CAPÍTULO QUINTO

ESTUDIOS DE CASO: LA IA FRENTE AL ESPEJO DE MÉXICO Y JAPÓN

El siguiente capítulo examinará el impacto de la IA en las plataformas de IA en México y Japón. Para ello, se analizarán algunos casos empíricos en torno a los problemas y retos de esta tecnología en ambos países, centrando el enfoque en Google, Uber y el sistema de vigilancia C5, y se examinará cómo estos casos afectan a los marcos y propuestas regulatorias en relación con los derechos humanos de privacidad, acceso a la información y no discriminación, así como los problemas de responsabilidad y entendimiento en torno a la IA.

1. Controversias legislativas globales y locales

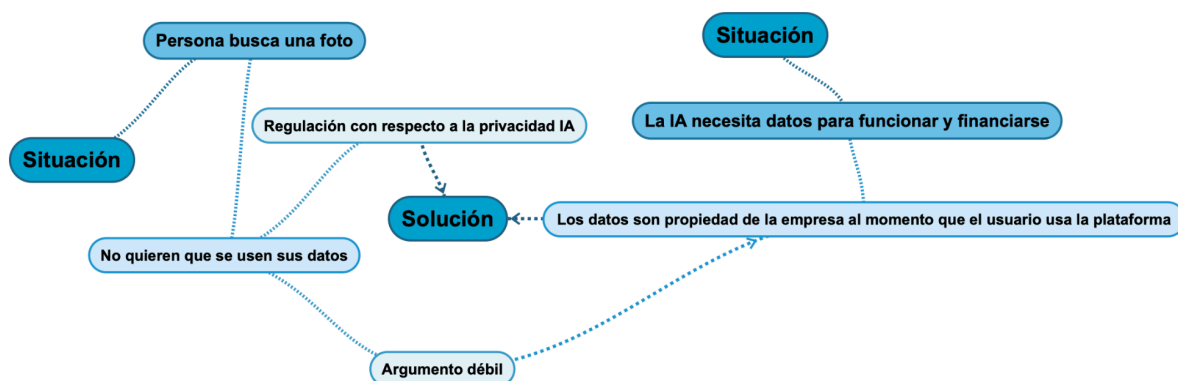
Es fundamental comprender las controversias locales y globales sobre la IA antes de analizar los casos de estudio de las plataformas, ya que estos fenómenos (en México y Japón) ponen a prueba nuestras herramientas teóricas y conceptuales. Los casos abordarán las transgresiones clave en la IA, como la toma de decisiones automatizadas y la equidad algorítmica, temas que, a su vez, afectan directamente derechos humanos como la privacidad, el acceso a la información y la no discriminación. Además, estos elementos generan interrogantes sobre la responsabilidad y el entendimiento de la IA. Por ejemplo, la compra de Twitter por Elon Musk en abril de 2022 destaca los desafíos de cómo los actores responden a estos dilemas éticos y protegen estos derechos.

La compra de Twitter (ahora X) por parte de Elon Musk plantea un desafío importante en torno a la IA, ya que la plataforma es vista como un espacio público de expresión, donde tanto humanos como agentes no humanos pueden participar. Sin embargo, esta diversidad corre el riesgo de ser coartada por el control de un solo individuo. Además, el uso de los datos que circulan en la red social puede vulnerar la privacidad de las personas; un cambio importante es que la plataforma X, formalmente Twitter, ahora hace uso de la “IA para resumir las tendencias en el

sitio, a menudo de manera inexacta, como se sabe qué hace la IA” (Tassi, 2024). Hay que resaltar que los datos incluyen tanto los estructurados (como el nombre, la edad, el género y el contenido textual) como los metadatos (como las imágenes, el video y el audio).

De acuerdo con Véliz (2022), la IA puede inferir mucha información sobre una persona a partir de los datos que produce, incluso si no los comparte explícitamente o tiene una actividad limitada. Para ello, utiliza dos técnicas principales: la de los “vecinos cercanos” y la de la clasificación semiótica. La primera consiste en relacionar un dato con otro para completar la información faltante. Por ejemplo, si alguien no indica su género, pero sube fotos de autos deportivos, la IA puede deducir que se trata de un hombre. La segunda consiste en agrupar a las personas en categorías basadas en los datos que generan. Por ejemplo, la IA puede asignar a alguien a la categoría “consumidor/ambientalista” y estimar su edad en función de otras personas que pertenecen a esa categoría. Estas técnicas se aplican tanto a los datos estructurados como a los metadatos, lo que permite a la IA inferir aspectos tan íntimos como la preferencia sexual de una persona.

En este sentido, hay que considerar que cualquier dato publicado en la plataforma puede comprometer la privacidad del usuario, especialmente si la plataforma cotiza en bolsa. Además, cuando la plataforma se convierte en propiedad privada y los datos publicados pertenecen de facto a una sola persona, las implicaciones son aún mayores, ya que se estaría entregando la identidad del usuario a un solo actor. Si una persona argumenta que sus datos no deben utilizarse, esta equivalencia puede convertirse en un objeto de mayor peso e incidencia.



Aunque exista un objeto regulatorio que proteja los datos de las personas, la plataforma, al tener un dueño único, puede imponer reglas internas que eviten el contenido de ciertos discursos e información, promoviendo discursos de odio. Además, algunos usuarios con capacidades técnicas pueden usar los datos públicos de otros usuarios en una IA para inferir información más personal, como su localización, las otras plataformas sociales en las que publican información y participan, entre otras actividades.

En este escenario, la equivalencia encarnada en un objeto regulatorio no es suficiente o no se puede modificar rápidamente. Por lo tanto, es posible que el usuario comprometido decida, como lo plantean Boltanski y Thevenot, alejarse del conflicto abandonando la plataforma y eliminando su cuenta. Este fenómeno de coyuntura permite poner a prueba los elementos teóricos y conceptos desarrollados a lo largo de la tesis frente a un escenario real, donde, como lo plantea el método de análisis de constelaciones, se puede crear un escenario para inferir la respuesta de las personas frente a una situación en la que las reglas de la plataforma pueden cambiar y afectar el derecho a la privacidad.

Lo anterior se relaciona con la noción justicia/equivalencia en Boltanski y Thevenot (1999) que se refiere a la capacidad de los actores sociales de establecer criterios comunes para evaluar y resolver los conflictos que surgen en diferentes ámbitos, *ciudades* o mundos de la vida social. Estos criterios se basan en principios que trascienden los intereses particulares y que permiten justificar las acciones y las demandas de los actores (Ramírez & Guerrero, 2011). En cuanto a los derechos

humanos de la IA, se puede argumentar que existe tensión entre diferentes principios de justicia/equivalencia, como el de la eficiencia, el de la seguridad, el de la dignidad o el de la creatividad, que se ponen en juego para regular el uso y desarrollo de la IA. Por ejemplo, ¿cómo se puede garantizar que la IA respete los derechos humanos sin limitar su potencial innovador? ¿cómo se puede proteger la privacidad de los usuarios sin afectar la calidad de los servicios basados en la IA? ¿cómo se puede asegurar que la IA no discrimine ni excluya a ciertos grupos sociales? Estas son algunas de las preguntas que podrían plantearse desde la perspectiva de Boltanski y Thevenot para analizar los problemas en los derechos humanos de la IA.

La otra perspectiva es la de programa, antiprograma y anti-anti-programa en Latour. De acuerdo con Latour (2008) un programa es un conjunto de acciones que un actor desea realizar, mientras que un antiprograma es un conjunto de acciones que otro actor realiza para impedir o modificar el programa. La resistencia es la fuerza que se opone al programa y que revela la existencia de otros actores involucrados en la red. Latour sostiene que tanto los actores humanos como los no humanos pueden tener programas y antiprogramas, y que la resistencia no es una propiedad intrínseca de los objetos, sino una relación entre actores. Desde esta perspectiva, los problemas en los derechos humanos de la IA se podrían analizar como el resultado de las interacciones entre diferentes programas y antiprogramas de actores humanos y no humanos, que generan resistencias y negociaciones en la construcción de la realidad social.

Ahora bien, como se mencionó arriba, cada vez hay más problemas en torno al poder e influencia en torno a las herramientas con sistemas de IA. En el caso de Twitter, su LLM Grok está diseñada para ser una herramienta de propaganda, y esto porque los modelos del lenguaje tienen la capacidad de influir en las decisiones y actitudes humanas (Marcus, 2025; Williams-Ceci et al, 2024). Pero es importante reconocer que estos sistemas también tienen repercusiones en los derechos humanos que se han mencionado a lo largo de la tesis.

Algunas controversias recientes respecto a los derechos de privacidad, acceso a la información y no discriminación derivados de la implementación y uso de la IA son

De regreso al tema de la compra de X (antes Twitter) por parte de Elon Musk, este proceso también pone en evidencia el problema de la equidad algorítmica, no solo en el sentido de cómo los sistemas de IA pueden generar sesgos, sino también en relación con qué actantes tienen el control de estas herramientas y cómo pueden ser utilizadas para influir en procesos de toma de decisiones. En este sentido, se ha señalado que, tras la adquisición de la plataforma por Musk, una granja de bots (usuarios artificiales) potenciados con sistemas de inteligencia artificial intentó influir en las elecciones presidenciales de Estados Unidos de 2024. Muchos de estos bots estaban basados en modelos avanzados desarrollados por empresas como OpenAI (Steffen, 2024).

Ahora bien, como se ha señalado a lo largo de este trabajo, la IA impacta de manera significativa tres derechos fundamentales: la privacidad, el acceso a la información y el derecho a la no discriminación. Por ejemplo, un estudio reciente reveló que los sistemas de IA generan sesgos discriminatorios en diversas áreas, como la financiera (afectando el acceso a créditos y préstamos), la laboral (vulnerando el derecho a la imparcialidad en contrataciones y ascensos), el ámbito jurídico, el acceso a la salud e incluso en la diferenciación de precios de alimentos (Bansal et al., 2023).

Las causas de estos sesgos pueden identificarse en tres niveles: primero, en la manera en que los modelos son diseñados y entrenados (sesgos inherentes al desarrollo de los modelos); segundo, en la selección y recopilación de los datos utilizados para entrenar estos sistemas (sesgos en la curaduría de datos); y tercero, en la existencia de sesgos latentes en los propios datos, los cuales pueden no haber sido detectados ni en la etapa de desarrollo ni en la de selección, lo que perpetúa las desigualdades de los humanos que generaron estos datos en primer lugar (Morton, 2024).

En cuanto a la privacidad, siempre existe el riesgo de que la información recolectada y almacenada por estos sistemas termine en manos equivocadas (Gomstyn & Jonker, 2024). Por ejemplo, una de las principales recomendaciones en torno al uso de ChatGPT es evitar compartir datos personales como el nombre completo, información financiera, planes para cometer un delito o la intención de hacerlo, detalles sobre el estado de salud o información protegida por derechos de propiedad intelectual (Miller, 2023). No obstante, muchas plataformas, como Uber, recopilan datos personales de sus usuarios, incluyendo la edad y el género de los conductores, así como otros datos sensibles como números de teléfono, generando preocupaciones en torno al uso y resguardo de esta información. Parte de este proceso se basa en la recolección de datos en línea y el registro de eventos móviles, donde la interacción del usuario con la aplicación es capturada a través de la recopilación de eventos desde dispositivos móviles. La arquitectura de registro de Uber se estructura en torno a principios clave como la estandarización de los registros, la consistencia en todas las plataformas (iOS, Android, Web), el respeto a la configuración de privacidad del usuario, la optimización del uso de la red y la confiabilidad del sistema sin comprometer la experiencia del usuario (Uber, 2021).

México y Japón presentan tanto diferencias como puntos de convergencia en el uso, los beneficios y las transgresiones asociadas con los sistemas de inteligencia artificial. La primera diferencia radica en que Japón necesita implementar herramientas de inteligencia artificial para automatizar servicios, debido al envejecimiento acelerado de su población y a la reducción demográfica que enfrenta. Además, su relación con los actores no humanos ha sido moldeada por siglos de prácticas religiosas como el budismo y, especialmente, el sintoísmo, que plantea un animismo posmoderno en torno a la interacción entre humanos y no humanos, así como la resiliencia (Yoneyama, 2018). Esta cosmovisión influye en la percepción y el uso de la tecnología, favoreciendo una integración en la que las herramientas tecnológicas son concebidas como parte de la sociedad y no como elementos que la dominan. Este enfoque puede estar relacionado con la noción de *wa*, un concepto que representa un entorno donde las personas pueden convivir

de manera armoniosa y cómoda, promoviendo el equilibrio, la comprensión mutua y la cooperación dentro de un grupo (Wa / The Government of Japan - JapanGov -, 2014); con base en esto, no es sorprendente que Japón esté invirtiendo fuertemente en el desarrollo e implementación de sistemas de inteligencia artificial en áreas como la manufactura (AI in Manufacturing | the Government of Japan - JapanGov -, 2025).

Por otro lado, México adopta una aproximación distinta en cuanto a la adopción, el desarrollo y la implementación de la inteligencia artificial. En primer lugar, el país aún cuenta con un bono demográfico, por lo que la IA no se percibe como un medio necesario para compensar la pérdida de fuerza laboral. En segundo lugar, México no ha mostrado urgencia en la creación de marcos regulatorios que orienten la integración de la IA en función de sus necesidades culturales y sociales, lo que sugiere una menor tendencia a concebirla como parte de las asociaciones sociales. No obstante, existen iniciativas que buscan aprovechar los modelos y sistemas de IA para abordar desafíos culturales, como la traducción de lenguas indígenas o el impulso al desarrollo de comunidades originarias (ESTILOGRÁFICA, 2024; Baños-González, 2024).

Ahora bien, más allá de los elementos culturales, y quizás entrelazados con estos, está el contraste de la defensa del derecho a la privacidad por parte de México y Japón. Japón otorga una gran importancia a la protección de la privacidad de sus ciudadanos. La Ley de Protección de la Información Personal (APPI), promulgada en 2003 y enmendada en 2017 y 2020, establece directrices estrictas para la recopilación y uso de datos personales (Comisión de Protección de la Información Personal, s.f.). Esta ley exige que las organizaciones obtengan el consentimiento explícito de los individuos antes de recolectar, utilizar o compartir su información personal, especialmente en casos que involucren datos sensibles o transferencias a terceros (CookieHub ehf, 2021).

Además, la APPI impone obligaciones como la notificación de violaciones de datos tanto a la Comisión de Protección de Información Personal (PIPC) como a los afectados, y la implementación de medidas técnicas y organizativas para

salvaguardar la información personal (Comisión de Protección de la Información Personal, s.f; (CookieHub ehf, 2021). Por ejemplo, en 2023, los *streamers* españoles IlloJuan y Plex fueron expulsados de un restaurante en Japón por comportarse de manera ruidosa e irrespetuosa. Además, grabaron a transeúntes sin su consentimiento para burlarse de ellos, lo que generó críticas en redes sociales debido a la invasión de la privacidad y la falta de respeto hacia las normas culturales japonesas (Rubén Martínez, 2024),

En contraste, México ha avanzado en la protección de datos personales con la promulgación de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares en 2010. Esta legislación establece principios y obligaciones para las entidades que manejan datos personales, incluyendo la necesidad de obtener el consentimiento de los individuos y garantizar la seguridad de la información. Sin embargo, en la práctica, la aplicación y el cumplimiento de estas disposiciones han enfrentado desafíos, y persisten preocupaciones sobre la efectividad de las medidas de protección de datos en el país. Por ejemplo, se han reportado casos de filtraciones de datos y uso indebido de información personal, lo que indica una necesidad de fortalecer la cultura de privacidad y la supervisión regulatoria en México (Cámara de Diputados del H. Congreso de la Unión, 2010).

2. Padrón de datos personales de telefonía móvil: privacidad y derechos humanos

La privacidad es el problema ético central de la IA. Según Véliz (2020): "La privacidad es la llave que abre los aspectos más íntimos y personales de uno mismo, lo que nos hace más vulnerables". Además, como se ha mencionado a lo largo de este trabajo, la privacidad es el elemento más fácil de violar con la IA debido a que a menudo se alimenta de datos y también puede inferir rasgos personales de un usuario. Por lo tanto, es importante examinar el Padrón de datos personales de telefonía móvil en México, sus consecuencias éticas en materia de privacidad y la respuesta de los actores ante ella.

El padrón básicamente consistía en el registro de datos biométricos en una base de datos de las compañías de telefonía móvil. Esto significa que, si este padrón se implementaba, el usuario estaba obligado a proporcionar muchos elementos de su identidad, incluyendo huellas dactilares y el iris de los ojos. Si alguien no quería proporcionar estos datos, no podría acceder a los servicios de telefonía móvil que son importantes para la vida cotidiana.

Ahora bien, Josefina Román Vergara, Comisionada del INAI, explica que en síntesis (INAI, 2021):

“El 16 de abril de 2021 se crea el Padrón Nacional de Usuarios de Telefonía Móvil. Se integra con una base de datos de titulares de cada línea telefónica móvil, cuyo único fin es el de colaborar con las autoridades competentes en materia de seguridad y justicia en asuntos relacionados con la comisión de delitos en los términos de las disposiciones jurídicas aplicables”.

El padrón de datos personales de telefonía móvil en México implicaba el registro de datos biométricos en una base de datos de las compañías de telefonía móvil. El objetivo era utilizar esta información como un elemento de seguridad para intervenir números telefónicos en caso de delito y facilitar la investigación. Sin embargo, esto también podría vulnerar la privacidad de las personas, ya que sus datos biométricos estarían vinculados con su información personal y número telefónico, permitiendo el rastreo y verificación de su identidad; asimismo, el padrón fue aprobado debido a la alta incidencia de crímenes como el secuestro, la extorsión, los fraudes y la desaparición forzada en México, pero hay que considerar que proporcionar estos datos puede poner en riesgo la privacidad de la población.

En este sentido, el argumento central del Senado de la República (2021) fue el siguiente: “El objetivo, frenar delitos como extorsión y secuestro que, en muchos casos, se cometen con la utilización de celulares y desde centros penitenciarios” Asimismo también contempló multas para las empresas de telefonía que disimularán o alterarán los datos relacionados con el padrón. Y de acuerdo con el Senado: “Los concesionarios de telecomunicaciones deberán recabar y validar la información sobre la identidad, datos biométricos y domicilio del usuario, así como

proporcionar la información con la cual se integrará el registro. Además, realizarán las altas, bajas y demás movimientos asociados a las líneas telefónicas” (idem).

Este padrón presenta varios problemas de privacidad. Primero, existe el peligro de que una sola entidad recolecte toda la información relacionada con una persona, ya que todo sistema de información es vulnerable a ataques informáticos. En segundo lugar, hay el riesgo de que esta información se utilice para fines nocivos, como rastrear o espiar sin consentimiento a una persona, o usar los datos para obtener más información sobre un individuo utilizando técnicas de Vecinos cercanos y clasificación semiótica en la IA

Otro argumento en contra del padrón fue presentado por la Comisión de Derechos Humanos de la Ciudad de México. Según la Comisión: “En el caso de esta reforma, no sólo no existe un nexo causal entre el registro de información y la reducción del delito que justifique la medida, sino que además se está poniendo en riesgo la democratización de la telefonía para las poblaciones más aisladas” (CNDHCM, 2021). Esto se argumentó que tener un mayor registro de datos personales de una población no ayuda a reducir el crimen; genera un mecanismo indirecto para desincentivar la adopción de telefonía móvil y vulnera a sus usuarios, ya que corren el riesgo de que se suplanten su identidad.

La CNDHCM también argumentó que: “Es deber del Estado proteger la privacidad de una población que, en sí misma, se encuentra inmersa en un contexto de digitalización en el que sus datos personales son perseguidos como uno de los bienes más valiosos comercialmente”. Este argumento es importante porque reconoce la importancia de los datos y la situación de su valor mercantil, y ancla al principio de privacidad y el bien que fácilmente puede vulnerarse con la recopilación de datos, especialmente los biométricos.

Conociendo lo que es el padrón de datos personales de telefonía móvil, cómo funciona y cuáles eran los problemas éticos que podía causar, es necesario entender cómo los actores, las personas y las sociedades civiles respondieron para detenerlo. El primer mecanismo para detener el padrón fue una serie de amparos, o medidas jurídicas para proteger a una persona de las herramientas de

una acción legal en contra de su libre expresión de personalidad. El segundo mecanismo fue la promoción de acciones de inconstitucionalidad en contra del padrón

En el segundo sentido resalta la acción emitida por el Instituto Nacional del Acceso la Información (INAI). En palabras de Román-Vergara:

“...en sesión celebrada el 27 de abril de 2021 el Pleno del INAI acordó por unanimidad, interponer Acción de Inconstitucionalidad en contra de dicho Decreto; al respecto, cabe destacar que la Reforma Constitucional de 2014 en materia de transparencia y protección de datos personales otorgó al INAI la facultad de velar por los derechos de acceso a la información y de protección de datos personales”.

Ahora bien, con base en la teoría de la red de actores, es necesario examinar conceptualmente cómo respondieron los actores humanos ante el padrón de datos personales, utilizando el objeto de contorno, la traducción y los conceptos de análisis de Boltanski y Thevenot. En primer lugar, el padrón de datos personales se presenta como un objeto programado con códigos que buscan recopilar datos personales para investigar a los usuarios de teléfonos móviles. En segundo lugar, la sociedad civil busca equivalencias en la constitución para proteger el derecho humano a la privacidad, lo que también buscan las personas que interponen recursos de amparo.

Por lo tanto, la solución del problema surge de una amenaza al derecho de la privacidad. Esta se busca resolver mediante actores no humanos como la constitución y sus intérpretes, como los jueces. Mientras que el poder legislativo y ejecutivo apelan a los códigos de la constitución y su interpretación. Y esto se refleja, según Boltanski y Thevenot, una disputa en el mundo civil a pesar de que las implicaciones éticas del problema se dan en un régimen de ágape entre el mundo civil y un posible mundo digital con sus propios valores.

De esta manera el padrón de datos personales podría ser visto como un objeto regulatorio que busca establecer una equivalencia entre los usuarios de teléfonos móviles y los investigadores, pero también como un actor no humano que tiene un programa de recopilar datos personales y que se enfrenta a los antiprogramas de los usuarios que reclaman su derecho a la privacidad. Y desde esta perspectiva, el

análisis de los problemas en los derechos humanos de la IA no se limita a identificar los principios de justicia/equivalencia que están en juego, sino que también implica seguir las redes de actores humanos y no humanos que se asocian o se oponen para realizar sus programas o antiprogramas.

Ahora bien, es importante conocer cómo se resolvió el conflicto. Según la Red en defensa de los derechos digitales (R3D) en abril de 2022: "El Pleno de la Suprema Corte de Justicia de la Nación aprobó el proyecto de sentencia que declara la invalidez total del decreto que crea el Padrón Nacional de Usuarios de Telefonía Móvil (PANAUT)". Por lo tanto, la organización comenta que, con base en esta resolución, se crea un precedente para la protección de la privacidad de las personas y el acceso a las tecnologías de información y comunicación (al que se puede sumar, desde la perspectiva de este autor, el acceso a la IA).

La resolución del conflicto protege el derecho a la privacidad de los usuarios y fortalece las equivalencias para protegerlo y el derecho a la no discriminación si un agente beligerante utiliza los datos para discriminar a las personas. Esto significa que, si se aprueba otra iniciativa contra estos derechos, una persona o grupo de personas que se quejen podrán anclarse tanto a las herramientas cívicas como a la equivalencia de que existe un caso donde ya se protegieron estos principios. En otras palabras, la resolución del caso incluye un elemento en el código que protege los derechos de privacidad y el derecho a la no discriminación mediante un antecedente que puede entrar en acción en caso de un conflicto futuro.

Aun así, en México, el derecho de privacidad está en riesgo de ser violado por nuevas herramientas y plataformas con IA. Sin embargo, existen objetos anclados a equivalencias que pueden proteger este principio si se presenta una queja, lo que incluso sirve como marco de referencia para analizar el impacto de la IA en las plataformas en México. Los actores pueden responder a escenarios en los que se transgreden estos derechos ya sea creando nuevas equivalencias ligadas a objetos o utilizando equivalencias existentes; por ejemplo, el Reglamento Europeo en torno a la regulación de la IA "contiene normas específicas para la protección de las personas en relación con el tratamiento de datos personales que restringen

el uso de sistemas de IA para la identificación biométrica” (Reglamento [UE] 2024/1689, 2024).

No obstante, la desintegración de organismos autónomos como el INAI, la concentración de todos los organismos de ejecutivo y legislativo federal sin contrapesos que puedan contener nuevas reformas, hacen de estos intentos legislativos algo más plausible. Tan sólo a finales de febrero del 2025 se resucitó la propuesta de la Padrón de telefonía móvil. Esta nueva propuesta de manera muy puntual busca que cualquier individuo que posea un teléfono móvil debe registrarlo y proporcionar su credencial del Instituto Nacional Electoral (INE) o su Clave Única de Registro de Población (CURP) (Mares, 2025). Estos datos son extremadamente sensibles, ya que facilitan la identificación sin mayor problema de un usuario dado que esta información contiene elementos importantes como el sexo, fecha de nacimiento, y hasta lugar de residencia en el caso de INE; además de que la iniciativa no contempla candados como la protección de datos personales (idem).

Con solo tu nombre y fecha de nacimiento, un delincuente puede rastrear filtraciones de tus datos personales en internet, accediendo a información pública como perfiles en redes sociales, propiedades e incluso la tenencia de mascotas, y utilizar ingeniería social para comprometer tus cuentas en plataformas digitales (Ravichandran, 2024). Esta situación se agrava con iniciativas gubernamentales en México que, a través de aplicaciones en teléfonos inteligentes, pueden invadir la privacidad de los ciudadanos. Por ejemplo, las aplicaciones bancarias en el país requieren que los usuarios compartan su ubicación para poder utilizarlas; si un criminal puede rastrearte con dicha información, es preocupante considerar el alcance que podría tener el gobierno, o peor aún, el riesgo de que estos datos se filtren y terminen en manos de intermediarios de datos o delincuentes con conocimientos técnicos avanzados (Velíz, 2022); aún así, el intento de padrón anterior era mucho más agresivo, al menos con base en lo que se sabe de este nuevo intento, ya que buscaba recolectar datos biométricos como huellas dactilares, escaneos de iris y reconocimiento facial para demostrar la identidad de los usuarios de teléfonos móviles e Internet (Bischoff, 2020).

De nueva cuenta, la diferencia en la protección de la privacidad entre Japón y México refleja variaciones en la implementación y el cumplimiento de las leyes de protección de datos. Mientras que Japón ha establecido un marco robusto y proactivo para salvaguardar la información personal, México continúa enfrentando desafíos en la aplicación efectiva de sus regulaciones. Esto se puede deber a una actitud tibia por parte de un segmento importante de la población mexicana en cada proceso que implica un menor nivel protección de sus datos personales, como cuando se aprobó la desaparición del INAI, un hecho que no generó una movilización social significativa a pesar de su impacto en la transparencia y la protección de datos personales.

Ahora bien, esto no significa que Japón sea una utopía en cuanto a la privacidad. Por ejemplo, está el caso de Shiori Ito, una periodista japonesa que denunció públicamente haber sido violada por un destacado periodista, ejemplifica las dificultades que enfrentan las víctimas al alzar la voz en Japón. Tras su denuncia en 2017 quien recibió amenazas y comentarios negativos, incluyendo insultos y cuestionamientos a su credibilidad, lo que refleja una tendencia en la sociedad japonesa a mantener el silencio sobre temas considerados “vergonzosos”, lo que dificulta que las víctimas de agresiones sexuales denuncien públicamente debido al temor a represalias y a la exposición pública (El Mercurio S.A.P, 2018).

Sin embargo, la sociedad japonesa tiende a ser extremadamente reactiva frente a las transgresiones de la privacidad. Un ejemplo claro es la medida implementada para abordar el problema de los hombres que toman fotografías no consentidas a mujeres en el transporte público. En respuesta, el gobierno japonés hizo obligatorio que todas las cámaras, incluidas las de los teléfonos móviles, emitan un sonido de obturador audible cuando se capture una imagen (Oto, 2024). Aunque esta política ha enfrentado ciertos desafíos, refleja cómo tanto la sociedad como las autoridades japonesas actúan con determinación ante violaciones de la privacidad. Esto también se evidencia en otras medidas, como las multas impuestas a quienes toman fotografías en calles residenciales o a geishas en el distrito de Gion en Kioto (fotografía 1) debido a que los turistas extranjeros acosaban a estas mujeres, y entraban a tomar fotografías en zonas residenciales

dentro de estos barrios turísticos alterando el *wa*, o la armonía de la comunidad.



(Fotografía, 1. Señal que indica que está prohibido tomar fotografías en Gion, Kioto. Imagen tomada por el autor).

Ahora bien, en Japón, no existe un registro gubernamental obligatorio de números de teléfono. Aunque los residentes pueden registrar su *My Number*, un sistema de identificación nacional diseñado para optimizar los procedimientos administrativos al vincular información personal entre distintas agencias gubernamentales—reduciendo así la necesidad de documentos adicionales en trámites oficiales—su uso no es un requisito para acceder a los servicios de telefonía móvil, y esta información no está directamente asociada con el número de teléfono (Digital Agency, s.f.). No obstante, la *My Number Card* digital permite acceder a diversos servicios, como la solicitud en línea de apoyo para el cuidado infantil y trámites de mudanza, la consulta de información médica y del Manual de Salud Materno-Infantil, y el uso de plataformas privadas para abrir cuentas

bancarias, contratar servicios financieros o realizar pagos móviles sin efectivo (implementados progresivamente desde el 11 de mayo de 2023); además, desde el 20 de diciembre de 2023, es posible emitir certificados oficiales en tiendas de conveniencia, y en el futuro se prevé su uso como tarjeta de seguro de salud (Digital Agency, s.f.).

El caso del padrón de datos personales de telefonía móvil en México plantea el papel de los actores humanos, y no humanos para proteger la privacidad y resistir regulaciones y conductas invasivas, estableciendo precedentes legales clave. En México, los contrapesos institucionales, como el INAI, intentaron frenar el avance de estas tecnologías mediante mecanismos legales y de transparencia; sin embargo, su debilitamiento limitó su impacto, dejando a los trabajadores en una posición vulnerable. En contraste, en Japón, la presión surgió desde la base: las movilizaciones de los conductores y las demandas sociales obligaron a la plataforma a modificar sus algoritmos y políticas.

Esta diferencia sugiere que la resistencia efectiva no proviene exclusivamente de las instituciones, sino de la capacidad de los trabajadores y la sociedad para organizarse y exigir cambios. Con ello, a pesar de las victorias obtenidas en México, la reaparición de iniciativas similares resalta la fragilidad de estos avances y la necesidad de un marco robusto que proteja los derechos digitales en un contexto donde la IA y la recopilación masiva de datos continúan expandiéndose. En este sentido, la comparación con otros países subraya que la protección de la privacidad no depende solo de leyes, sino también de la conciencia social y la disposición de los ciudadanos para defender sus derechos en contextos donde los contrapesos institucionales son débiles o inexistentes, la movilización social se vuelve un factor clave para equilibrar el poder de las plataformas; esto plantea un desafío más amplio: si las reglas que gobiernan los algoritmos siguen diseñándose sin considerar las condiciones locales, ¿qué estrategias pueden adoptar los trabajadores y la sociedad para contrarrestar sus efectos negativos?

3. Los enfoques de I + D de la IA en México y Japón

La IA es una tecnología que influye en las sociedades que la adoptan y la desarrollan; en particular, afecta a sus perspectivas y actitudes sociales, como sus costumbres y sus formas de resolver los conflictos. Así, el nivel de inversión, desarrollo e implementación de la IA en un país determina en parte cómo su población y sus organizaciones se relacionan con esta tecnología y con los desafíos que plantea. Este es el caso de México y Japón, dos países con diferentes grados de avance en el campo de la IA. Por eso, es importante analizar el contexto y las condiciones en los que se desarrolla e implementa la IA en ambos países, para comprender mejor cómo sus sociedades se adaptan y reaccionan ante esta herramienta tecnológica.

En primer lugar, es necesario examinar por qué se tiene que invertir y desarrollar herramientas, plataformas y algoritmos de IA, en particular en el aprendizaje de máquina. En segundo lugar, hay que comprender cómo esta inversión ayuda a evitar la dependencia económica y a influir en el proceso de desarrollo de regulaciones en torno a las estrategias de IA, como se vio en el capítulo anterior. Finalmente, se debe vincular esto brevemente con las actitudes sociales en torno a la IA como resultado de la preparación de la población de cada país.

Hay opiniones divididas sobre por qué es importante invertir y desarrollar herramientas, plataformas y algoritmos de IA. Algunos argumentan que es un motor económico importante, mientras que otros argumentan que puede ayudar a reducir la dependencia de otros países. Gómez Mont et al (2020) comentan que América Latina y el Caribe (que incluye a México) "necesitan soluciones innovadoras para mejorar la calidad de sus servicios sociales. A ello puede contribuir la tecnología, haciéndolos más eficientes y sostenibles. Dentro del instrumental disponible, la IA puede desempeñar un papel clave, pero solo si cumple con determinados requisitos".

Aunque el aprendizaje de máquina ha alcanzado un punto de saturación, todavía hay oportunidades para desarrollar soluciones de IA relevantes. Según Andrew Ng en Coursera (s.f.), esto significa que todavía hay oportunidades para preparar y

desarrollar capital humano y ajustar la creación de nuevos trabajos capacitados para hacer frente a los puestos de trabajo que serán reemplazados por la automatización. En este sentido, la IA todavía representa un abanico de posibilidades para muchos países y personas emprendedoras.

Hay que resaltar que no se debe desarrollar ni implementar ni adoptar la tecnología de IA de manera aislada. Por lo tanto, es necesario crear una ruta para su uso responsable, como se ha visto en distintos países. Esto implica promover el uso responsable de la IA no solo en grandes compañías tecnológicas sino también en empresas emergentes e innovadoras en sectores como la salud, la educación y la seguridad social para que diseñen soluciones tanto para el sector privado como para el público.

Según Gómez Mont, et al (2020), para mejorar la infraestructura digital y, por tanto, el desarrollo de la IA, México implementó una Estrategia Digital Nacional (EDN) durante el período 2013-2018. Esta estrategia incluyó objetivos como la "transformación gubernamental", la "economía digital", la "transformación educativa", la "salud universal y efectiva", y la "innovación cívica" (Estrategia Digital Nacional: Transformamos México, 2018). Sin embargo, durante el mandato del presidente López Obrador, esta estrategia ha sido descuidada, a pesar de que existe una oficina adjunta a la presidencia encargada de mantener la comunicación al respecto (ídem). Además, no se han implementado nuevas medidas para abordar los desafíos relacionados con la IA.

A pesar de esto, se han implementado algunas estrategias para mejorar la situación digital en México. Por ejemplo, en 2019 se lanzó Data México, una iniciativa impulsada por el secretario de Economía que busca mejorar los datos públicos y basar la política económica en evidencia (ídem). Además, existen diversas estrategias locales enfocadas en crear *hubs* de internet municipales en diferentes áreas urbanas y de transporte, como la red de internet de la Ciudad de México.

El país cuenta con una amplia oferta educativa en el ámbito de la IA, tanto a nivel de licenciatura como de posgrado. Por ejemplo, la UAM Iztapalapa imparte una

asignatura de IA en la carrera de computación, que busca que los alumnos aprendan a aplicar los conocimientos de esta disciplina para resolver problemas (UAM-I, s.f); asimismo, la asignatura de Inteligencia Artificial se ofrece en la Licenciatura en Sistemas de Información de la DCCD y en la Ingeniería en Computación de la DCNI, ambas pertenecientes a la UAM Unidad Cuajimalpa; y a finales de 2024, la UAM entro en planes para lanzar una carrera en IA. La UNAM ofrece un posgrado especializado en IA, que define esta área como un campo de investigación y desarrollo que resuelve problemas complejos sin soluciones algorítmicas exactas o factibles, por su gran tamaño, complejidad estructural o la incertidumbre de los datos que manejan (UNAM, s. f.). Además, el país destaca por la capacidad de sus universidades para acoger a un gran número de estudiantes en programas gratuitos, como los que ofrecen el Tecnológico Nacional de México (TecNM), que agrupa a 266 instituciones en todo el territorio nacional; la UNAM; y el Instituto Politécnico Nacional (IPN) (Gómez Mont et al., 2020).

¿Ahora bien cuál es la estrategia de inversión de México y Japón en torno a la inteligencia artificial? Japón busca invertir 65 millones de dólares en microchips e inteligencia artificial (AFP, 2024). En este sentido, Japón ha demostrado un compromiso sólido con el avance de la IA siendo un ejemplo de ello el anuncio del Ministerio de Industria de la asignación de hasta 72,500 millones de yenes (aproximadamente 65 millones de dólares) en subsidios para cinco proyectos destinados a la fabricación de supercomputadoras enfocadas en el desarrollo de tecnología de IA en abril de 2024; asimismo, empresas como KDDI Corp. y Sakura Internet Inc. fueron beneficiadas con 10,200 millones y 50,100 millones de yenes respectivamente, con el objetivo de mejorar la infraestructura computacional del país y promover el desarrollo de sistemas de IA generativa (Nippon.com, 2024); además, iniciativas como *Woven City* de Toyota buscan crear entornos urbanos experimentales para probar robótica avanzada, IA y transporte autónomo de cero emisiones (Reuters, 2024).

Japón enfrenta desafíos en el desarrollo de talento especializado en inteligencia artificial, entendido como la generación de conocimiento y modelos propios que reduzcan la dependencia de software, modelos y propiedad intelectual extranjera.

A esto se suman dificultades en la producción de semiconductores avanzados, esenciales para la implementación de estos sistemas. Además de la escasez de talento, existen barreras socioculturales y lingüísticas que limitan la atracción de especialistas, a diferencia de países como Estados Unidos. La resistencia a cambios tecnológicos y el envejecimiento de la población agravan estos problemas, junto con preocupaciones clave en materia de derechos humanos, como la privacidad, la no discriminación y el acceso a la información (Deep Knowledge Group, 2024).

Japón cuenta con una red sólida de hubs y centros clave para la investigación y desarrollo en inteligencia artificial. A nivel gubernamental, el Ministerio de Economía, Comercio e Industria (METI) desempeña un papel central, mientras que en el sector privado destacan empresas como SoftBank, el Instituto de Investigación de Toyota y NEC Corporation. En el ámbito académico y científico, sobresalen la Universidad de Tokio, el RIKEN Center for Advanced Intelligence Project (AIP), el National Institute of Informatics (NII) y la Sociedad Japonesa para la Inteligencia Artificial.

Además, iniciativas como Sociedad 5.0 y la reciente alianza entre OpenAI y SoftBank Group reflejan una fuerte apuesta por la IA como eje estratégico del desarrollo tecnológico y económico del país. SoftBank ha comprometido una inversión de 3 mil millones de dólares anuales para desarrollar y comercializar “Cristal Intelligence”, una solución avanzada de IA empresarial. Para ello, estableció la empresa conjunta “SB OpenAI Japan”, con el objetivo de adaptar esta tecnología a grandes corporaciones en el país (SoftBank Group, 2025)

Es importante destacar nuevamente el aspecto demográfico de Japón. Ante el envejecimiento de su población y la reducción de la fuerza laboral, el país ha comenzado a adoptar e investigar aplicaciones de IA, especialmente en el desarrollo de robots para la atención hospitalaria y el cuidado de adultos mayores (Takenaka, 2025), así como en la seguridad aeroportuaria y otras áreas estratégicas. A esto se añade la dificultad de integrar fuerza laboral extranjera, ya sea por barreras culturales, el idioma o las restricciones de acceso al país.

En México, la inversión en I+D de IA ha sido impulsada en gran medida por actores internacionales; por ejemplo, Amazon Web Services (AWS) ha iniciado una inversión de 5.000 millones de dólares en Querétaro para construir una “ciudad digital” que ofrecerá servicios de computación en la nube, almacenamiento de datos y proyectos de inteligencia artificial. Este centro proporcionará más de 440 servicios tecnológicos a México y Latinoamérica, y se espera que genere más de 10.000 millones de dólares en el PIB en 15 años, además de crear un promedio de 7.000 empleos anuales (Suárez, 2025). Sin embargo, México enfrenta desafíos en innovación y generación de conocimiento, falta de talento, exceso de regulación y la creación de valor agregado (Palafox, 2025; MBN Staff, 2024).

Mientras Japón adopta una estrategia proactiva con inversiones sustanciales y colaboración entre gobierno, empresas y universidades para liderar en áreas específicas de IA, México se beneficia principalmente de inversiones extranjeras directas para impulsar su desarrollo en este campo. Estas diferencias se explican por las prioridades económicas y tecnológicas de cada país, así como por la infraestructura existente y el apoyo gubernamental hacia la innovación. Además de que Japón tiene una estrategia de innovación y desarrollo coordinada entre el sector privado y el gobierno; a modo de ejemplo, para impulsar la investigación y desarrollo en inteligencia artificial, la AI Japan R&D Network fue creada en 2019 por AIST, RIKEN y NICT como un consorcio de universidades e institutos de investigación pública, y en 2023 se reestructuró como una organización voluntaria que incluye empresas privadas para fomentar la colaboración y la aplicación de sus resultados en IA (MEXT, 2024).

4. Alphabet-Google: la IA y los derechos humanos en México y Japón

Alphabet-Google es una plataforma infraestructural. Como los plantean Van Dijck et al (2018), las plataformas infraestructurales han comenzado a infiltrarse en las configuraciones sociales existentes, ya que el ecosistema se entremezcla cada vez más con las estructuras institucionales establecidas. Dentro de las plataformas infraestructurales se encuentran las llamadas Cinco Grandes: Amazon, Microsoft, Apple, Meta y Alphabet-Google (idem), varias de estas empresas entre otras

cosas se dedican a servicios de servidores en la nube (Alphabet-Google, Amazon, y Microsoft con Azure) y hasta el momento de escribir la tesis, dos de estas dominan el espacio de las llamadas plataformas de IA generativa como es el caso de Microsoft que está bancando su importante inversión en OpenAI, y por supuesto, Alphabet-Google y sus laboratorios como *DeepMind*.

Ahora bien, más allá del gran abanico de productos y servicios que ofrece Alphabet-Google, es importante resaltar su peso a nivel estatal y en torno a las transgresiones de a nivel de asociaciones humanas en torno a la teoría de actor-red. En este sentido, quizás el caso más rescatado es el de El Salvador. En 2023, la asamblea legislativa presentó la Ley general para la modernización digital del estado. El objetivo de esta ley es “establecer las disposiciones necesarias para la implementación, desarrollo, transformación digital y modernización del Estado Salvadoreño” Asamblea Legislativa de El Salvador (2023).

A grandes rasgos, la ley busca establecer las disposiciones necesarias para la implementación, desarrollo, transformación digital y modernización del Estado. Esto es, al reconocer los compromisos internacionales que el Estado adquiera con entidades públicas o privadas. Además de los beneficios sociales y económicos que se derivan de la alianza con Google LLC. La ley también define los conceptos, principios, objetivos, competencias, obligaciones y derechos relacionados con la modernización digital del Estado, así como los mecanismos de coordinación, supervisión, evaluación y transparencia que se aplicarán.

En el caso puntual de Alphabet-Google, la propuesta de ley para la modernización digital del Estado salvadoreño, resalta en muchas instancias su alianza con esta empresa. En especial, dentro del párrafo VII donde grosso modo se menciona que, Mediante el Órgano Ejecutivo, el Estado salvadoreño firmó un acuerdo de colaboración estratégica con Google LLC, una empresa de tecnología, con el propósito de promover la transformación digital en el país, enfocándose en la reducción de costos y en el mejoramiento de los servicios que ofrece el gobierno, aprovechando los servicios y la tecnología que proporciona dicha empresa (Idem). Sin embargo, esta alianza estratégica también plantea una serie de interrogantes y

desafíos sobre el rol, la responsabilidad y la regulación de Google LLC como socio tecnológico del Estado salvadoreño.

Otro apartado en el que resalta el papel de Alphabet-Google dentro de la propuesta de ley es el Capítulo Primero, párrafo IV. En esta sección se menciona como Socio Estratégico a Google LLC, o cualquiera de sus empresas asociadas, filiales, oficinas, o entidades jurídicas que estén bajo su dominio o influencia, así como sus herederos o cesionarios en relación con la Alianza Estratégica, incluyendo sin restringir los compromisos, contratos, acuerdos o instrumentos que se usen para ejecutar la misma (idem). Así pues, de nueva cuenta, se ve cómo claramente la forma en la que Alphabet-Google no sólo aparece como una opción dentro de la estrategia de ley para un estado, sino como un socio estratégico a nivel infraestructural en un área de ecosistemas de servicios que ya es tan relevante como la producción y transmisión de electricidad, más en torno al papel de los datos y los algoritmos.

Dado este escenario, vale la pena preguntar ¿Qué implicaciones tiene para la soberanía, la seguridad, la privacidad y la transparencia de los datos públicos y personales que se alojan y procesan en los servidores de Google LLC? ¿Qué garantías existen de que Google LLC no hará un uso indebido, abusivo o ilegal de esos datos para fines comerciales, políticos o de otra índole? ¿Qué mecanismos de rendición de cuentas, control y sanción se aplicarán en caso de que Google LLC incumpla con los términos y condiciones de la alianza estratégica o viole los derechos humanos o las normas legales vigentes? Y a pesar de que no todos los estados y asociaciones plantean mecanismos jurídicos tan estrechamente relacionados con Alphabet-Google, este tipo de preguntas son pertinentes en torno al uso de los datos, el rol de las plataformas, y el empleo de la IA en México y Japón.

A. Política ética de la IA en Google

Es necesario conocer los lineamientos regulatorios de Google en torno a la IA para poder compararlos con las políticas y propuestas regulatorias que permiten explicar cómo y por qué la plataforma de movilidad integrada de Google se

implementa de manera diferente en México y Japón. También es útil examinar estos lineamientos regulatorios para comprender la perspectiva de la empresa en relación con la IA y sus desafíos, lo que permite comprender mejor cómo implementa sus productos y herramientas y cómo estos se relacionan con los derechos humanos.

De acuerdo con el sitio de Google AI (s.f.), la plataforma busca “crear tecnologías que resuelvan problemas importantes y ayuden a las personas en su vida diaria. Somos optimistas sobre el increíble potencial de la IA y otras tecnologías avanzadas para empoderar a las personas, beneficiar ampliamente a las generaciones actuales y futuras y trabajar por el bien común”. Esto muestra una perspectiva entusiasta de la empresa sobre el potencial social de la IA. Por lo tanto, se entiende que, al menos en la plataforma de Google Maps, la empresa del Valle del Silicio busca generar experiencias de flujo a través de herramientas y recursos tecnológicos.

El comunicado de Google AI reconoce que la tecnología de la IA plantea importantes desafíos, por lo que los desarrolladores explican la necesidad de establecer criterios de necesidad. Más importante e interesante aun, plantean áreas en las que no se piensa ni se desarrolla e implementa la IA, como una medida de prevención que va más allá de los principios éticos expuestos en este trabajo.

En cuanto a los objetivos sociales, la empresa hace énfasis en los avances que está haciendo la IA en áreas como la salud, el transporte, la fabricación y el entretenimiento. A pesar de los riesgos de la IA, el comunicado de Google AI establece que “tendremos en cuenta una amplia gama de factores sociales y económicos y procederemos cuando creamos que los posibles beneficios generales superan sustancialmente los riesgos y desventajas previsibles” (ídem).

Además, Google AI reconoce que la IA ayuda a comprender el significado de los contenidos a escala (es decir, con grandes cantidades de datos). No obstante, también se compromete a respetar las regulaciones y normativas, así como aspectos culturales locales de los países en los que opera. Y finalmente expresan

que continuarán “evaluando cuidadosamente cuándo hacer que nuestras tecnologías estén disponibles de forma no comercial” (ídem).

Google AI dedica un apartado completo en el que plantea su compromiso para evitar que la IA adquiera sesgos a partir de los datos en que se le entrena o que aprende por sí misma. Esto ayuda a prevenir prácticas o fenómenos como la discriminación automatizada por parte de la máquina. Y en palabras de la empresa: “Buscamos evitar impactos injustos en las personas, particularmente aquellos relacionados con características sensibles como raza, etnia, género, nacionalidad, ingresos, orientación sexual, capacidad y creencias políticas o religiosas” (ídem).

Asimismo, Google AI plantea el diseño de soluciones que promuevan y protejan la privacidad de los usuarios. Por lo tanto, la empresa explica que integrará los principios de privacidad desde la etapa del diseño de sus herramientas y plataformas de IA. Esto es importante, porque sirve como medida preventiva, y según la empresa: “Brindaremos la oportunidad de notificación y consentimiento, fomentaremos arquitecturas con garantías de privacidad y proporcionaremos transparencia y control adecuados sobre el uso de los datos” (ídem).

Otro punto central es el constante monitoreo de las prácticas de investigación, desarrollo y afinamiento de las herramientas de IA. El compromiso de probar constantemente las herramientas con esta tecnología para evitar resultados nocivos para las personas y la sociedad. Por lo tanto, la empresa explica: “En los casos apropiados, probaremos las tecnologías de IA en entornos restringidos y monitorearemos su funcionamiento después de la implementación” (ídem).

En relación con la responsabilidad, este es otro principio planteado en el sitio de Google AI. En particular, el apartado se centra en la responsabilidad social o responsabilidad con las personas. Y explican que: “Diseñaremos sistemas de IA que brinden oportunidades apropiadas para comentarios, explicaciones relevantes y atractivas. Nuestras tecnologías de IA estarán sujetas a la dirección y el control humanos apropiados” (ídem).

Otro punto para resaltar es que Google AI dedica un apartado completo a explicar que la empresa desarrolla la IA con base al método científico. Según el comunicado del sitio, el método científico se basa en el rigor, la integridad, la consulta y la colaboración. Asimismo, plantean los enormes beneficios que puede tener la IA en distintas ramas de la ciencia como la medicina, la biología y la química, y hacen énfasis en que se colaboran con distintos actores dentro y en torno a la comunidad científica:

“Trabajaremos con una variedad de partes interesadas para promover un liderazgo reflexivo en esta área, aprovechando enfoques científicamente rigurosos y multidisciplinarios. Y compartiremos responsablemente el conocimiento de la IA mediante la publicación de materiales educativos, mejores prácticas e investigaciones que permitan a más personas desarrollar aplicaciones útiles de IA” (ídem).

Otro punto de interés es el énfasis que Google AI le da nueva cuenta a la prevención de prácticas que puedan hacer un daño. Para ello, el sitio de la empresa deja claro que evitará desarrollar IA que se aplique en armas, en sistemas de vigilancia que atenten con la privacidad de las personas, y en medidas que contravengan las regulaciones internacionales. Además, plantea una serie de incisos en torno a medidas prevengan fenómenos que generen daño:

- “Propósito y uso principal: el propósito principal y el uso probable de una tecnología y una aplicación, incluido el grado de relación o adaptabilidad de la solución a un uso nocivo.
- Naturaleza y singularidad: si estamos poniendo a disposición tecnología que es única o más generalmente disponible.
- Escala: si el uso de esta tecnología tendrá un impacto significativo
- Naturaleza de la participación de Google: si proporcionamos herramientas de propósito general, integramos herramientas para clientes o desarrollamos soluciones personalizadas.” (ídem).

En este sentido los mediadores a favor de sus sistemas de IA en Alphebt-Google son aquellos que valoran la innovación, la diversificación, la eficiencia y el liderazgo de esta empresa en el mercado global. Dentro de estos mediadores se

encuentran los inversores que buscan mejorar sus ganancias, clientes, socios, empleados, académicos, medios de comunicación, entre otros, que reconocen el potencial y el impacto de los productos y servicios de Google, como el buscador, YouTube, Gmail, Google Maps, Google Cloud, entre otros. No obstante, quizás el grupo de mediadores más importantes dentro de Alphabet-Google, son los voceros que promueven la integración de sistemas de IA y dataficación, no sólo para mejorar los sistemas, sino también para aumentar y mantener su base usuarios, y con ello promover aún más herramientas en el mercado de anuncios.

Ahora bien, los mediadores en contra de sus sistemas de IA son aquellos que cuestionan la ética, la transparencia, la privacidad y la responsabilidad de esta empresa en el manejo de los datos personales, la competencia desleal, la evasión fiscal, la censura, la manipulación, etc. Dentro de estos mediadores se pueden encontrar activistas, reguladores, competidores, usuarios, organizaciones de derechos humanos, entre otros actores que denuncian y critican las prácticas y los riesgos de Alphabet-Google. Por ejemplo, cualquier persona que exponga una crítica informada, o que promueva una regulación a modo de antiprogramación respecto a las transgresiones de los sistemas de IA en las plataformas de esta empresa o su manejo de datos, se puede considerar como un mediador en contra.

ASPECTO	TRANSGRESIONES	MEDIADORES A FAVOR	MEDIADORES EN CONTRA
Derechos humanos	Los sistemas de IA de Google pueden violar el derecho a la privacidad, a la libertad de expresión, a la no discriminación, a la protección de datos personales y a la participación democrática de las personas que usan sus productos y servicios, o que son objeto de sus	Los mediadores a favor de los sistemas de IA de Google son aquellos que valoran la innovación, la diversificación, la eficiencia y el liderazgo de esta empresa en el mercado global. Dentro de estos mediadores se encuentran los inversores que buscan mejorar sus ganancias, clientes, socios, empleados, académicos,	Los mediadores en contra de los sistemas de IA de Google son aquellos que cuestionan la ética, la transparencia, la privacidad y la responsabilidad de esta empresa en el manejo de los datos personales, la competencia desleal, la evasión fiscal, la censura, la

	prácticas comerciales y de gestión de datos (Amnesty International, 2019; Zuboff, 2019).	medios de comunicación, entre otros, que reconocen el potencial y el impacto de los productos y servicios de Google, como el buscador, YouTube, Gmail, Google Maps, Google Cloud, entre otros.	manipulación, etc. Dentro de estos mediadores se pueden encontrar activistas, reguladores, competidores, usuarios, organizaciones de derechos humanos, entre otros actores que denuncian y critican las prácticas y los riesgos de Google.
Asociaciones humanas	Los sistemas de IA de Google pueden alterar las formas de relacionarse e interactuar de las personas en el ámbito digital y social, al generar una dependencia de sus plataformas, una pérdida de autonomía y una influencia en sus opiniones y comportamientos. Esto puede afectar la diversidad, la creatividad, la solidaridad y la participación social de las personas (Couldry & Mejias, 2019; Lanier, 2018).	Los mediadores a favor de los sistemas de IA de Google son aquellos que se benefician de las ventajas y las oportunidades que ofrecen sus plataformas, como el acceso a información, entretenimiento, educación, comunicación, etc. Dentro de estos mediadores se pueden encontrar usuarios satisfechos, creadores de contenido, educadores, periodistas, entre otros, que aprovechan las herramientas y los servicios de Google para sus fines personales o profesionales.	Los mediadores en contra de los sistemas de IA de Google son aquellos que se ven perjudicados o amenazados por las desventajas y los riesgos que implican sus plataformas, como la dependencia, la pérdida de autonomía, la influencia, la manipulación, entre otros. Dentro de estos mediadores se pueden encontrar usuarios insatisfechos, creadores de contenido censurados, educadores críticos, periodistas independientes, entre otros, que rechazan o cuestionan las herramientas y los servicios de Google por

			sus efectos negativos en la sociedad.
Nivel técnico	Los sistemas de IA de Google presentan limitaciones y sesgos técnicos que afectan su precisión, su fiabilidad y su transparencia. Estos sistemas dependen de la calidad de los datos, de los algoritmos y de los dispositivos que los implementan, y pueden ser vulnerables a errores, manipulaciones o hackeos,	Los mediadores a favor de los sistemas de IA de Google son aquellos que confían en la calidad, la eficacia y la seguridad de sus tecnologías, y que las utilizan para mejorar sus capacidades, sus procesos y sus resultados. Dentro de estos mediadores se pueden encontrar desarrolladores, investigadores e investigadoras, empresas y organizaciones, entre otros, que se apoyan en las soluciones y los recursos de Google para sus proyectos y sus objetivos.	Los mediadores en contra de los sistemas de IA de Google son aquellos que desconfían de la calidad, la eficacia y la seguridad de sus tecnologías, y que las evitan o las reemplazan por otras alternativas. Dentro de estos mediadores se pueden encontrar desarrolladores, investigadores e investigadoras, empresas y organizaciones, entre otros, que se oponen o se resisten a las soluciones y los recursos de Google por sus problemas y sus amenazas.
Nivel sociotécnico	Los sistemas de IA de Google implican una compleja red de actores humanos y no humanos que interactúan y se influyen mutuamente, generando efectos sociales, políticos, económicos y culturales que van más allá de la mera aplicación técnica. Estos sistemas están	Los mediadores a favor de los sistemas de IA de Google son aquellos que comparten sus fines, sus usos y sus consecuencias, y que los apoyan o los promueven. Dentro de estos mediadores se pueden encontrar gobiernos, instituciones, organizaciones, entre otros, que se alían o se	Los mediadores en contra de los sistemas de IA de Google son aquellos que discrepan de sus fines, sus usos y sus consecuencias, y que los rechazan o los impugnan. Dentro de estos mediadores se pueden encontrar gobiernos, instituciones, organizaciones, entre

	insertos en contextos históricos, institucionales y normativos que determinan sus fines, sus usos y sus consecuencias	benefician de las actividades y los proyectos de Google en diferentes ámbitos y sectores.	otros, que se enfrentan o se perjudican de las actividades y los proyectos de Google en diferentes ámbitos y sectores.

En resumen, Alphabet con su filial Google AI plantea lineamientos regulatorios en torno a la IA que claramente se anclan en los derechos a la privacidad, al acceso a la información y a la no discriminación. Además, la empresa expresa una serie de lineamientos adicionales para prevenir prácticas y fenómenos nocivos producto del desarrollo de la IA, particularmente en el desarrollo de armas y herramientas que dañen a los humanos. Sin embargo, en ningún momento Google AI expresó explícitamente respetar el principio de la entendibilidad, aunque se comprometió a crear y explicar herramientas educativas y prácticas de investigación que ayuden a comprender cómo funciona su IA y la forma en que toma decisiones; aun así, esto no elimina del todo la "caja negra" con la que trabajan sus sistemas

B. Alphabet y su ecosistema de recolección de datos: cookies y perfiles

El problema de las plataformas de Google en cuanto a la transgresión de derechos de privacidad, acceso a la información y derecho a la no discriminación se da en dos sentidos. El primero tiene que ver con los sistemas y ecosistemas que utiliza la plataforma, como aplicaciones, cookies y buscadores, para recolectar información. El segundo problema gira en torno a la recolección de datos a partir de los servicios y aplicaciones en las plataformas de Google que requieren que el usuario acceda a ellos mediante una cuenta, lo que también implica una mayor recolección de datos.

Así pues, es relevante dedicar un espacio para responder la forma en la que algunas funcionan las herramientas de recolección de Alphabet/Google. Además de examinar cómo éstas se aplican en distintos espacios y plataformas en México,

particularmente en algunos sitios web de gobierno. Y comparar los resultados con las propuestas regulatorias en torno a la IA en México y Japón, así como el libro blanco sobre la IA de Google donde se exponen propuestas y herramientas que la empresa ha desarrollado para abordar los problemas alrededor de esta tecnología.

Ahora bien, primero hay que conocer qué tipo de datos de usuarios y cómo los recolecta Alphabet-Google. En este caso, la empresa emplea una serie de herramientas para recolectar datos sobre sus usuarios. Estas herramientas, grosso modo, son *cookies* para rastrear la actividad de los usuarios en distintos sitios de internet, herramientas de rastreo en su sistema operativo Android en donde muchas veces se tiene acceso al distinto hardware donde está instalado (cámaras, micrófonos, entre otros), y el mismo buscador de internet de la empresa el cual está en proceso de implementar APIs con IA.

De acuerdo con Google (s.f): “Las cookies son pequeños fragmentos de texto que los sitios web que visitas envían al navegador. Permiten que los sitios web recuerden información sobre tu visita, lo que puede hacer que sea más fácil volver a visitar los sitios y hacer que estos te resulten más útiles”. En este sentido, las *cookies* se convierten en una herramienta con la que se recolecta una especie de huella digital cada vez que la persona visita un sitio. En este sentido, la empresa emplea cookies para distintos propósitos: funcionalidad, publicidad, seguridad, análisis y personalización (“se usan con fines de personalización mejoran tu experiencia proporcionándote contenido y funciones personalizados, dependiendo de tu configuración de [g.co/privacytools](https://www.google.com/privacytools)”) (ídem).

En cuanto al sistema operativo Android, la plataforma permite a diferentes aplicaciones acceder a la geolocalización de los usuarios, micrófonos y cámaras, entre otros aspectos del hardware. Sin embargo, según el Centro para la Organización Ética de la Cultura de la Universidad de Auburn (2022), esta característica se ha convertido en un problema: “Por ejemplo, se ha revelado que los teléfonos Android contienen funciones de registro de ubicación que permiten a la empresa recopilar las coordenadas GPS de sus usuarios, así como las coordenadas de las redes wifi-cercanas”. Por lo tanto, muchos servicios y

aplicaciones dentro del sistema operativo, como Google Maps, también pueden hacer uso de estas herramientas del dispositivo.

Otro servicio de Alphabet es su navegador Chrome, una herramienta importante en su ecosistema. Esta plataforma está preinstalada en la mayoría de los dispositivos con sistema operativo Android y es uno de los navegadores más utilizados en el mundo. Sin embargo, Alphabet planea dejar de utilizar cookies en Chrome mediante su iniciativa Privacy Box, que utiliza tecnología de IA. En su lugar, han introducido un nuevo algoritmo basado en IA llamado API de temas (Pisani, 2022).

Cabe destacar que Alphabet-Google cuenta con otras plataformas y servicios que forman parte de su ecosistema, como Google Maps y Google Analytics. Estas herramientas ofrecen diferentes funciones para los clientes, como direccionar a clientes potenciales (Bundeskartellamt - Homepage - Condition of objections issued against Google's data processing terms, s.f.). Otros servicios importantes son sus servicios de correo electrónico y la plataforma de videos Youtube. Este último recolecta información de la actividad de los usuarios y permite a las empresas colocar anuncios, siendo éstos su principal fuente de ingresos.

Asimismo, es necesario hablar de una de las piezas más importantes de la maquinaria de plataformas y ecosistemas de Alphabet-Google: su buscador web. Pero este elemento es quizás uno de los más delicados y controvertidos dentro de la empresa. Por ejemplo:

En Francia, multaron a Google con 57 millones de dólares por el Reglamento General de Protección de Datos (RGPD). La autoridad francesa de protección de datos afirmó que Google no reveló cómo se recopilan correctamente los datos en sus servicios (ídem).

Otro servicio relevante que ofrece Alphabet-Google es el de la publicidad. La publicidad es esencial para las finanzas de esta empresa, y al igual que la mayoría de los servicios gratuitos en internet, se alimenta de los datos de los usuarios. Según la Organización Ética de la Cultura de la Universidad de Auburn (2022): "La principal fuente de ingresos de Google es la publicidad. La compañía genera

aproximadamente \$134 mil millones en ingresos por publicidad. Google AdWords, ahora conocido como Google Ads, fue introducido por primera vez en 2000".

Alphabet-Google es capaz de ejercer un modelo de publicidad altamente eficiente gracias a los datos recolectados a través de su ecosistema de aplicaciones. Esto se traduce en que "Google no fija los precios de los anuncios, sino que los pone a subasta en su limitado espacio publicitario; las empresas envían ofertas sobre cuánto pagarán por acción del cliente, y las ofertas más altas generalmente obtienen más tiempo publicitario" (Organización Ética de la Cultura de la Universidad de Auburn, 2022). Sin embargo, el modelo de publicidad de Alphabet-Google, al igual que gran parte de su ecosistema, es controvertido, tanto así que a principios del 2023 el Departamento de Justicia de los Estados Unidos emitió una demanda antimonopolio en torno a este tema. De acuerdo con el Departamento de Justicia de los Estados Unidos (2023):

“Durante los últimos 15 años, Google se ha involucrado en un curso de conducta anticompetitiva y excluyente que consistió en neutralizar o eliminar a los competidores de tecnología publicitaria a través de adquisiciones; ejercer su dominio en los mercados de publicidad digital para obligar a más editores y anunciantes a utilizar sus productos; y frustrar la capacidad de utilizar productos de la competencia.

Entre los principales puntos anticompetitivos que señala el Departamento de Justicia de los Estados Unidos (2023) se encuentran:

- Adquisición de competidores: Participar en un padrón de adquisiciones para obtener control sobre las herramientas publicitarias digitales clave que utilizan los editores de sitios web para vender espacios publicitarios;
- Obligación de la adopción de las herramientas de Google: obligar a los editores de sitios web a utilizar sus herramientas recientemente adquiridas al restringir su demanda de anunciante única e imprescindible a su intercambio de anuncios y, a su vez, condicionar el acceso efectivo en tiempo real a su intercambio de anuncios en el uso de su servidor de anuncios del editor;
- Distorsionar la competencia de subastas: limitar las ofertas en tiempo real en el inventario del editor a su intercambio de anuncios e impedir la

capacidad de los intercambios de anuncios rivales para competir en los mismos términos que el intercambio de anuncios de Google; y

- Manipulación de subastas: manipular la mecánica de subastas en varios de sus productos para aislar a Google de la competencia, privar a los rivales de escala y detener el auge de las tecnologías rivales.

Las implicaciones de esta resolución serán relevantes para el contenido de esta tesis. Esto significa que si el Departamento de Justicia de los Estados Unidos (2023) decide dividir a Alphabet-Google en diferentes empresas, se tendrá que abordar el análisis de la compañía de manera diferente. Hay que resaltar que, aunque se eligió a Alphabet como uno de los casos empíricos de este trabajo debido a sus antecedentes en torno al derecho a la privacidad en México y Japón, no es la única empresa que ha enfrentado controversias legales relacionadas con el uso de datos personales; por ejemplo, a finales del 2022 Apple Inc enfrentó una demanda de la corte francesa debido al uso de datos personales sin el consentimiento de sus usuarios, según un sitio de noticias relacionadas con tecnologías de la información, *Techcrunch* (2023):

Una rara sanción de privacidad para Apple: el organismo de control de protección de datos de Francia, la CNIL, ha anunciado que impuso una sanción de 8 millones de euros (~8,5 millones de dólares) al fabricante del iPhone por no obtener el consentimiento de los usuarios móviles locales antes de colocar (y/o leer) identificadores de anuncios en sus dispositivos en violación de la ley local de protección de datos.

Una de las plataformas más usadas de Alphabet Google es la herramienta de *Google Maps*. la arquitectura de Google Maps en torno a la recopilación de datos para el funcionamiento de su IA plantea dudas sobre cómo respeta los derechos de privacidad, acceso a la información y derecho a la no discriminación. Además de que responde a dudas en torno a los problemas de responsabilidad y explicabilidad.

Google Maps es una plataforma de movilidad que utiliza una combinación de herramientas de geolocalización en tiempo real y APIs (interfaces de programación de aplicaciones) para comunicar y obtener información de diferentes fuentes (Amazon AWA, s.f.). Estas herramientas incluyen IA, especialmente

aprendizaje automático, para proporcionar retroalimentación al usuario a través de interfaces gráficas y recopilar datos de usuarios y gobiernos. Estos datos pueden incluir información sobre tráfico en tiempo real, horarios de transporte público y condiciones climáticas.

Google Maps utiliza una serie de herramientas, como la geolocalización en tiempo real y las APIs, para brindar información precisa y actualizada sobre la ubicación del usuario, así como estimar el tiempo de viaje y verificar si hay cierres en rutas o sistemas de transporte público. Además, la plataforma reconoce la ubicación de diferentes elementos y los clasifica en categorías como sitios turísticos, históricos, hoteles, escuelas y restaurantes. Como una plataforma de movilidad integrada, Google Maps está en constante actualización, añadiendo nuevos productos y servicios a su ecosistema.

En cuanto a la predicción del tiempo, esta es una de las tareas que la plataforma actualiza regularmente. Derrow-Pinion et al (2021) explican que la labor para estimar el tiempo de arribo “se puede describir como proporcionar un tiempo de viaje estimado dada una ubicación de inicio proporcionada por el usuario y una ruta sugerida por un sistema anterior (red), así como información temporal (la evolución de las condiciones del tráfico a lo largo del tiempo)”. Ahora bien —y como se mencionó arriba— Google Maps emplea distintas herramientas, y entre las que usa para estimar el tiempo de viaje destacan las GNN o redes neuronales gráficas: un subcampo del aprendizaje de máquina profundo.

“Al ser una red neuronal de gráficos, nuestro trabajo se basa directamente en los años anteriores de progreso en el aprendizaje de representación de gráficos. En particular, nuestro modelo se basa fundamentalmente en el marco Graph Network y sigue el proceso de codificación-decodificación y sigue el paradigma codificar-procesar-decodificar. Esto nos permite alinearnos mejor con la naturaleza iterativa de los cálculos de tráfico, así como con los algoritmos de búsqueda de rutas. Se sabe que mejorar dicha alineación ayuda a que los GNN generalicen mejor a una distribución diversa de grafos” (ídem).

Así pues, la IA es una herramienta fundamental para Google Maps. En primer lugar, porque le permite actualizar y entrenar de forma constante sus modelos; en segundo lugar, porque esto ayuda a generar procesos de predicción estimación

cada vez más finos, lo que contribuye a una mejor experiencia de flujo en torno al viaje. Por ello, Dicker, R. (2021, 18 mayo) argumenta:

“La IA es una parte fundamental de lo que hace que Google Maps sea tan útil. Con él, podemos mapear carreteras 10 veces más rápido que hace cinco años, y podemos llevar mapas llenos de información útil a prácticamente todos los rincones del mundo. Hoy, le brindamos una mirada detrás de escena de cómo la AI hace posibles dos de las características que anunciamos en I/O”¹⁰.

Otro punto de interés es cómo funcionan las GNN de Google Maps. Como toda herramienta de aprendizaje de máquina y redes neuronales artificiales, las GNN funcionan como un triple de proceso de recolección de datos, los algoritmos ordenan esos datos y generan una predicción o resultado, y la capacidad de cómputo para procesar los datos y poner en marcha los algoritmos. Pero esclarecer más este punto, hay recuperar la explicación de la propia de compañía:

“...entrenamos nuestros modelos de aprendizaje automático en dos conjuntos de datos. El primer conjunto de información proviene de teléfonos que utilizan Google Maps. Los sensores de los teléfonos móviles pueden determinar la desaceleración a lo largo de una ruta, pero estos datos son muy propensos a generar falsas alarmas porque su teléfono puede moverse independientemente de su automóvil” (ídem).

La otra fuente de datos con la que Google Maps entrena a su plataforma proviene de las rutas que recorren de estos cuando se detectan en pantalla un automóvil de la plataforma de Android Auto (ídem). En este sentido, los investigadores de Alphabet inc aclaran que, si bien este conjunto particular de datos es pequeño, no deja de ser preciso debido a que es sumamente estable al estar conectado a la pantalla de los automóviles tienen aplicaciones ligadas a su ecosistema. Por ello, Google Maps es capaz de aprender información en tiempo real sobre situaciones falsas y reales como aceleración y desaceleración, lo cual hace que su modelo sea más preciso.

Para crear un flujo entre los sistemas de transporte multimodales (a pie, en bicicleta, en metro, en tren, entre otros), Google Maps también utiliza la IA. Según Dicker (2021), “cuando camina, anda en bicicleta o toma el transporte público, la

¹⁰ I/O es una abreviatura en inglés del concepto del Internet de las cosas, o del internet como parte de un ecosistema más allá de los ordenadores (computadoras) y de los teléfonos inteligentes.

IA también está allí para ayudarlo a moverse de manera segura y fácil”. Cabe destacar que para implementar este flujo en modos de transporte más allá del automóvil, la plataforma ha buscado la forma de integrar datos estructurados y no estructurados de mayor complejidad. En este sentido Dicker (2021) explica:

“El modelo debe comprender no sólo qué son los objetos, sino también las relaciones entre ellos, como dónde termina exactamente una calle y comienza una acera. Con estos nuevos modelos de escena completa, podemos detectar y clasificar amplios conjuntos de características a la vez sin sacrificar la precisión, lo que nos permite mapear una sola ciudad más rápido que nunca”.

Google Maps utiliza grandes cantidades de IA para poder reconocer la mayor cantidad de objetos y elementos en un espacio. Esto demuestra que el proceso de entrenamiento de un modelo de movilidad integrada o inteligente de Google Maps es sumamente complejo y requiere una gran cantidad de datos, poder de cómputo y herramientas de IA avanzadas para funcionar. Sin embargo, una vez que se entrena un modelo, este se puede escalar y trasladar a otras ciudades: “Por ejemplo, las aceras, los bordillos y los semáforos se ven similares en Atlanta y Ciudad Ho Chi Minh, a pesar de estar a más de 9,000 millas de distancia. Y en Madrid funciona el mismo modelo que en Dallas, algo que a primera vista puede resultar difícil de creer” (idem).

Otro punto importante para destacar es que Google Maps no es un ecosistema que funcione por separado. En este sentido, Lau (2022) menciona que la plataforma se asoció con el laboratorio de DeepMind, una compañía de investigación en IA que compró Alphabet y que ahora está ayudando a mejorar tanto la capacidad de predicción del tráfico de Google Maps como la estimación y precisión del tiempo estimado de llegada o ETA: “Nuestras predicciones de ETA ya tienen una barra de precisión muy alta; de hecho, vemos que nuestras predicciones han sido consistentemente precisas para más del 97% de los viajes”.

En relación con la pandemia del Covid-19, la flexibilidad de Google Maps como plataforma de movilidad integrada con IA resultó relevante. Esto es, porque a diferencia de los productos de software tradicionales con instrucciones de programación rígidas, Alphabet Inc puede afinar los modelos de Google Maps y las fuentes de datos que utiliza para entrenarlos, con el objetivo de mantener un

flujo en torno al tiempo estimado de llegada. Por lo tanto, los usuarios obtienen certeza constante, sin importar cómo las condiciones globales afectan a las situaciones locales.

Lau (2022) explica que "recientemente actualizamos nuestros modelos para hacerlos más ágiles, priorizando automáticamente los patrones de tráfico históricos de las últimas dos a cuatro semanas y des-priorizando los patrones de cualquier momento anterior". En cuanto a las nuevas fuentes de datos u otras fuentes de datos, el investigador de Alphabet Inc menciona que se incluyen datos provenientes de gobiernos y otras organizaciones, y datos fiables de los comentarios de los usuarios. Estos datos permiten a Google Maps integrar las reglas locales de tráfico, por un lado, y los elementos de contingencia de viaje como accidentes e incidentes viales reportados por usuarios, por otro.

Señala Yakobovitch, que además de actualizar su plataforma de movilidad integrada o inteligente en situaciones de contingencia como una pandemia, Google Maps también actualiza su plataforma para ofrecer nuevas herramientas y servicios. En este sentido, la plataforma está proporcionando nuevas funciones a las personas con diferentes capacidades motrices para acceder a mejores rutas mediante mecanismos de recopilación de datos de crowdsourcing (2020). Y al ser escalable, esto no se limita a una ciudad o región en la que se entrenaron las bases para la actualización, sino que también se expande a diferentes regiones del mundo:

"Los usuarios ahora pueden acceder al historial de sus viajes en Google Maps, explorar áreas e incluso participar en la generación de contenido, como calificar lugares y edificios. La función de guardar facilita la organización de viajes más tarde al recordar a los usuarios. Mediante el crowdsourcing de datos, Google ofrece información sobre una ubicación determinada que los usuarios desean conocer" (idem).

En el caso de México, es importante conocer de dónde obtiene Google Maps los datos para hacer más eficiente su IA. De acuerdo con el sitio de Avisos legales de la plataforma, Google Maps obtiene datos del INEGI y del "Mapa de Ubicación de Códigos Postales para la Ciudad de México" publicado por Correos de México (accesible en <https://datos.gob.mx/> el 28-03-2022) y del "Mapa de Ubicación de

Códigos Postales para Jalisco" publicado también por Correos de México (accesible en <https://datos.gob.mx/busca/> el 28-03-2022). (Alphabet Inc, s.f.).

Aun así, es necesario destacar que muchos de estos datos están sesgados hacia infraestructuras formales. Por lo tanto, en ciudades con sistemas de transporte que tienen distintos elementos en los que no participa el gobierno, como los sistemas de transporte concesionados o informales en el área metropolitana de la Ciudad de México, Google Maps recopila suficientes datos para hacer predicciones de tiempo estimado de llegada. Según el sitio de la propia organización, "WhereIsMyTransport fue la primera organización en mapear la extensa red de transporte público de la Ciudad de México en su totalidad. En la mayoría del mundo, los datos precisos y confiables de la red de transporte público raramente están disponibles desde una fuente central. La colaboración entre los datos de *WhereIsMyTransport* ha permitido fortalecer las opciones de ruta y la capacidad de predicción de Google Maps.

La interfaz gráfica de la plataforma también estima el costo total de los pasajes con base en los datos que va actualizado de las tarifas de estos. Esto permite que la persona usuaria no solamente planee la mejor ruta para llegar a su destino, sino también optar por la opción más económica sin necesidad de realizar el cálculo del costo total de todos los transbordos. ¿Ahora bien cuáles son los rasgos particulares de Google Maps en Japón? Al momento de entrar a la página de avisos de privacidad de avisos legales de Google Maps sobre Japón aparecen una serie de apartados en torno al derecho de autor. Esto contrasta enormemente con el caso mexicano donde solamente se mencionan las fuentes gubernamentales de los datos y ni siquiera se menciona a organizaciones como *WhereIsMyTransport* y con ello, el primer párrafo sobre el aviso legal de plataforma en Japón se lee así:

“Los servicios de Google Maps en Japón incluyen datos de listados de empresas modificados por Google en función de datos protegidos por derechos de autor proporcionados por terceros. El uso no autorizado de dichos datos puede infringir los derechos de autor de Google u otros. Además, el uso comercial está estrictamente prohibido. Los usuarios no pueden recuperar datos de listados de empresas” (Alphabet Inc, s. f.).

Asimismo, en el apartado legal sobre Japón, la plataforma es más precisa y clara en torno a problemas relacionados con la responsabilidad. En este sentido, el sitio de aviso legal explica: “Google y los terceros con licencia para usar los datos que se muestran en Google Maps no serán responsables de ningún daño (incluidos los daños indirectos) que resulten del uso de estos datos” (ídem). De esta manera, la plataforma se está previniendo de cualquier acción legal, y por ello de cualquier responsabilidad, en torno al uso de los datos que se encuentran en ella, así como de las empresas y de las organizaciones que los proveyeron.

Google Maps en Japón también tiene un mayor número de fuentes de datos. A diferencia de México, la plataforma emplea datos de un crisol de organizaciones, empresas e institutos de investigación, y para cada uno de ellos genera un apartado donde explica las condiciones de uso y propiedad intelectual de esos datos. Por ejemplo:

“El archivo de diccionario KANJIDIC utilizado para crear imágenes de mapas de Japón o contenido basado en mapas tiene derechos de autor del Grupo de Investigación y Desarrollo de Diccionarios Electrónicos (© por el Grupo de Investigación y Desarrollo de Diccionarios Electrónicos). Se utiliza bajo la licencia de Investigación y Desarrollo de Diccionarios Electrónicos. Grupo” (ídem).

Entre las empresas que consultó Google Maps para obtener datos para su plataforma en Japón se encuentran grupos que forman parte de los llamados *keiretsu* (系列) que “abarcan a los proveedores y clientes de una sola gran empresa, como Toyota Motors” (Mork y Nakamura, 2004). En el caso de Toyota Motors, el *keiretsu* comparte datos del centro de regulación de tráfico basados en el Centro nacional de promoción de actividad vial en torno a la plataforma de Toyota *Mapsmaster*. De acuerdo con el sitio, se destacan las siguientes condiciones legales en torno a los datos de Toyota Motors: “Los datos de regulación de tráfico no se pueden copiar, duplicar, procesar o modificar sin permiso. Si los datos de regulación de tráfico difieren de la regulación de tráfico en el sitio, dé prioridad a la señal / señal de regulación de tráfico real, etc. cuando conduzca. Los datos de regulación de tráfico no se pueden copiar, duplicar, procesar o modificar sin permiso” (Alphabet Inc, s.f).

Con respecto a los datos del sistema de movilidad, Google Maps tuvo que recurrir a las fuentes de una gran cantidad de organizaciones y empresas. Esto es, porque el sistema de transporte en Japón está integrado por una gran cantidad de empresas estatales y privadas que muchas conviven no solamente en el mismo espacio geográfico, sino que también alrededor de mismas soluciones de movilidad. Pero entre estas empresas, la que más resalta por su extensa red de trenes, autobuses y trenes de alta velocidad (Shinkansen) se encuentra Japan Rail o JR y sus distintas divisiones (JR East, JR West, entre otras); y en el caso de JR, el sitio de aviso legal de Google Maps plantea lo siguiente:

“El contenido de los datos de JR Bus se basa en los datos creados por Kotsu Shimbunsha Co., Ltd. para el 4 de junio del año de Reiwa. Se prohíbe la reproducción/copia o procesamiento no autorizado de estos datos de tiempo en medios electromagnéticos” (ídem).

Con respecto a los derechos humanos como la privacidad, el acceso a la información y el derecho a la no discriminación, así como los problemas de responsabilidad y explicabilidad, es relevante examinar hasta qué punto Google Maps y sus herramientas de IA han transgredido. En lo que respecta a la privacidad, el caso más reciente donde Google Maps violó este principio en Japón ocurrió en 2010. De acuerdo con el sitio del ministro de Asuntos Internos y Comunicaciones de Japón, en 2010 se reportó lo siguiente:

“El 14 de mayo de 2010, Google Inc. de los Estados Unidos anunció que había recopilado por error parte de la comunicación a través de LAN inalámbrica al tomar imágenes alrededor de la carretera con un automóvil con vista a la calle (*). En respuesta, el ministerio solicitó a Google Inc. informar sobre los hechos a la luz de lo dispuesto en el artículo 4 (Protección de la Confidencialidad) de la Ley” (邱丞漁達, s. f.).

Como respuesta, el gobierno japonés aplicó una serie de medidas cautelares citando su ley de protección de datos personales vigente:

- “(A) Eliminación de registros (excluidos los anteriores al establecimiento de la comunicación) relacionados con la comunicación a través de LAN inalámbrica proporcionada por el operador de telecomunicaciones almacenados en el servidor actualmente provisto para el servicio.

- (B) Recopilación y registro de información relacionada con la comunicación después de que se establezca la comunicación. Prevención de la recurrencia de casos y formulación de medidas para el cumplimiento legal futuro
- (C) Difusión de las circunstancias, estado de respuesta, medidas de prevención de reincidencia, etc. con respecto a este asunto en japonés” (ídem).

Con respecto a las herramientas de IA, Alphabet-Google ha enfrentado controversias. Un ejemplo es cuando la empresa estaba desarrollando un sistema de IA para mejorar el diagnóstico de problemas médicos. En 2019, Google y la Universidad de Chicago fueron demandados en una demanda que acusaba a la empresa y a la universidad de no eliminar los datos identificables de los registros médicos.

Otro aspecto relevante es el derecho a la privacidad, específicamente el derecho al olvido. Según Esposito (2022), el Tribunal de Justicia Europeo dictó una sentencia en 2014 que otorgaba a los ciudadanos el derecho a solicitar la eliminación de los resultados de búsqueda web relacionados con su nombre. Esto se conoce como derecho al olvido. Sin embargo, este derecho también ha generado controversias. En México, algunos críticos temen que este derecho sea abusado por personas poderosas para eliminar información desfavorable de su pasado (Bundeskartellamt, s. f.). Por lo tanto, es importante señalar que, aunque existe la posibilidad de eliminar datos personales de navegación en Google en México, este sistema puede ser abusado con fines maliciosos.

Hay que recordar que actualmente Alphabet-Google desarrolla herramientas con IA que planean sustituir las cookies en varios de sus sistemas bajo la forma de *Topics API for Privacy Sandbox*. De acuerdo con Google:

Comenzamos la iniciativa Privacy Sandbox para mejorar la privacidad web de los usuarios, al tiempo que brindamos a los editores, creadores y otros desarrolladores las herramientas que necesitan para construir negocios prósperos, asegurando una web segura y saludable para todos (Goel, 2022).

La forma en que estas API de Google generan perfiles de los usuarios para mandar anuncios con base en estos y mantener control más estricto de la privacidad se da mediante la IA. En este caso, Pasini explica:

La integración de IA en la publicidad es la clave para el reemplazo de cookies. Al puede procesar varios tipos diferentes de datos en muchos formatos. Al recopilar datos del usuario en tiempo real, como el historial de búsqueda, las preferencias personalizadas, la demografía y la información meteorológica, la IA puede analizar el comportamiento de los usuarios, encontrar patrones y personalizar las recomendaciones de anuncios para orientar los intereses de productos actuales de los usuarios (Pasani, 2022).

En resumen, la API funciona mediante un mecanismo donde los algoritmos aprenden y almacenan los intereses de los usuarios de forma local, para luego eliminarlos en un plazo de tres semanas. Esto es similar a cómo muchos sistemas crean perfiles de usuarios basados en sus intereses o interacciones en el sitio. Según Pasani (2022), a diferencia de los sistemas con cookies, en Topics API for Privacy Sandbox el “...algoritmo procesa toda la información del usuario localmente sin compartir información potencialmente confidencial o privada con servidores externos”.

Esto se ve ligado a la forma y el lugar donde se procesa la información, y la capacidad de decidir sobre esta: “al proporcionar a los sitios web sus temas de interés, las empresas en línea tienen una opción que no involucra técnicas de seguimiento encubiertas, como la toma de huellas digitales del navegador, para continuar publicando anuncios relevantes” (Goel, 2022). No obstante, la iniciativa de *Topics API for Privacy Sandbox* aun está en proceso experimental. Por ello, muchos sitios de internet, entre sitios de gobierno y universidades públicas, aun emplean sistemas de cookies (algunas invasivas como las de fingerprinting) que vulneran la privacidad de las personas usuarias.

Por ejemplo, en los resultados de un estudio realizado por el autor como voluntario en la función de científico ciudadano del CSI-COP¹¹, se analizaron 20 sitios web relacionados con la actividad laboral del autor. 11 de estos sitios usan cookies de

¹¹ Es importante agradecer al maestro Jaimz Winter por desarrollar la herramienta en Excel que se empleó para este análisis. Asimismo, es necesario reconocer el trabajo de la doctora Huma Shah por permitir que se usará este recurso dentro del corpus de la tesis.

Google, tres de ellos son instituciones públicas y educativas. Esto representa un 55% de los sitios analizados que usan cookies de Google y un 15% de instituciones públicas y educativas.

En relación con Alphabet-Google y la privacidad, este hallazgo sugiere una amplia adopción de tecnologías de Google en sitios web de instituciones públicas y educativas. Además, el uso de cookies de Google por parte de estos sitios puede tener implicaciones en cuanto a privacidad y transparencia, ya que permiten el seguimiento y recopilación de datos de los usuarios. Por ende, se requieren mecanismos regulatorios que atiendan esta problemática, así como las consecuencias en los derechos humanos de los sistemas que está desarrollando e implementando la empresa.

Además, es importante señalar otras controversias jurídicas que ha enfrentado Google. En México, uno de los casos más relevantes fue el de Ulrich Richter Morales, quien solicitó a la empresa borrar sus datos personales. El afectado:

“...envió una carta a Google en México, en la que solicitaba *amablemente* a su entonces director, Lino Cattaruzzi, la eliminación del blog del historial de búsqueda de Google, por considerar que el contenido, además de constituir un delito por suplantación de identidad, contenía –una serie de comentarios y expresiones insultantes e insinuaciones insidiosas en mi contra, correspondientes a los meses de mayo y junio de 2014, que atentan en contra de mis derechos al honor, reputación e integridad profesional, haciéndose un uso indebido de mi imagen–” (Galindo, 2017).

Con base en esto, se plantean antecedentes en México donde existe la posibilidad de exigir una especie de derecho al olvido que garantice la privacidad y protección de datos personales. Ahora bien, la pregunta si esto aplica a los otros sistemas y plataformas de Google, y con base en su política de privacidad, la respuesta es sí. Esto es, porque a partir de 2012 la empresa estableció una política de privacidad única para gestionar todas sus plataformas, lo cual tuvo otras consecuencias como la capacidad de compartir datos que recolecta entre todos sus servicios (Bundeskartellam, s. f.).

Antes de analizar brevemente el caso de Alphabet-Google en torno a México y Japón, hay que responder cómo se articulan estos problemas en torno a la IA,

dado que la empresa busca usar API con IA para combatir problemas de privacidad derivadas de las cookies. Y también es importante explicar brevemente la perspectiva de Alphabet-Google con respecto a la regulación de la IA. En este caso, la empresa busca usar la IA para proteger y mejorar la privacidad de las personas, además su perspectiva en torno a la regulación de la IA plantea estos puntos:

- Esbozar los flujos de trabajo básicos y los estándares de documentación para contextos específicos suficientes para demostrar diligencia por comprobaciones de seguridad.
- Establecer marcas de certificación de seguridad para indicar que un servicio ha sido evaluado pasando pruebas específicas para aplicaciones críticas.
- Colaboración humano-AI
- Determinar contextos en los que la toma de decisiones no debe ser completamente automatizada por un sistema AI, sino que requeriría un "humano en el bucle" significativo.
- Evaluar diferentes enfoques para habilitar la revisión y supervisión humana de los sistemas AI.
- Estructuras de responsabilidad
- Evaluar las debilidades potenciales en las reglas de responsabilidad existentes y explorar reglas complementarias para aplicaciones específicas de alto riesgo.
- Considerar marcos de refugio seguros específicos de la industria y límites de responsabilidad en los dominios en los que hay preocupaciones de que las leyes de responsabilidad puedan desalentar la innovación beneficiosa para la sociedad.
- Explorar alternativas de seguros para los escenarios en los que las reglas de responsabilidad tradicionales son insuficientes o inviables (Google, s.f).

En su libro blanco *Las perspectivas sobre los problemas en la Gobernanza de la IA*, Google propone soluciones a los problemas de responsabilidad de la IA que implican transgresiones a los derechos humanos. La empresa plantea el dilema de cómo equilibrar la necesidad de regulación para proteger la privacidad con el potencial positivo de la IA en la sociedad. Estos mismos dilemas también se discuten en las propuestas regulatorias de la Unión Europea, Japón y en la Agenda Nacional para la IA en México, donde se plantean áreas de aclaración y acciones sugeridas como estándares de explicabilidad, evaluación de niveles de justicia, consideraciones de seguridad, colaboración humano-IA y marco de responsabilidad (Sánchez Rodríguez & Rodríguez, 2020).

En México y Japón, la regulación ha sido un tema importante en la discusión sobre los problemas de privacidad relacionados con Google. Ambos países han tomado medidas para abordar las preocupaciones relacionadas con el uso de datos personales por parte de la empresa, como el problema de reconocimiento facial en Japón y las resoluciones de tribunales mexicanos en torno a una persona que solicitó la cancelación, bloqueo y supresión de su información en el motor de búsqueda de Google (Expediente PPD.0094/14; INAI, 2014). En México, la Autoridad Federal ha investigado a Google por posibles violaciones a la Ley de Protección de Datos Personales, y ha emitido recomendaciones para mejorar la transparencia y privacidad en los servicios de la empresa; en Japón, la Agencia de Protección de Datos Personales ha llevado a cabo inspecciones de Google en relación con la recolección y uso de datos personales, y ha emitido directrices para garantizar el cumplimiento de las leyes de protección de datos personales.

En cuanto a las propuestas regulatorias en torno a las herramientas de IA de Google, México y Japón han propuesto medidas para garantizar que se cumplan los principios de transparencia, responsabilidad e imparcialidad en el uso de la tecnología. En México, se ha propuesto la creación de una autoridad independiente encargada de supervisar el uso de IA en el sector público y privado, y se ha promovido la formación de un marco ético para el desarrollo y uso de IA. Por otro lado, en Japón, se han propuesto medidas para garantizar la transparencia en el uso de datos personales por parte de las empresas de

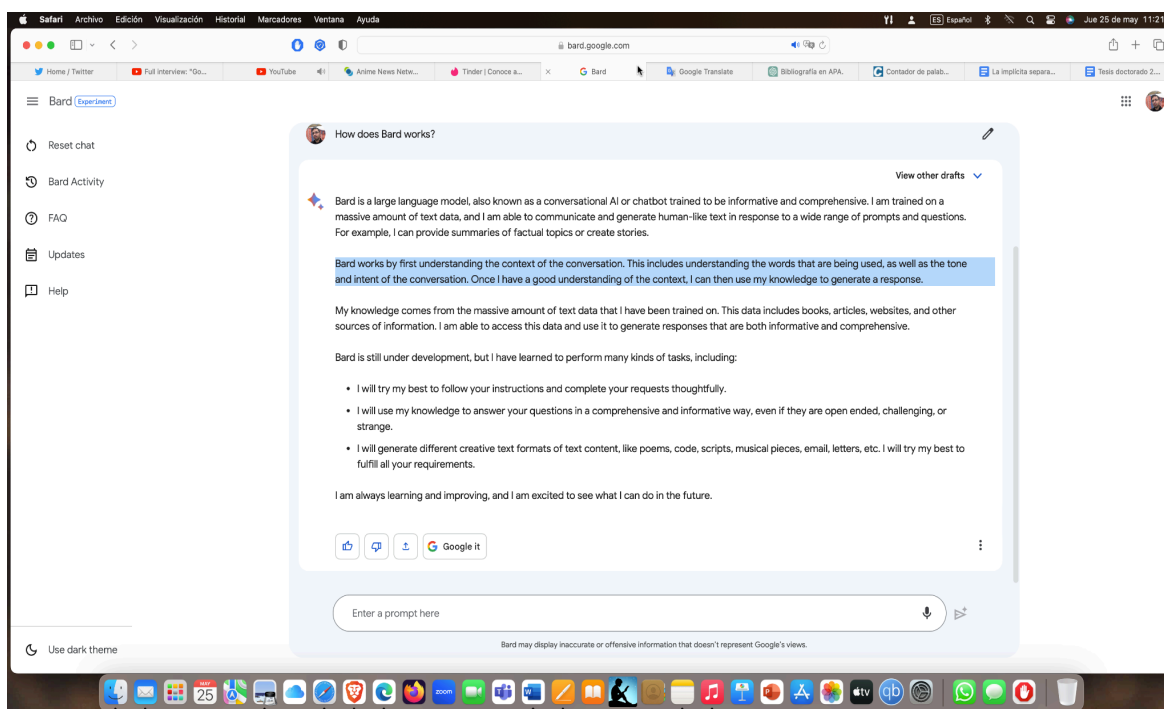
tecnología, y se ha promovido la formación de un marco ético para el desarrollo y uso de IA: estas propuestas buscan garantizar un uso responsable y ético de la tecnología de IA, y proteger los derechos humanos de los ciudadanos.

Alphabet-Google es una empresa que ha transgredido el derecho a la privacidad en México y Japón. Sin embargo, los mecanismos reguladores existentes en ambos países han resultado eficaces para resolver las quejas, y en algunos casos los argumentos presentados por las personas que se quejaron se han elevado al nivel de autoridad. Sin embargo, estos mecanismos reguladores resultan limitados frente a los nuevos desafíos de la IA en cuanto a derechos y responsabilidad. Mientras Japón ha avanzado en una propuesta reguladora basada en el marco de la Unión Europea, es importante conocer si México ha articulado propuestas reguladoras más allá de las recomendaciones de la Agenda Nacional de IA, especialmente para enfrentar desafíos relacionados con plataformas como Uber, sistemas de reconocimiento facial y gigantes tecnológicos como Alphabet-Google.

¿Pero cómo se explica la resistencia ante la transgresión? Según Latour (2008), un programa es lo que un actor quiere hacer, un antiprograma es lo que otro actor hace para impedirlo o modificarlo, y un antiprograma es lo que el primer actor hace para contrarrestar al segundo y seguir con su programa a estos derechos. En el caso de Alphabet-Google, el programa es recolectar datos de sus usuarios por medio de servicios hasta cierto grado gratuitos para vender anuncios.

No obstante, este programa de recolectar datos ya implica en sí una transgresión al derecho a la privacidad. Además, estos datos que son recolectados de los usuarios pueden generar sesgos en sistemas y aplicaciones como Bard, la aplicación de Redes Naturales de Lenguaje de Alphabet Google que funciona "...entendiendo primero el contexto de la conversación. Esto incluye la comprensión de las palabras que se utilizan, así como el tono y la intención de la conversación. Una vez que tengo una buena comprensión del contexto, puedo usar mi conocimiento para generar una respuesta" (Bard sesión de 25 de mayo, 2023) (imagen 10). Y finalmente, dentro del antiprograma se encuentra el

problema del derecho al acceso a la información sobre cómo funciona el modelo, con qué datos fue entrenado, y qué hace con los datos que recopila los usuarios.



(Imagen 10)

Ante esto, en México y Japón se han incorporado legislaciones existentes a modo de actantes para generar antiprogramas para hacerle frente al programa y antiprograma de los modelos de Alphabet Google. Por ejemplo, en México se ha aprobado la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP), que regula el tratamiento de los datos personales por parte de las personas físicas o morales de carácter privado (Secretaría de Economía, s.f.). En Japón se ha reformado la Ley de Protección de Información Personal (APPI), que establece los principios y las obligaciones para el manejo de los datos personales por parte de las entidades públicas y privadas (Secretaría de Economía, s.f.). Estas legislaciones buscan garantizar el derecho a la privacidad y el acceso a la información de los usuarios frente a la recolección de datos por parte de Alphabet Google.

Desde la perspectiva de la sociología pragmática de la crítica de Boltanski, se podría analizar cómo los actores humanos involucrados en el uso y el desarrollo

de la IA en Alphabet Google recurren a diferentes principios o mundos para evaluar y resolver los conflictos que surgen en torno a los derechos humanos. De esta manera se podría estudiar cómo los actores humanos que participan en Alphabet Google apelan a distintos principios de justicia para defender o cuestionar la recolección de datos o la generación de sesgos por parte de la IA, y cómo negocian o confrontan sus puntos de vista con otros actores que tienen principios diferentes o contrarios como han sido los casos en México y Japón ante la resistencia de los programas y antiprogramas de Alphabet Google al incorporar actantes legislativos para frenar las transgresiones de sus derechos humanos.

Ahora bien, Alphabet-Google opera principalmente en el ámbito del mercado y la industria, buscando aumentar su margen de ganancias y ofrecer soluciones basadas en IA que agilicen los procesos industriales. Sin embargo, la compañía ha desarrollado, al menos de manera tentativa, principios y prácticas de contingencia sobre el uso, desarrollo e implementación de sus modelos y sistemas de IA. Estos principios se dividen en tres áreas principales: regulación, desarrollo de IA y normativas sobre el uso de la IA generativa.

Los principios de Alphabet-Google para una regulación responsable: Estos principios incluyen el fomento de la innovación responsable, la garantía de que los resultados de la IA sean positivos y minimicen los riesgos, y la identificación de lagunas legales que deben abordarse en relación con las aplicaciones de la IA. También se destaca la importancia de empoderar las agencias regulatorias existentes, adoptar un modelo radial de experiencia en IA y alinearse con los estándares internacionales. Por su parte, la Política de Uso Prohibido de la IA Generativa de Google establece pautas éticas que restringen actividades peligrosas o ilegales, el abuso de la privacidad y propiedad intelectual, y la creación de contenido dañino, como discursos de odio o desinformación. Finalmente, Google reafirma su compromiso con el uso de la IA para avanzar en la investigación científica y resolver problemas globales, asegurando que se mantengan los más altos estándares de ética y seguridad (Walker, 2024; (Google, 2024b; Google, 2023).

Transgresiones de la IA de Google en Japón: Aunque no se han registrado violaciones directas de los sistemas de IA de Google en Japón, han surgido controversias tras la implementación de sus modelos avanzados. Un caso relevante es el señalado por la Asociación de Editores y Periódicos de Japón, que alertó sobre la posible infracción de derechos de autor por parte de los motores de búsqueda con IA de Google y Microsoft, argumentando que reutilizan contenido sin autorización, reducen el tráfico hacia los medios y ponen en riesgo el periodismo y la democracia, lo que ha llevado a exigir reformas en las leyes de propiedad intelectual (Kyodo News, 2024).

Además, los algoritmos de Google han generado sospechas en torno a posibles sesgos en Japón relacionados con prácticas anticompetitivas. En este contexto, la Comisión Federal de Comercio de Estados Unidos emitió una orden administrativa contra la empresa, acusándola de emplear tácticas para restringir la capacidad de Yahoo Japan de competir en el ámbito de los anuncios de búsqueda dirigidos (Dan-Awoh, 2024). Además, en temas de privacidad, acceso a la información y no discriminación, Alphabet-Google debe ser particularmente cuidadoso en la implementación de sus sistemas de IA, ya que la transparencia es un aspecto clave tanto para la sociedad japonesa como para sus regulaciones gubernamentales (The Yomiuri Shimbun, 2023; Ioku et al., 2024).

Transgresiones de la IA de Google en México: En México, el impacto de la IA de Google ha generado preocupaciones principalmente en torno a la privacidad y los sesgos algorítmicos en publicidad y búsqueda de información. Investigaciones han documentado casos donde los sistemas de anuncios dirigidos de Google han reforzado estereotipos de género y discriminación económica, lo que ha suscitado críticas sobre la falta de supervisión en el desarrollo de estos algoritmos (R3D, 2023). Asimismo, organismos de derechos digitales han señalado que la recopilación masiva de datos personales por parte de Google podría entrar en conflicto con la Ley Federal de Protección de Datos Personales, especialmente en sectores vulnerables donde la opacidad en el uso de la información representa un riesgo significativo (Arturo et al., 2024). A pesar de estas preocupaciones, el marco regulatorio mexicano sigue siendo más laxo en comparación con el

japonés, lo que permite a Google operar con menor presión para garantizar transparencia y equidad en sus sistemas de IA.

Políticas de Google en torno a la IA en México: Aunque Google ha enfrentado diversos problemas relacionados con la privacidad, el acceso a la información y la no discriminación en México, como se ha evidenciado en los casos mencionados en el apartado anterior, estos no están directamente vinculados con sus sistemas de inteligencia artificial. Sin embargo, pueden ilustrar posibles conflictos y prácticas futuras. En cuanto al uso y desarrollo de la IA en México, Alphabet-Google ha implementado diversas herramientas y políticas internas, aunque no ha establecido principios éticos ni normativas locales sobre derechos humanos. Entre estas iniciativas se incluyen: el apoyo al progreso económico, las habilidades digitales para la era de la IA a través de modelos como Gemini Education y Generative AI for Educators, y el uso de la IA para abordar desafíos globales, como el apoyo al Instituto Internacional de Manejo del Agua, con una donación de 1.5 millones de dólares para desarrollar una plataforma de gestión hídrica sostenible impulsada por IA (Coulter, 2024). Además, la empresa ha creado herramientas que promueven la privacidad, permitiendo a los usuarios eliminar información importante, como su nombre, de su motor de búsqueda mediante plataformas como “About You” (Google, 2024).

5. Uber y los grandes problemas de la IA en México

Uber es una empresa cuyo modelo de negocio se basa en la desregulación (Morton, 2020). Esto significa que busca transgredir los códigos regulatorios nacionales y locales mediante un modelo de crecimiento que la haga demasiado grande para fracasar. De esta forma, la empresa comienza siempre una campaña de expansión y crecimiento sin tener en cuenta los códigos jurídicos de las regiones en las que se inserta, lo que genera muchos problemas sociales a través de sus sistemas con IA.

Los sistemas de IA desempeñan un papel clave en la operación y funcionalidad de Uber, estructurando su red de relaciones con trabajadores y usuarios (Climent et al., 2024). En el centro de esta infraestructura se encuentra Michelangelo, una

plataforma de aprendizaje automático (ML) que optimiza la creación y gestión de modelos a gran escala (Wang et al., 2024). Con la incorporación de modelos de aprendizaje profundo (DL) y arquitecturas avanzadas como redes neuronales convolucionales (CNNs), Uber ha expandido sus capacidades en IA, facilitando proyectos de alto impacto desde 2019 (Purwono et al., 2023).

Desde 2023, Uber ha integrado sistemas de IA generativa, incluyendo modelos de lenguaje, para mejorar la experiencia del usuario y la productividad interna (Wang et al., 2024). Además, la plataforma utiliza IA para detectar fraudes, gestionar precios dinámicos y optimizar la asignación de conductores en función de la demanda (Ghandi, 2024). Desde la perspectiva de la Teoría del Actor-Red (ANT), estos sistemas actúan como mediadores en la red de Uber, conectando clientes y conductores, gestionando pagos y ajustando precios, lo que refleja la influencia de la IA en la dinámica del servicio (Morton, 2024).

Sin embargo, a pesar de que en la última mitad de 2022 la empresa finalmente reportó ganancias después de un período de crecimiento exponencial sin importar las pérdidas que generó, y de los enormes problemas comerciales causados por un objeto que literalmente paró el mundo y gran parte del negocio de taxis bajo demanda de la empresa: el SARS-CoV-2, que causó la pandemia Covid-19 (2019-actualidad). Aun así, la empresa sigue siendo uno de los casos más agresivos y visibles de los peligros de los sistemas y plataformas con IA. Antes de examinar los casos a nivel local en México, es necesario tener una visión general de qué es Uber, cómo funciona y cuáles han sido algunas de sus mayores controversias.

Uber es una compañía que ha sufrido muchas controversias por su modelo de negocios y su forma en que ha operado en el pasado. Su servicio de taxi a demanda y su servicio de entrega de comida, UberEats, son dos de sus servicios más populares y visibles. Además, la empresa ha intentado incursionar en otros campos, como los vehículos autónomos, sin embargo, este esfuerzo no ha tenido éxito.

La empresa ha enfrentado problemas ligados a su modelo de negocios, como los revelados en los llamados *Uber Files*, documentos publicados por el diario británico *The Guardian* que muestran algunas de las prácticas de la empresa con reguladores y líderes gubernamentales. A pesar de su expansión agresiva, Uber ha tenido dificultades para penetrar en mercados como el japonés debido a las prácticas regulatorias y los gremios locales. En resumen, Uber es un importante caso de estudio en torno a los problemas de la IA en México debido a su gran penetración en el país y a los problemas sociales que ha causado a nivel global.

A. Las controversias de Uber y Uber eats a nivel global

Las plataformas digitales de trabajo hacen un uso controvertido de la IA. Esto es de especial relevancia en torno a las plataformas digitales de trabajo o de la economía gig que emplean repartidores para entregar comida. A simple vista, esta actividad parece tener grandes beneficios para las personas que hacen las entregas, al prometerles un trabajo no supervisado que complementa sus ingresos; sin embargo, este tipo de actividad es menos no supervisada de lo que parece.

Al considerar los tres elementos de la IA (datos, algoritmos y capacidad de cómputo), se observa que los repartidores alimentan constantemente a la plataforma con datos sobre su localización, horas activas y otros datos estructurados como edad y género. Pero más importante aún, estos datos junto con los que se recolectan de los consumidores sirven para que la plataforma planee rutas de entrega, presione sobre las horas de trabajo y determine los ingresos de los repartidores. En resumen, el algoritmo se convierte en el supervisor predeterminado de los trabajadores: un supervisor que no es empático no tiene capacidad reflexiva y que solo se enfoca en optimizar un resultado al transgredir bajo un vago consentimiento de las personas, el derecho a su información privada.

La IA también genera problemas en torno a la responsabilidad en Uber, ya que, si un algoritmo provoca un accidente de un repartidor al hacer que este trabaje demasiadas horas o al sugerir una ruta inadecuada, la culpa no será del algoritmo

ni de la empresa. El repartidor tendrá que hacerse cargo de los daños causados por un tercer actor que, debido a problemas regulatorios y epistémicos, no asume responsabilidad.

En este sentido, se tiene que examinar cómo funciona Uber. Como se mencionó al principio de este párrafo, el modelo de negocios de Uber se basa en la regulación y en ser demasiado grande como para fracasar. Esto quedó particularmente evidente en los documentos de Uber Files, publicados por el diario británico *The Guardian*, que mostraron la verdadera naturaleza de la empresa más allá de su imagen engañosa de simple intermediario entre trabajadores y clientes. De acuerdo con Davies et al., (2022) Más de 124.000 documentos confidenciales filtrados a *The Guardian*:

“El tesoro filtrado de archivos confidenciales ha revelado la historia interna de cómo el gigante tecnológico Uber se burló de las leyes, engañó a la policía, explotó la violencia contra los conductores y presionó en secreto a los gobiernos durante su agresiva expansión global.... El caché de archivos, que abarca de 2013 a 2017, incluye más de 83,000 correos electrónicos, *iMessages* y mensajes de WhatsApp, incluidas comunicaciones a menudo francas y sin adornos entre Kalanick y su equipo de ejecutivos principales...”.

Los Uber Files revelan información importante sobre el período de expansión agresiva de la empresa, que ocurrió antes de su crisis como modelo de negocios causada por la pandemia de Covid-19 y la crisis de imagen pública relacionada con los excesos y desplantes de Kalanick (ídem). Los documentos también muestran que la empresa utilizó estrategias para evadir los mecanismos regulatorios de los gobiernos e influir en ellos, o crecer tan rápido que resultara difícil implementar mecanismos regulatorios para abordar los problemas sociales que causaba. Uno de los mecanismos legales y conceptuales que permitió a la empresa eludir obligaciones legales fue calificar a sus trabajadores como agentes libres o trabajadores independientes, aunque los algoritmos y programas influyen constantemente en su comportamiento y distribución de ganancias y asignación de tarifas (Morton, 2020).

Entre las revelaciones en torno los procesos de desregulación, los autores del diario de *The Guardian* exponen:

“Los mensajes filtrados sugieren que los ejecutivos de Uber al mismo tiempo no se hacían ilusiones sobre la violación de la ley por parte de la compañía, con un ejecutivo bromeando sobre que se habían convertido en *piratas* y otro admitiendo: —Simplemente somos jodidamente ilegales—” (Davies et al., 2022).

Uno de los mecanismos que empleó la empresa para transgredir los aparatos regulatorios locales y nacionales fue el *lobbismo*. La práctica del *lobbismo* se puede resumir como el actor de convencer a instancias y agentes legisladores en aras de que estos cambien los códigos regulatorios en favor o en contra de un actor. Es decir, cambiar el contenido de las leyes o cuerpos regulatorios antes de que se presente un reclamo en su contra desde el mundo civil.

En este sentido, las filtraciones de los *Uber files* exponen como Kalanick y otros ejecutivos de Uber intentaron convencer a distintos actores legislativos para que crear o cambiar leyes en favor de Uber. Por ejemplo:

“Además de reunirse con Biden en Davos, los ejecutivos de Uber se reunieron cara a cara con Macron, el primer ministro irlandés, Enda Kenny, el primer ministro israelí, Benjamín Netanyahu, y George Osborne, el canciller del Reino Unido en ese momento. Una nota de la reunión describió a Osborne como un *firme defensor*” (ídem).

Otro de los mecanismos que empleaba Uber durante la gestión de Kalanick fue simplemente ignorar las leyes o mecanismos regulatorios locales. Esto está ligado al modelo de crecer hasta que la empresa sea demasiado grande para fallar: es mejor permitir que la empresa haga lo que quiera porque el beneficio y su influencia social la han hecho indispensable para la economía, el empleo, entre otros sectores. Con base a las más de 83,000 correos electrónicos, iMessages y mensajes de WhatsApp filtradas, Davies et al (2022) ilustran este punto al mencionar: “el personal se refirió al “estado diferente al legal” de Uber u otras formas de incumplimiento activo de las regulaciones en países como Turquía, Sudáfrica, España, República Checa, Suecia, Francia, Alemania y Rusia”.

Asimismo, los archivos exponen cómo Kalanick no dudaba en incitar la violencia en los conductores de sus plataformas para sofocar las manifestaciones en contra de los problemas sociales que estas ocasionan. Por ejemplo: “En medio de huelgas de taxis y disturbios en París, Kalanick ordenó a los ejecutivos franceses

tomar represalias alentando a los conductores de Uber a realizar una contraprotesta con desobediencia civil masiva” (ídem). Este tipo de revelaciones expone el nivel de pensamiento que tenía el entonces ejecutivo de la empresa bajo la falsa premisa erróneamente asignada a Maquiavelo en torno a vulgar —*e/ fin justifica los medios*.

Pero además los documentos señalan la forma en los mismos algoritmos de plataforma tenían motores para evadir las leyes locales. En las mismas plataformas se había implementado herramientas para apagar funciones que permitieran que fuera auditable; lo que viola tanto el principio de acceso la información, y hace que la empresa evite la responsabilidad en torno a sus transgresiones jurídicas y sociales. David et. al (2022) explican lo anterior de la siguiente manera:

“Uber desarrolló métodos sofisticados para frustrar la aplicación de la ley. Uno era conocido internamente en Uber como un “interruptor de apagado”. Cuando se allanó una oficina de Uber, los ejecutivos de la empresa enviaron frenéticamente instrucciones al personal de TI para cortar el acceso a los principales sistemas de datos de la empresa, evitando que las autoridades pudieran reunir las pruebas”.

Más allá de las controversias reveladas en los *Uber Files*, la empresa se ha generado otra serie de problemas concernientes a su imagen pública, en particular, en ciernes a la discriminación. Quizás es uno de los casos más emblemáticos de esto, y en donde el algoritmo de las plataformas de Uber estuvo directamente involucrado, fue en torno a la discriminación de personas transgenero en Estados Unidos. Y uno de los testimonios de las víctimas expone lo siguiente:

“...después de un descanso para almorzar en su primer día, Escobedo se quedó sin acceso a la aplicación; su documentación no había sido aprobada. Escobedo intentó al menos 20 veces volver a presentar los registros, que incluían una foto de su rostro, copias de su identificación y comprobante de seguro de automóvil, dijo. Cada vez, fueron negados” (Hussain, 2021).

Luego, al revisar sus documentos, Escobedo se dio cuenta: es un hombre transgénero, y su apariencia en la foto de su licencia de conducir anterior no coincide con las fotos actuales que envió, mostrando su bigote ralo y perilla” (Hussain, 2021).

Casos como el de arriba exponen un problema dentro de la interfaz de usuario y los algoritmos de la plataforma. La interfaz de usuario no contiene o facilita las condiciones para que una persona que transita hacia su género verdadero suba los datos que expongan su nueva condición: nuevo nombre, nueva fotografía, entre otros. El algoritmo de la plataforma levanta banderas para señalar que esta persona usuaria no muestra su verdadera identidad, una experiencia económica y psicológicamente violenta para quienes caminaron el largo trecho emocional para hacer su transición de género. En palabras de Hussain (2021), Uber en algunos casos la plataforma:

“...ha bloqueado a personas transgénero y no binarias de los trabajos de conductor y repartidor al tratar sus documentos como fraudulentos, suspender sus cuentas y no rectificar la situación, según entrevistas con conductores y documentación proporcionada por la Unión Estadounidense de Libertades Civiles del Sur de California”.

Muchas personas transgénero que han sido bloqueadas de la plataforma se han quedado sin acceso a esta hasta durante más de un año. Y como en muchos casos de sesgos algorítmicos, la discriminación de las personas transgénero en Uber tiene que ver con una combinación entre el diseño del algoritmo, su implementación, y algunos sesgos por parte de los humanos que los desarrollan. En este sentido, la justificación de que muchas personas transgénero hayan sido dadas de bajas al subir los documentos que evidencian la transición a su verdadero género tiene que ver con los algoritmos de identificación de documentos.

Esto es, porque los algoritmos de supervisión de identidad de las plataformas de Uber están ligados a la base de datos del gobierno. Y sí alguien sube documentos con información distinta a la de esos datos, los algoritmos de la plataforma los etiquetan como fraudulentos. En suma, si alguien sube documentos que confirmen que la persona ya se identifica con su nuevo género, y estos no coinciden con el género que tiene identificado el gobierno de Estados Unidos, entonces la plataforma dará de baja a esta persona usuaria.

“El portavoz de Uber, Zahid Arab, dijo que hacer coincidir las fotos de perfil con las identificaciones del gobierno es una medida de prevención del fraude que la

empresa lleva a cabo como parte de sus protocolos de seguridad. Arab dijo en un correo electrónico que: —en ocasiones, las solicitudes pueden desviarse y dar como resultado una experiencia lamentable para el cliente que estamos trabajando para abordar—” (idem).

En muchos casos, las personas que fueron discriminadas por la plataforma desistieron de intentar trabajar más en esta. Ahora bien, Uber acepta dos elementos importantes en este sentido; primero, que el proceso para agilizar los procesos de contratación de personal recae en los sistemas creados por una empresa llamada Checkr; segundo, que: “Checkr no determina los estándares utilizados para evaluar la elegibilidad para asociarse con Uber y no está involucrado en la decisión de la asociación” (Legal Uber, s.f, en Hussain, 2021).

Los problemas de discriminación en contra de las personas transgénero no son los únicos problemas de este tipo que ha enfrentado Uber. De acuerdo con una investigación realizada por The Guardian, muchos conductores reportaron haber sido víctimas de discriminación algorítmica por parte de la plataforma en torno a su tono de piel y grupo étnico. En este sentido, el diario reportó: “Uber enfrenta más reclamos de compensación por discriminación racial por parte de conductores que dicen que fueron despedidos falsamente debido al mal funcionamiento de la tecnología de reconocimiento facial” (Butler, 2021).

En síntesis, Uber es una plataforma con un claro modelo anti-regulación. Con base en este modelo, la empresa busca crecer sin importar los marcos regulatorios locales y hasta que sea lo suficientemente grande para que los efectos de cualquier marco regulatorio también logren afectar directamente a las poblaciones locales (debido a los empleos que genera y los servicios que presta), y este mismo crecimiento le permite acabar con mucha de su competencia. Además, se han reportado casos de discriminación algorítmica por parte de los sistemas de la plataforma que afecta los derechos humanos.

Aun así, hay explorar las políticas de privacidad, sobre todo después de las lecciones que la empresa aprendió en torno los escándalos y problemas de imagen pública debido a los excesos y estilo de su antiguo director. Así pues, en el siguiente apartado examinarán a grandes rasgos las condiciones de usuario de la

plataforma. Y después, a mayor profundidad, los apartados en estas condiciones en torno al caso mexicano.

B. Políticas de privacidad de Uber

Es importante conocer las políticas de privacidad y uso de datos de Uber. Esto es, porque estos elementos proporcionan una idea general y actual de cómo opera la plataforma, qué datos utiliza de los usuarios y, en general, qué hace con ellos. Para ello, examinaremos la política de privacidad de la empresa a nivel global y los apartados dedicados al marco regulatorio mexicano.

Desde el principio, la empresa deja claro que las personas que acepten sus condiciones de uso permitirán que Uber haga uso de sus datos. Además, Uber explica que es importante que las personas lean este aviso (aunque la mayoría de los usuarios no lo hace) para conocer los puntos clave de su política de privacidad. A continuación, la empresa comienza a explicar tanto los conceptos clave de usuario y socio como las prácticas en torno a los datos personales de la siguiente manera:

- “Usuarios: Personas que piden viajes o que se transportan, incluidos quienes reciben los servicios programados por un tercero.
- Socios de la App: personas que brindan servicios de transporte a los usuarios de manera individual o a través de empresas asociadas.
- Clientes: personas que piden o reciben comida u otros productos y servicios para llevar o con entrega, p. ej., a través de Uber Eats y Postmates, incluso aquellas que usan el pago como invitado.
- Socios repartidores: personas que prestan servicios de entrega a domicilio, p. ej., a través de Ubereats y Postmates” (Legal | Uber, s. f.)

¿Qué datos utiliza Uber? En el caso de México, la plataforma deja claro que recupera datos estructurados al momento de crear una cuenta, datos en torno a la ubicación y el uso de datos cuando se emplean sus plataformas, e información de otras fuentes como conductores y plataformas que emplean las API de Uber. Al aceptar las condiciones de usuario de Uber, las personas permiten que esta

plataforma acceda a información como edad, nombre, sexo y residencia en el perfil; así como la ubicación y otros datos extraídos de sus socios y otras plataformas integrantes de sus API.

¿Cómo usa estos datos? De nueva cuenta, para la situación de México, Uber explica que recopila estos datos para mejorar el funcionamiento y la calidad de su ecosistema de servicios. Pero, además, la empresa expresa que también usa los datos en torno los siguientes objetivos:

- Para mejorar la seguridad y la protección de los usuarios y los servicios
- Para soporte al cliente.
- Para comunicarnos con los usuarios o permitir que estos se comuniquen entre ellos. Para la investigación y el desarrollo.
- Para ofrecer promociones o concursos y brindar descuentos y noticias” (ídem).

Asimismo, Uber explica que la empresa no está obligada a —obtener el consentimiento de los usuarios— para hacer esas transferencias de datos. Esto es, porque ese tipo de transferencias están permitidas bajo la legislación mexicana de protección de datos personales. De acuerdo con Uber, estas transferencias también son necesarias para generar un vínculo legal con el usuario y con ello poder proporcionar sus productos y servicios, y porque está sujetas a los códigos regulatorios vigentes.

Hay que considerar que la plataforma también usa cookies. De acuerdo con la plataforma, estos son fragmentos de código o texto que las plataformas o empresas emplean para archivar el historial de navegación de los usuarios. En este sentido, Uber explica que usa las cookies para:

“...para recordar su navegador o dispositivo durante y entre las visitas al sitio web. También utilizamos otras tecnologías que pueden identificarlo a usted o los dispositivos que utiliza. Por ejemplo, las "etiquetas de píxel" (también llamadas balizas) son pequeños bloques de código instalados en (o llamados por) una página web, una aplicación o un anuncio” (Sic: ídem).

Otro de los apartados del aviso de privacidad es el de intercambio y divulgación de información. En este apartado, la plataforma explica de manera más detallada

cómo se utilizarán los datos recopilados tanto por usuarios como por los trabajadores que utilizan sus servicios, y más importante aún, con quién se pueden compartir. Por tanto, Uber expone lo siguiente:

“Con otros usuarios, para que Uber pueda proporcionar sus productos y servicios. Con otras personas, cuando usted lo solicite (por ejemplo, cuando usted comparte su Llegada estimada o divide la tarifa con un amigo) y con los socios comerciales de Uber (por ejemplo, si solicita un servicio a través de una asociación u oferta promocional realizada por un tercero).

Con el público en general, cuando envía contenido a un foro público.

Con el propietario de una cuenta de Uber que usted pueda usar.

Con otros usuarios, para que Uber pueda proporcionar sus productos y servicios.

Con otras personas, cuando usted lo solicite (por ejemplo, cuando usted comparte su Llegada estimada o divide la tarifa con un amigo) y con los socios comerciales de Uber (por ejemplo, si solicita un servicio a través de una asociación u oferta promocional realizada por un tercero).

Con el público en general, cuando envía contenido a un foro público.

Con el propietario de una cuenta de Uber que usted pueda usar” (ídem).

De esta manera se aprecia que Uber comparte de manera constante sus datos con actores dentro de su red y más allá de su red. Asimismo, plantea los distintos mecanismos en que los comparte y los divulga; por ejemplo, fotos de internet dentro del ecosistema de la plataforma. De esta manera se puede ver que las personas dejan de ser las dueñas exclusivas de los datos que suben y producen en la plataforma; aun así, la empresa permite que esta información se borre si los usuarios solicitan la cancelación del servicio.

Además, dentro de las condiciones de usuario y aviso de privacidad, también se plantean los derechos que tienen las personas usuarias bajo la protección de la legislación mexicana. Los derechos a los que tienen los usuarios mexicanos de acuerdo con sitio de la plataforma Uber son:

(a) el derecho a acceder a su información personal y solicitar explicaciones sobre cómo Uber usa esta información, (b) el derecho a rectificarla si cree que está desfasado, incorrecto o incompleta, (c) el derecho a cancelar sus datos si cree que no se usan según los principios, deberes y obligaciones establecidos en las leyes mexicanas con respecto a la protección de datos, y (d) el derecho a oponerse al procesamiento de sus datos por parte de Uber para determinados propósitos.

En este caso, se respeta el derecho humano de acceso a la información, el derecho a instar procesos de eliminación de datos e incluso un contrapeso frente al uso de datos por parte de las plataformas de Uber. Pero como bien lo explica Boltanski, y como plantea el aviso de privacidad de la empresa, en el caso del punto (d) con respecto a oponerse al procesamiento de datos se requiere de una autoridad –encarnada o no en un objeto– que contenga los códigos para justificar ese reclamo. En este sentido, Uber sostiene que estos derechos provienen de los derechos ARCO, y de acuerdo con la Secretaría de la Función Pública (s.f): “como titular de sus datos personales o a través de su representante, tiene derecho a acceder a ellos, a rectificarlos, a solicitar su cancelación u oponerse a su tratamiento”.

Los derechos ARCO a los que está sujeto el aviso de privacidad de Uber son: derecho de acceso, derecho a la rectificación de datos; derecho a la cancelación, lo cual implica la cancelación del uso de los datos o la eliminación de estos (con algunas excepciones como "cuando sean necesarios para el cumplimiento de las atribuciones de la SFP y de obligaciones legales"; y derecho a la oposición, o solicitar a la Secretaría de la Función Pública que no use los datos para ciertos fines salvo para casos legales como excepción (ídem)". Y en este sentido, al menos en papel, la empresa articula su aviso de privacidad y condiciones de usuario en torno a la legislación local. Esto contrasta mucho con el modelo de negocio de regulador que promovía Kalanick.

Finalmente, Uber plantea que es posible que las condiciones de servicio, así como el aviso de privacidad, cambien con el tiempo. Puede ser por cambios en los modelos tecnológicos, avances tecnológicos, cambios en el modelo de negocio de la empresa y cambios en las legislaciones. Además, la plataforma deja claro que: “En la medida de lo permitido por las leyes aplicables, usted acepta las actualizaciones al usar nuestros servicios luego de recibir esta notificación” (Legal Uber, s.f).

En resumen, el aviso de privacidad y las condiciones de usuario de Uber explican cómo, cómo y para qué se utilizan los datos. Además, presentan apartados

relacionados con el sistema regulatorio mexicano basado en los derechos humanos; en particular, los derechos en materia digital conocidos como ARCO. Esto significa que, al menos en esta etapa de la empresa, se respetan los derechos de privacidad y acceso a la información, aunque plantea problemas relacionados con los sistemas de IA como la explicabilidad.

Además, en el aviso de privacidad no se menciona el compromiso de evitar prácticas discriminatorias desde una perspectiva social (con daño a los seres humanos). Así, estos avisos explican claramente el mecanismo constante de recopilación de datos de la plataforma (aunque, de nuevo, esto no explica cómo funciona). Esto al mencionar:

“...podemos recopilar información de contacto de las personas que usan las cuentas de clientes de Uber para Empresas o de propietarios o empleados de restaurantes u otros comercios, datos personales de aquellos que inician solicitudes para ser socios conductores o socios repartidores (aunque no las completen) u otra información personal relacionada con nuestras funciones y tecnología de mapeo” (idem).

C. Uber en México

Tan sólo en México hay más de 200 mil conductores activos de Uber hasta el año de 2021 cuando la plataforma reveló sus últimas cifras (Comunicación Uber, 2021). Por ende, Uber es un caso de estudio importante para México en torno al impacto de la IA con respecto a los derechos a la privacidad, acceso a la información y derecho a la no discriminación, además de los problemas de responsabilidad y entedibilidad. A pesar de que, al momento de la redacción de este trabajo, atraviesa una gran crisis de imagen pública y financiera, se ha consolidado en la sociedad y el mercado mexicano, especialmente en relación con sus servicios de entrega de alimentos y taxis bajo demanda. Uber ha logrado establecerse en regiones importantes como Cancún y la Ciudad de México.

Es necesario preguntarse si la plataforma transgrede los derechos de privacidad, acceso a la información y no discriminación en México, además de otros problemas sociales relacionados con su funcionamiento, como la explicabilidad, la rendición de cuentas y la responsabilidad. Para ello, se examinará brevemente

cómo la plataforma afecta a sus trabajadores, usuarios y a los problemas sociales que ocasiona en el país. Todo en marco que examine los problemas de Uber en sus principales plataformas: Uber y Uber Eats.

Pero primero hay que explicar qué tan grande es realmente Uber en México. En este sentido, de acuerdo con el sitio de Uber, hasta 2021 la plataforma contaba con alrededor ocho millones de usuarios activos —aquellos que usan de manera recurrente las plataformas— (Uber celebra 8 años de generar oportunidades a través del movimiento y la tecnología en México, 2021). Asimismo, hay que retomar brevemente la historia de la plataforma en el país:

“El primer viaje en Uber en la Ciudad de México, entonces Distrito Federal, fue realizado en junio 2013 por —...*Daniela Schmidt*...— quien —...tuvo la suerte de llegar a la inauguración de FashionTV en una fabulosa Escalade negra— (Uber Blog, 2020). La plataforma ofrecía una experiencia más cómoda anclada a promociones de cien pesos o cuatro dólares para que las personas usaran el servicio. Uber promocionaba a la CDMX como un lugar con un panorama bello, con zonas congestionadas y con áreas atractivas en donde conviven personas con interés culturales y culinarios amplios” (Morton, 2020).

Como es el caso de otros países, la estrategia de mercado de Uber en México fue la desregulación. Por ejemplo, una de las estrategias que empleó para evadir los marcos regulatorios existentes fue presentarse como una empresa de transporte y no de tecnología (Hernández y Galindo, 2016). Pero más adelante, se determinó que, en la Ciudad de México, la plataforma operaba bajo un modelo similar al de los taxis bajo demanda: “Por ende, las PDT se sujetaron a la Ley De Movilidad aprobada en 2014 y tuvieron que cumplir con sus todos sus estatutos destacando —...la definición de taxi como vehículo destinado al servicio de transporte público individual de pasajeros— (Ávalos y Sofía, 2015)” (Morton, 2020).

Más adelante, Uber lanzaría una serie de productos y servicios en México, destacando su plataforma de entrega de comida llamada Ubereats. Algunas preguntas que se pueden hacer sobre cómo funcionan estas plataformas de Uber son: ¿cómo operan las plataformas de Uber? Tanto Uber como Ubereats funcionan a través de plataformas digitales de trabajo en las que un cliente solicita un servicio a un trabajador (que la empresa llama socio).

En principio, las plataformas de Uber actúan como intermediarias. La plataforma actúa como intermediaria, también interviene y afecta las relaciones y cómo las personas se relacionan, cuánto ganan, qué rutas tomar y en qué momentos es mejor trabajar, entre otros aspectos. Por lo tanto, Uber evita la responsabilidad de ser tanto un empleador como una empresa, bajo el pretexto de proporcionar un punto medio en el que los actores humanos pueden interactuar, y de esta manera:

“Uber — y otras plataformas como Didi—, en la Ciudad de México, operan con sus propios códigos que son ejecutados por diferentes actores humanos y no humanos. Mientras que, en un modelo de taxi pirata como las Panteras negras, se establece una tarifa a partir de una maquina o lo que considere el gremio o el conductor, la tarifa en PDT es ejecutada a partir de las decisiones gestadas por su IA con base en un algoritmo” (ídem).

Se puede apreciar que los sistemas con IA juegan un rol importante en el funcionamiento de Uber. Y esto es el humo y los espejos detrás de la ilusión de agencia libre en torno al rol de trabajadores que ocupan esta. Además, como se ha visto a lo largo de la tesis, el uso de estos algoritmos trae consigo el riesgo de problemas en torno a la transgresión de los derechos humanos: ¿Pero ¿cómo se da esto en México, y cómo se le hace frente y se puede hacer frente?

Los Uber Files revelan información importante sobre el período de expansión agresiva de la empresa, que ocurrió antes de su crisis como modelo de negocios causada por la pandemia de Covid-19 y la crisis de imagen pública relacionada con los excesos y desplantes de Kalanick. Los documentos también muestran que la empresa utilizó estrategias para evadir los mecanismos regulatorios de los gobiernos e influir en ellos, o crecer tan rápido que resultara difícil implementar mecanismos regulatorios para abordar los problemas sociales que causaba. Uno de los mecanismos legales y conceptuales que permitió a la empresa eludir muchas obligaciones legales fue calificar a sus trabajadores como agentes libres o trabajadores independientes, a pesar de que los algoritmos y programas constantemente influían en su comportamiento y distribución de ganancias y asignación de tarifas (Morton, 2020).

García-Tejeda (2021) explica que hasta el momento la regulación en torno a las plataformas se ha centrado en cobrar una especie de impuesto de dos por ciento,

que al final es asimilado tanto por usuarios o trabajadores. En este sentido, el autor explica que, por ejemplo: “A partir de junio de 2020, los usuarios de Netflix en México pagarán el impuesto al valor agregado (IVA) por los servicios remotos de películas y series que contratan”. Además, menciona:

“Desde 2015, los gobiernos estatales han regulado los servicios de transporte de pasajeros de empresas como Uber, y en la Ciudad de México existe una contribución de 1.5%, que cobra a los usuarios por cada viaje realizado a través de esta empresa de red de transporte” (ídem).

Este tipo de regulaciones son de carácter tributario. A pesar de que existan elementos como la ley de protección de datos personales, estos quedan cortos frente a la cada vez más compleja naturaleza de las herramientas con IA que implementa Uber. Así pues, determinar si Uber transgrede la privacidad de sus conductores es complejo porque no se puede determinar exactamente cómo, y si esta plataforma toma una decisión que afecte abiertamente a los usuarios, no existen los mecanismos para determinar al responsable.

Tal como lo plantean Boltanski y Thevenot, no es que los actores sean agentes pasivos sujetos a las condiciones de un campo. Si se presentan las circunstancias estos pueden presentar una inconformidad y manifestarse contra un actor que les hizo algún daño. Esto se vio claramente durante el Quinto Paro Internacional de Repartidores en noviembre de 2021.

Durante este paro, el mayor reclamo fue la discriminación (Laureles, 2021). Asimismo, muchas mujeres manifestaron sufrir de violencia sexual durante sus jornadas de trabajo. Pero si bien estos problemas son de suma importancia en lo laboral, es importante conocer cómo la IA en las plataformas como Ubereats impactan a los trabajadores de manera directa o indirecta.

En el caso de la plataforma de taxis bajo demanda. Si bien hubo una serie de manifestaciones y cabildeos en torno al reclamo de la participación de Uber en el mercado, y de sus intentos de regulación, la plataforma logró consolidarse en el país. Y esto contrasta con lo que pasó en Japón, y que a pesar de que ambos gremios ondeaban banderas de protesta similares —en particular con respecto al discurso de que los conductores de Uber y la plataforma eran taxis

piratas disfrazados— el gremio de taxistas mexicano no logró contener el avance de la participación de Uber en el mercado, salvo en espacios contados como algunos aeropuertos y centrales de autobuses.

“...el taxista Palafox en su entrevista con Mendoza (2015) comentó otro motivo por el que decidieron manifestarse: “Que se combata la piratería, como en el caso de los irregulares, Uber y los Panteras”. Con base en las entrevistas, se puede ver que gran parte de lo que ocasionó la movilización de los taxistas fue la entrada de una nueva serie de actores que no seguían los códigos normativos de la comunidad encarnada en las leyes de movilidad de la Ciudad de México. Ahora bien, a diferencia de otros actores irregulares, Uber contaba con la fachada de ser una empresa de tecnología más no una empresa de transporte por lo cual no se debía adaptar a los mismos códigos normativos que regían a los taxistas regulares” (Mendoza 2015 en Morton, 2021).

Los algoritmos en Uber dirigen a los usuarios. Y en el caso de México esto se traduce en cambiar al jefe de trabajo por la supervisión, así como toma de decisiones constantes de estas herramientas que en muchos casos tienen IA. No obstante, y pesar de algunos casos que, en torno a problemas de discriminación dentro de las plataformas de Uber en México, es necesario conocer si hay más elementos en torno a casos donde la plataforma transgredió los derechos de privacidad, acceso a la información y derecho a la discriminación, además de cómo responde a los problemas de responsabilidad y entedibilidad.

Asimismo, los obtener ingresos suficientes, los conductores suelen trabajar entre 10 y 12 horas diarias, superando la jornada laboral permitida en el país, que es de máximo 48 horas semanales (Flores, 2024). Históricamente, los conductores han carecido de acceso a prestaciones como seguro médico, pensiones o licencias por maternidad. Sin embargo, en diciembre de 2024, el Senado mexicano aprobó reformas para mejorar las condiciones laborales de los conductores de aplicaciones, garantizándoles acceso a seguridad social y aguinaldo. Estas reformas, propuestas por la presidenta Claudia Sheinbaum, buscan reconocer derechos laborales básicos para los trabajadores de plataformas digitales (Reuters Staff, 2024; Presidencia de la Republica n.d).

Más allá de los beneficios legales y regulatorios obtenidos por los conductores de Uber, persisten problemas estructurales derivados de los sistemas de IA que

utiliza la plataforma. Como mencionamos anteriormente, Uber emplea modelos basados en aprendizaje automático y redes neuronales convolucionales (CNNs), los cuales funcionan como una forma de software capataz (bossware) que supervisa y condiciona las decisiones de los conductores (De los Santos et al., 2023). Esto es evidente en prácticas como la tarificación dinámica (surge pricing), un mecanismo central de la IA de Uber que reduce los tiempos de espera de los pasajeros e incentiva a los conductores a operar en zonas de alta demanda (Ghandi, 2024).

Ahora bien, en el caso del derecho a la privacidad de Uber en México, por lo menos en torno al derecho al olvido, la plataforma en su política de privacidad establece los mecanismos en los que los usuarios le pueden pedir a la empresa que borre sus datos. De acuerdo con Uber en México, el usuario:

Puede solicitar que no se use información personal para los propósitos secundarios aquí. Sírvase indicar “Derechos del usuario de México” al principio de su solicitud. Le informamos que puede solicitar no recibir correos electrónicos promocionales de Uber aquí. También puede optar por no recibir correos electrónicos y otros mensajes de Uber mediante las instrucciones proporcionadas en ellos (Uber México, s.f).

Uber México es un ejemplo de una plataforma que resquebraja Derechos de privacidad, acceso a la información y derecho a la discriminación, además de los problemas de responsabilidad y entedibilidad. Particularmente, es posible que esta plataforma no logre respetar completamente la privacidad de los datos de sus trabajadores ni de sus usuarios, aunque investigar si esto es cierto queda pendiente. Y finalmente, los objetos regulatorios en torno a la responsabilidad de los incidentes ligados a los algoritmos de Ubereats no resultan ser eficientes para resolver controversias.

D. Uber en Japón

Es necesario examinar el caso de Uber en Japón para investigar cómo se desarrolla la plataforma en ese país y cómo han reaccionado los actores sociales ante los problemas relacionados con sus herramientas de IA. Además, este

análisis nos ayudará a entender qué tanto influyen los rasgos sociales nacionales en el modelo de regulador que la empresa tenía bajo su antiguo director.

Uber en Japón comenzó con un proyecto impulsado por un grupo de inversionistas y por el antiguo ministro Shinzo Abe. Al principio, la plataforma buscó evadir los modelos del país al presentarse como una empresa de interés social capaz de brindar transporte a las comunidades más alejadas o poco conectadas. Sin embargo, a pesar del lobbying que algunos sectores hicieron a su favor, la empresa rápidamente encontró resistencia y su tecnología finalmente fue cooptada por las empresas locales de taxis cuyo modelo de negocio se transformó debido a la incursión de Uber.

En última instancia, Uber no tuvo éxito en Japón debido a la fuerte respuesta de distintos actores que se manifestaron en contra de la intención desreguladora de la empresa, que estaba respaldada por las tendencias en esa dirección del ministro Shinzo Abe. Esto impidió que las plataformas de Uber y sus sistemas de IA se desplegaran ampliamente en el país, lo que contuvo los efectos negativos de estos sistemas, herramientas y algoritmos.

En un momento, Uber tenía el viento a su favor. Contaba con el apoyo de importantes actores como el ministro Abe y sus partidarios en el parlamento, que buscaban liberalizar mercados tradicionales, así como el interés de líderes empresariales de cooperativas como Rakuten y Softbank en reformas que legalizaran servicios como los de Uber. Además, Uber se presentó de manera atractiva con un discurso de intereses sociales y no comerciales, en especial porque mencionaron correctamente que las empresas de taxis tradicionales descuidaban las zonas rurales y alejadas. Sin embargo, es importante conocer por qué Uber fracasó en Japón y cómo lo hizo.

Según Altura et al. (2020), la explicación de este comportamiento de resistencia de los actores ante la implementación de nuevas herramientas está ligada a los contextos históricos. Es decir:

“Los actores operan en contextos históricos que les otorgan poder ex ante de coaliciones establecidas. El famoso *triángulo de hierro* en Japón —vínculos entre

los zoku (grupos de política dietética), los burócratas ministeriales y las asociaciones industriales— proporciona a los actores estructuras organizativas de apoyo que protegen sus intereses”.

Este contexto histórico y social también puede explicar cómo Uber rápidamente enfrentó resistencia. En particular, la plataforma chocó con una fuerte oposición de grupos de empresas de organizaciones de taxistas como el *Zentakuren*, que se manifestaron en contra de lo que consideraban taxis piratas (*shirotaku*) que violaban las normativas de seguridad establecidas en un mercado regulado. Además, muchas empresas y asociaciones contaban con conductores que tenían casi todas las prestaciones y beneficios de un empleo formal, lo que intensificó el rechazo hacia Uber. Por eso: “El encuadre de Uber como 'shirotaku que destruye la seguridad' ayudó a movilizar la resistencia y ofreció una poderosa contranarrativa” (ídem)

Pero esto no significa que los gremios, empresas y organizaciones de taxistas no estuvieran preocupadas por las ventajas que Uber tenía. Aunque la empresa buscaba desregular el mercado al hacer alianzas con diferentes actores y empresarios importantes que promovían la liberación del mercado (como fue el caso de Son, fundador de la importante empresa Softbank), también eran conscientes de que esta empresa y plataformas similares ofrecían procesos de innovación importantes. Según Altura et al. (ídem), se instruyó a los miembros del *Zentakuren*: “De ahora en adelante, si reciben solicitudes de cooperación de áreas despobladas o rurales, quiero que brinden asesoramiento de inmediato, sin excepción”.

Más allá de la estrategia de atender áreas despobladas, como se mencionó arriba, los corporativos de taxis sabían perfectamente que Uber había propuesto elementos de innovación importantes. Así pues, después de una larga jornada de protestas y de cabildos con ministros y miembros de la dieta parlamentaria, Uber quedó relegado a un papel secundario dentro de la industria de taxis bajo demanda japonesa al tener que funcionar en alianza con alianzas con las organizaciones taxistas, pero sus procesos de innovación no pasaron desapercibidos; de esta manera:

“La estrategia de adoptar las “partes buenas” de Uber dio sus frutos con el desarrollo de plataformas tecnológicas estilo Uber para la industria del taxi. Nihon Kotsu creó su propia aplicación, JapanTaxi, y formó una alianza con Toyota, que invirtió \$69 millones. JapanTaxi incluye la mayor cantidad de taxis, más de 70,000, de cualquier aplicación en Japón. Y en 2019, JapanTaxi se asoció con la empresa de viajes compartidos respaldada por Softbank, Grab, con sede en el sudeste asiático. Sony también entró en la refriega, asociándose con otras seis compañías de taxis para desarrollar su propia plataforma de taxis”.

El análisis sociológico muestra cómo un nuevo actor (Uber y su tecnología) creó movimiento entre actores más tradicionales (gremios y organizaciones de taxistas japoneses) y provocó una reacción. Mientras que Uber, como empresa, trató de impulsar una acción entre miembros del parlamento, ministros y el primer ministro para liberalizar la industria de taxis. Al final, ganó el movimiento de los actores encarnados en los gremios y organizaciones de taxistas japoneses. Sin embargo, aunque Uber no logró desregular el mercado a su favor, sus procesos de innovación no pasaron desapercibidos. Los gremios y organizaciones de taxistas adoptaron algunas de estas herramientas tecnológicas y modificaron sus códigos de conducta.

Desde la perspectiva de Boltanski, hubo un proceso de disputa en torno a un mundo civil. La intención de Uber de desregular el mercado de taxis enfrentó la oposición de los gremios y organizaciones de taxistas japoneses, que se apoyaron en la autoridad de la regulación existente y se manifestaron en contra de Uber como "taxis piratas". Esto desencadenó una disputa entre actores y objetos con autoridad, que terminó desarticulando la operación de Uber en Japón. A diferencia de lo que sucedió en otros países, los actores sociales se manifestaron con fuerza y supieron usar los objetos y actores de autoridad a su favor. El gremio de taxistas ejerció suficiente presión para mantener los códigos en torno a objetos de autoridad que estaban de su lado, y a pesar de las presiones de Uber, ni siquiera el ministro de trabajo se atrevió a cambiarlos.

Dado el corto periodo de vida que Uber vivió en Japón, no se presentaron muchos de los efectos adversos en torno a la empresa y sus herramientas de IA. Es decir, no hubo tiempo suficiente para que la plataforma alterara la estructura social ni para que se manifestaran problemas en torno a sus algoritmos y herramientas de

IA, como la transgresión de los derechos a la privacidad, acceso a la información y derecho a la no discriminación, además de problemas de responsabilidad y explicabilidad. Como sugirió Davis (2016: 139), “la tecnología no es el destino: la misma tecnología se implementará de diferentes maneras dependiendo de las instituciones circundantes”.

E. Uber y teoría de actor-red y la teoría de las capacidades

Con base en el análisis de Uber en México y Japón se puede apreciar que el programa de Uber es el de la expansión bajo cualquier costo: ignorar la regulación, usar los algoritmos para maximizar sus ganancias, y trastocar gremios locales. Este programa genera el mismo antiprograma que se ve en otras plataformas que usan datos para el funcionamiento de sus algoritmos: privacidad, discriminación y opacidad (falta de acceso a la información). No obstante, en el caso de Japón, dentro de la red actantes humanos se resistió ante estos programas y antiprogramas mediante un antiprograma que básicamente desbarató a la plataforma de transporte de Uber para incorporar dentro de las distintas asociaciones de taxistas.

Pero en el caso de México, la resistencia ante el programa y el antiprograma de Uber ha sido menor y más dispersa. A pesar de las protestas y las demandas de los taxistas tradicionales, que acusan a Uber de competencia desleal y de evadir impuestos y regulaciones (Uber, s.f.), la plataforma ha logrado expandirse y consolidarse en más de 40 ciudades del país (Uber, s.f.). Asimismo, a pesar de las denuncias y las críticas de los usuarios y las organizaciones civiles, que señalan a Uber por violar la privacidad, discriminar o generar inseguridad (El Universal, s.f.), la plataforma ha mantenido su popularidad y su rentabilidad en el mercado mexicano (Marketing4Ecommerce, s.f.). Estos hechos sugieren que el antiprograma de Uber en México se basa en ofrecer un servicio accesible, eficiente y conveniente para los usuarios, que les hace tolerar o ignorar los problemas que plantea la IA de Uber.

Desde la perspectiva de la sociología de la traducción de Callon, se podría analizar cómo Uber logra imponer su programa de expansión bajo cualquier costo

mediante un proceso de traducción que implica cuatro momentos: la problematización, el interés, la enrolación y la movilización. Según Callon (1986), la problematización es el momento en que un actor define el problema y se presenta como una solución indispensable para los demás actores; el interés es el momento en que el actor trata de convencer a los demás actores de que su solución es la mejor y de que deben seguirlo, la enrolación es el momento en que el actor establece relaciones estables con los demás actores y les asigna roles y funciones, la movilización es el momento en que el actor habla y actúa en nombre de los demás actores y consolida su red. Así, se podría estudiar cómo Uber problematiza la movilidad urbana y se ofrece como una solución innovadora y eficiente para los usuarios, cómo interesa a los socios conductores y a las autoridades para que acepten sus condiciones y regulaciones, cómo enrola a los usuarios y a los socios conductores mediante su plataforma tecnológica y sus algoritmos, y cómo moviliza a sus aliados y defiende sus intereses frente a las críticas y las resistencias.

Ahora bien, al igual que en el caso de Alphabet-Google, los actantes humanos japoneses empujaron un antiprograma apelando a diferente justicia para resistir los programas y antiprogramas de Uber. Para ello, también hicieron uso de diferentes actantes no humanos para fortalecer su posición y su argumentación. Por ejemplo, los taxistas japoneses se apoyaron en la legislación vigente, en las asociaciones gremiales, en las tradiciones culturales y en la confianza de los clientes para defender su mundo doméstico frente al mundo del proyecto de Uber. Así, lograron limitar la expansión de la plataforma y obligarla a adaptarse a las condiciones locales.

En cuanto a los mediadores a favor a favor de sus sistemas de IA se encuentran aquellos que aprecian la conveniencia, la accesibilidad, la flexibilidad y la rentabilidad de esta plataforma de transporte. Pero en especial, sus voceros comerciales que promueven la implementación de estas tecnologías como un arma competitiva a sus accionistas. Dentro de estos mediadores resalta los directores ejecutivos, tecno-entusiastas, aliados que se benefician de la implementación de estas tecnologías más allá de sus transgresiones, medios de

comunicación que ganan audiencia al transmitir estas notas, y accionistas e investigadores que se interesan por el modelo de negocio y la tecnología de Uber, como la aplicación móvil, el algoritmo de precios, la conducción autónoma entre otras cosas.

Sin embargo, los mediadores en contra de sus sistemas de IA son aquellos que rechazan la precariedad, la inseguridad, la ilegalidad y la injusticia de esta plataforma de transporte. Estos mediadores pueden ser taxistas, sindicatos, gobiernos, jueces, usuarios, organizaciones de derechos laborales, y hasta reguladores y legisladores que se oponen o se ven afectados por el funcionamiento y los efectos de Uber. Y en el caso de México y Japón, estos voceros en contra se han manifestado en gremios de taxistas, organizaciones, e incluso clientes que básicamente han resistido a las transgresiones a nivel de derechos humanos, a nivel de asociaciones humanas en torno a la teoría de actor-red, a nivel técnico, y a nivel sociotécnico.

NIVEL	TRANSGRESIÓN	EJEMPLO
Derechos humanos	Violación de los derechos laborales, de la seguridad, de la privacidad y de la no discriminación de los conductores y los pasajeros.	Uber no reconoce a los conductores como empleados, sino como socios independientes, lo que implica que no les ofrece prestaciones sociales, seguridad social, protección legal, etc. Uber también ha sido acusado de poner en riesgo la seguridad de los pasajeros y los conductores al no verificar adecuadamente sus antecedentes, al no ofrecer seguros suficientes, al no responder por accidentes, robos, agresiones, etc. Uber también ha sido criticado por violar la privacidad de los usuarios al recopilar y compartir sus datos personales, de geolocalización, de viajes, etc. sin su consentimiento o con fines comerciales o de vigilancia. También se ha señalado a Uber de discriminar a usuarios y conductores por su género, raza, edad, discapacidad, etc. con algoritmos que determinan los precios, las rutas, las calificaciones, etc.
Asociaciones humanas	Alteración de las relaciones sociales, económicas, políticas y culturales entre los actores humanos y no	Uber ha generado conflictos y tensiones entre los actores humanos y no humanos que participan en el sector del transporte, como los taxistas, los conductores, los pasajeros, los gobiernos, los reguladores, los jueces, los medios de comunicación, las organizaciones sociales, entre otros. Estos actores tienen intereses, valores y expectativas

	humanos involucrados en el transporte.	diferentes o contrapuestos sobre el transporte, y se ven afectados o beneficiados por la presencia y la acción de Uber y sus sistemas de IA. Asimismo, la plataforma ha alterado el equilibrio y la estabilidad de las redes de transporte existentes, y ha creado nuevas redes de transporte alternativas o disruptivas; y en torno a la crueldad algorítmica que impacta asociación entre humanos, plataforma, y actores humanos, resalta: El uso de la desactivación, que consiste en bloquear o suspender a los conductores que reciben bajas calificaciones de los usuarios, lo que puede generar situaciones de injusticia, estrés o precariedad (Éticas Foundation, 2021). El uso de la asignación de viajes, que consiste en emparejar a los conductores con los usuarios según la distancia, el tiempo y la rentabilidad, lo que puede generar situaciones de manipulación, presión o riesgo (Aguar, 2023).
Técnico	Falla, error, vulnerabilidad o ataque de los sistemas de IA que afectan el funcionamiento, la calidad o la seguridad del transporte.	Uber ha enfrentado problemas técnicos con sus sistemas de IA, como la falla de la aplicación móvil, el error del algoritmo de precios, la vulnerabilidad de los datos, o el ataque de los hackers. Estos problemas técnicos han causado inconvenientes, pérdidas, daños o riesgos para los conductores, los pasajeros, la empresa o la sociedad. Uber ha tenido que resolver, corregir, mejorar o proteger sus sistemas de IA para evitar o mitigar estos problemas técnicos. Entre estos problemas técnicos están los relacionados con la discriminación.
Sociotécnico	Inadecuación, incompatibilidad, incompreensión o inaceptabilidad de los sistemas de IA con respecto al contexto, la cultura, la norma o el valor del transporte.	Uber ha tenido que adaptar, modificar, explicar o justificar sus sistemas de IA para que se ajusten, se integren, se entiendan o se acepten en los diferentes contextos, culturas, normas o valores del transporte. Uber ha tenido que lidiar con la resistencia, el rechazo, la crítica o la protesta de los actores que no comparten, no aprueban, no confían o no se benefician de sus sistemas de IA. Uber ha tenido que negociar, dialogar, colaborar o competir con los actores que tienen diferentes visiones, intereses, demandas o expectativas sobre el transporte y sus sistemas de IA.

¿Cómo se comparan las operaciones de Uber en México y Japón? En México, particularmente en la Ciudad de México, Uber logró establecerse gracias a una política desreguladora que enfrentó pocas resistencias efectivas, como la Ley de

Trabajadores de Plataforma, y que ha funcionado como un amortiguador en términos de oportunidades laborales (Morton, 2021). En contraste, en Japón, los gremios y asociaciones de taxistas lograron frenar la expansión del modelo de taxi bajo demanda, y aunque el país ha legalizado los viajes compartidos, lo hizo bajo estrictas regulaciones que impiden a Uber competir directamente con las compañías de taxis, ya afectadas por la escasez de conductores y el envejecimiento de su fuerza laboral. La resistencia a una desregulación total sigue siendo fuerte, incluso cuando los taxis autónomos ganan terreno a nivel global (Suzuki, 2024). Gran parte de estas diferencias en la adopción de Uber entre ambos países se debe a su distinta tolerancia al riesgo: mientras que en México la aversión al riesgo es menor, en Japón, una sociedad más reticente a la incertidumbre, Uber ha enfrentado mayores obstáculos (France 24, 2024).

En suma, Uber al ser una plataforma utiliza la IA para ofrecer un servicio de transporte urbano innovador y eficiente, pero que también genera problemas y conflictos relacionados con los derechos humanos de la IA, como la privacidad, la discriminación y la opacidad. Las perspectivas teóricas de la teoría del actor-red, la sociología de la traducción y la sociología pragmática de la crítica ayudaron a comprender cómo Uber logra imponer su programa de expansión bajo cualquier costo mediante un proceso de traducción que implica problematizar, interesar, enrolar y movilizar a los actores humanos y no humanos que participan en su red. Los actores humanos que se oponen al programa de Uber recurren a diferentes principios o mundos de justificación para resistir y criticar la recolección de datos o la generación de sesgos por parte de la IA de Uber; y en los casos de México y Japón se mostró cómo las condiciones sociales, culturales y legales influyen en la capacidad de resistencia y negociación de los actores humanos frente a Uber.

6. Sistemas metropolitanos de reconocimiento facial metropolitano

Una de las herramientas más controvertidas en el ámbito de la IA son los sistemas de reconocimiento facial. Estos sistemas gestionan datos biométricos que, una vez procesados, permiten la identificación de individuos (Contreras y Soto, 2023). En lo que respecta a derechos fundamentales como la libertad de expresión y el

derecho de reunión, estas tecnologías pueden tener un “efecto inhibitor” sobre las personas (idem). Además, propician transgresiones relacionadas con derechos como la privacidad, la no discriminación y el acceso a la información.

Los sistemas de reconocimiento facial han mejorado gracias a los avances en el procesamiento de la visión por computadora. Esto implica la implementación de técnicas que permiten a las máquinas aprender y reconocer patrones en el flujo de datos de imágenes, beneficiándose de los avances en la tecnología de cámaras de video. Además, las aplicaciones de visión artificial emplean inteligencia artificial y machine learning (IA/ML) para procesar estos datos con precisión, permitiendo la identificación de objetos y rostros, así como la clasificación, recomendación, supervisión y detección (Amazon Web Services, s.f; Mohsienuddin & Sabri, 2020).

Con esto en cuenta, es necesario a examinar algunos sistemas de vigilancia metropolitano con sistemas reconocimiento facial, en especial el C5, también conocido como Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la Ciudad de México. Es el sistema de videovigilancia más grande de la Ciudad de México y su adopción cada vez mayor, junto con el uso de tecnologías de reconocimiento facial y herramientas de IA para la prevención y control del delito, plantea implicaciones importantes alrededor de los derechos de privacidad, acceso a la información y derecho a la discriminación, además de los problemas de responsabilidad y explicabilidad. Según el sitio oficial del gobierno, hasta octubre de 2022 (Cdmx, s.f.):

El Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la CDMX (C5), es la dependencia del Gobierno de la Ciudad de México encargada de captar información integral para la toma de decisiones en materia de seguridad pública, urgencias médicas, medio ambiente, protección civil, movilidad y servicios a la comunidad en la capital del país a través del video monitoreo, de la captación de llamadas telefónicas y de aplicaciones informáticas de inteligencia, enfocadas a mejorar la calidad de vida de las y los capitalinos.

Como se mencionó en el libro *Weapons of Math Destruction*, la IA puede tener un gran impacto en sistemas como el C5. Los sesgos en estos sistemas pueden llevar a que muchas personas sean falsamente acusadas y encarceladas por delitos debido a los sesgos en los algoritmos y datos con los que fueron

entrenados. Parte de esto se debe a la confianza del público en la tecnología, que confía más en las predicciones de la IA que en las de una persona; y también tiene que ver con el desconocimiento de cómo funcionan estas tecnologías, especialmente en lo que respecta a los algoritmos, que, como señala Trejo et al. (2019), no siempre reflejan la realidad.

Los sistemas reconocimiento facial metropolitano están desplegados so pretexto de prevenir el crimen y asistir en la persecución del delito. De acuerdo con Wenger et al (2023), un sistema de vigilancia famoso en China emplea el reconocimiento facial para controlar el comportamiento de los civiles y aplicar el sistema de puntaje de crédito social. Asimismo, los autores plantean que, en Rusia, las autoridades compraron más de 100.000 cámaras en Moscú para construir un sistema de cuarentena de COVID basado en el reconocimiento facial; además de los casos de uso gubernamentales, los sistemas de reconocimiento facial se utilizan ahora para múltiples fines, como verificar la identidad de los viajeros en los aeropuertos y de los empleados al entrar en las oficinas corporativas (idem).

Wenger et al (2023), explican que los sistemas de reconocimiento facial se han convertido en parte de la vida cotidiana de muchas personas. Sin embargo, esta tecnología también plantea importantes desafíos éticos, legales y sociales, como la privacidad, la seguridad, la precisión y la equidad. En este sentido, los autores explican que básicamente cualquier institución puede rasgar imágenes o video de plataformas sociales para entrenar a sus algoritmos en torno a herramientas de reconocimiento facial; además de la gran cantidad de *data brokers* que contienen importantes bases de datos, pero de acuerdo a Wenger et al también existen otros informes han revelado cómo las fotos tomadas en lugares inesperados - aeropuertos, calles de la ciudad, edificios gubernamentales, escuelas, oficinas corporativas —pueden terminar en sistemas de reconocimiento facial sin que los sujetos lo sepan o lo autorice.

Además, los sistemas de reconocimiento facial transgreden la intimidad de las personas, ya que el rostro es el rasgo más distintivo y reconocible. Esto hace que la protección de datos personales sea de suma importancia (Domingo, 2022).

Aunque estas herramientas pueden ser útiles en situaciones susceptibles a la violencia, como en eventos deportivos, la creación de bases de datos vinculadas a las imágenes recopiladas por los sistemas y la identificación de personas a través del reconocimiento facial pueden erosionar y limitar los derechos de los individuos (idem). Además, surgen otros problemas asociados con estos sistemas, como la función de arrastre o *function creep*, en la que el uso inicialmente previsto de los datos se extiende a contextos no autorizados, incrementando así el riesgo de abusos y prácticas indebidas (idem).

Ahora bien, es muy fácil seducir con el discurso de seguridad en torno a los sistemas de reconocimiento facial. En 2023, el precandidato a la presidencia Marcelo Ebrard anunció que como propuesta de campaña implementaría un sistema de reconocimiento facial y de prevención del delito. De acuerdo con el exfuncionario público: “Un delincuente tiene una cierta forma de actuar y caminar, y podemos establecer patrones. Si tuviéramos esa tecnología, podríamos haber evitado muchos feminicidios, por el tipo de caminar y de comportarse” (Riquelme, 2023).

En cuanto a la adopción de sistemas de reconocimiento facial en la Ciudad de México, la jefa de Gobierno Claudia Sheinbaum anunció en 2020 que el sistema C5 estaría instalando 18,500 cámaras con esta tecnología para 2022. Estas cámaras se conectarán a 3,800 postes en la ciudad y también estarían disponibles para ser utilizadas por negocios y bancos. Esto significa un aumento significativo en la videovigilancia en la ciudad; no obstante, hay que conocer si todas estas cámaras cuentan con tecnología de reconocimiento facial o IA que el tema pertinente de esta tesis.

Sin embargo, la tecnología de reconocimiento facial plantea problemas éticos y de derechos humanos. Como menciona O'Neil (2016), las tecnologías relacionadas con el reconocimiento facial comenzaban a crear perfiles para prevenir la delincuencia que a menudo etiquetaban a las personas afroamericanas como posibles perpetradores. A pesar de los avances en la tecnología de aprendizaje automático y el reconocimiento facial y la prevención del delito, los sesgos en los

algoritmos aún están presentes. Además, es difícil evitar estos sesgos con modelos de aprendizaje automático entrenados con datos centrados en contextos sociales de países como Estados Unidos.

Además, responder a las quejas de sesgos algorítmicos que producen discriminación es complejo si no se tienen en cuenta las leyes locales. En otras palabras, el problema de la tecnología de reconocimiento facial y prevención del delito en sistemas como el C5 es que están diseñados en base a un modelo social y datos relacionados con una realidad particular, lo que produce sesgos específicos. Para aclarar la problemática jurídica, Muñoz (2020) explica: “Los conceptos de igualdad y discriminación tienen diferentes contenidos y alcances en cada sistema jurídico (Hagerty y Rubinov, 2019) y tienen particularidades propias en cada país, siendo en definitiva conceptos elaborados en el contexto social (Hellman y Moreau, 2013)”.

A Introducción a las problemáticas de los sistemas de reconocimiento facial

Ahora bien, aunque el alcance de esta tecnología en la Ciudad de México es limitado, aún está sujeta a los problemas sociales mencionados a lo largo de esta tesis. Por eso es necesario plantear mecanismos de respuesta tanto jurídicos como tecnológicos. Y, es necesario seguir abogando por fortalecer y reconocer los instrumentos de acceso a la información, como los del INAI, para mantener la transparencia en cuanto a cómo y dónde se implementan y operan estas tecnologías.

En contraste, el uso de sistemas de reconocimiento facial en Japón también ha sido escaso, aunque se ha utilizado en aeropuertos internacionales y para el acceso tentativo a estadios olímpicos. Incluso, la implementación de esta tecnología en líneas de transporte de pasajeros como JR East para los Juegos Olímpicos de 2020 ha generado controversia. Según Mainichi Japan:

“consentimiento para la adquisición de datos que describan sus rasgos faciales. Aun así, la Comisión de Protección de Datos Personales del gobierno, que supervisa la implementación de la ley, indicó en su sitio web que considera que

—debe haber notificaciones y anuncios sobre el objetivo del uso— en caso de que se utilicen cámaras de seguridad con capacidad de reconocimiento facial” (Do facial-recognition security cameras in Japan’s stations really keep us safe?, s. f.).

Para hacer frente concretamente a los problemas de discriminación y de sesgos en torno a los sistemas de videovigilancia desde una perspectiva técnica y de diseño, se realizó un ejercicio en el minicurso de Human Centered AI (IA centrada en el ser humano). Este ejercicio fue parte de la Escuela de Interacción Humano Computadora organizada por ACM SIGCHI Latinoamérica en 2022. El objetivo del ejercicio era cómo afrontar los sesgos en los algoritmos de tecnologías de reconocimiento facial en la Ciudad de México, utilizando los principios normativos para el diseño de plataformas con IA de Microsoft.

Para realizar el ejercicio, se propuso considerar la experiencia de calle, es decir, considerar la realidad de las experiencias micro para afinar los datos entrenan a los algoritmos de plataformas de prevención del delito. Además, se mencionaron específicamente los puntos cinco y seis de las Guías para la interacción humano-máquina de Microsoft (2019).

El punto cinco, que forma parte de la contingencia cuando una interacción humano-máquina sale mal, menciona que las plataformas con IA deben ajustarse a las normas socialmente relevantes. Esto es, adaptarse a la situación particular en aras de crear interacciones que respondan a las necesidades de la población y no necesariamente a los mecanismos con los que se desarrolló el sistema en otros mercados. Asimismo, como lo hemos visto en los casos de Japón, este punto ayuda a integrar los sistemas de IA con elementos que ayuden a solucionar controversias, incluyendo los sesgos discriminatorios.

Ahora bien, el punto seis menciona que hay que mitigar los sesgos sociales. Sí, se pueden ajustar los datos de una IA para prevenir el crimen con datos nacionales o intermunicipales, pero estos datos todavía pueden contener sesgos locales. Es decir, aunque los nuevos datos no tendrán los sesgos norteamericanos de que una persona afrodescendiente tenga más probabilidades de cometer un delito, sí pueden contener sesgos particulares de la región, como el hecho de que personas

con cierto estilo de ropa, con un tono de piel más oscuro y que circulen por ciertas zonas, sean más propensas a cometer delitos.

Por lo tanto, es necesario primero reconocer los sesgos generales de una ciudad y un municipio para tratar de contrarrestar los algoritmos con técnicas como el ruido de los datos. Pero eso no es suficiente, ya que incluso se pueden filtrar sesgos cuando se aplican motores de aprendizaje generales al momento de predecir el delito a nivel local. Esto significa que se deben recopilar datos a nivel de calle, ya que una persona que se considera sospechosa en una zona puede ser una trabajadora habitual en otra y viceversa.

Hay que resaltar qué tipo de sesgos produce la IA particularmente en los sistemas de vigilancia y de prevención de delitos. De acuerdo con Muñoz, el primero de estos es el de los sesgos en los propios datos. La autora explica que ese tipo de sesgos se da porque:

“los datos rara vez son neutros, están ligados a experiencias e historias de personas, por lo que reducirlos a modelos matemáticos sin tener en consideración las circunstancias que los rodean con el objeto de darle una aparente neutralidad, lleva ineludiblemente a resultados incompletos y equivocados” (ídem).

Por su parte, Trejo (2020) explica que los sesgos presentes en los datos y en su estructuración pueden limitar cómo los sistemas de IA interpretan y entienden la realidad. Estos sistemas basan su interpretación en los datos que reciben, pero estos datos no representan la realidad de manera exacta. Según Muñoz (2020), otro tipo de sesgo presente en la IA es el de la representación, que se manifiesta de diversas maneras, como el uso de datos que sólo representan a un tipo de población: este sesgo es común en especial en los sistemas de toma de decisiones automatizadas, que suelen ser estandarizados.

Finalmente, existen sesgos en el diseño y desarrollo de los sistemas de IA. Estos sesgos pueden deberse tanto a las limitaciones del sistema como a las decisiones de las personas que lo desarrollan. Por ejemplo, si una plataforma o sistema no tiene la capacidad de procesar una gran cantidad de datos, tomará decisiones basándose en menos variables y esto puede resultar en sesgos. Por otro lado, si un diseñador no revisa los sesgos en los datos utilizados para entrenar la

plataforma, este sistema reproducirá esos sesgos al momento de su implementación.

Muñoz (2020) menciona que en Europa y América del Norte se han detectado diferentes tipos y formas de discriminación. Según Muñoz, hay dos tipos de discriminación: directa e indirecta. La discriminación directa es fácilmente perceptible, como cuando alguien recibe un trato diferente basándose en sus características personales (género, edad, religión, tono de piel, orientación sexual, entre otros). Por otro lado, la discriminación indirecta se manifiesta cuando una práctica pone en desventaja a ciertos grupos de personas. Por ejemplo:

“De la misma forma, la discriminación también puede ser indirecta, la cual ocurre cuando una disposición o práctica aparentemente neutra pone a las personas que comparten una característica protegida en una desventaja desproporcionada en comparación con otras personas (Kitching, 2005)”.

En el caso de los tipos de discriminación, la IA puede aprender la de manera directa o indirecta. Por ejemplo, la IA puede aprender a discriminar al no reconocer ciertos tonos de piel al momento de procesar una fotografía en un dispositivo móvil, como fue el caso de las cámaras de los Iphone, o de manera indirecta al considerar que ciertas personas que viven en algunas regiones de la ciudad no pueden ser sujetos a crédito. Además, Muñoz (ídem) señala que “es importante destacar que, dentro de una discriminación mediante sistemas de IA, tanto directa como indirecta, una persona puede ser afectada por error, no perteneciendo a un grupo protegido, pero siendo asociada a alguno”.

Es importante recordar que la IA no es un actor moral. Aunque puede influir en la acción de actores humanos en su papel de actante, depende de los datos y de la arquitectura proporcionados por los seres humanos. Para evitarlo, las plataformas de IA y las del C5 deben afinarse constantemente antes de su implementación en público, y que se basen en datos que reflejen la diversidad de escenarios cotidianos locales para evitar nuevos sesgos.

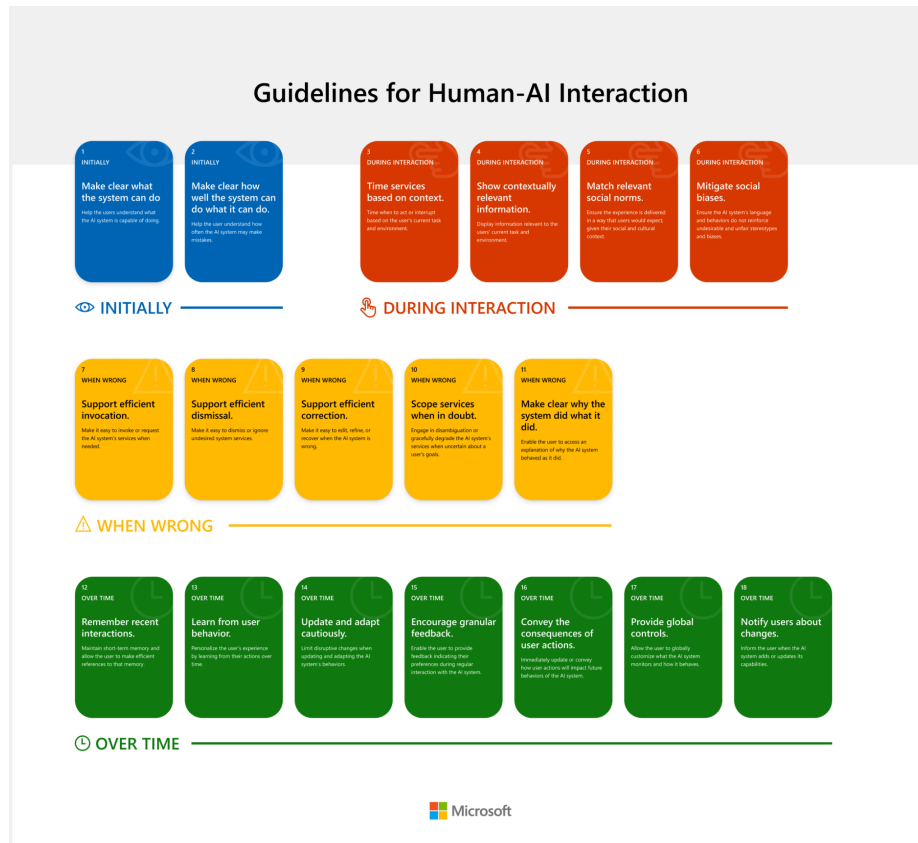
En este sentido, es importante señalar que mitigar los sesgos basados en las condiciones culturales de Estados Unidos y Europa es complicado en México y otros países fuera de estas regiones. Esto se debe a que los contrapesos jurídicos

y las herramientas para detectar la discriminación por parte de la IA están adaptados a contextos específicos. Muñoz explica lo anterior de la siguiente manera:

“un mecanismo para detectar discriminación considerando el tono de piel usado como marcador de racismo, hipotéticamente puede no ser capaz de detectar discriminación en una población donde el tono de piel está compuesto por matices (Mehrabi y otros, 2019). Lo anterior significa no solo que no sería útil, sino que podría generar un resultado de «falso negativo» de discriminación, por lo que se considera primordial el desarrollo de estudios técnicos y legislativos regionales sobre este tema” (ídem).

La autora señala que los mecanismos para prevenir la discriminación en grupos protegidos en Europa y Estados Unidos no funcionan de la misma manera en lugares como México. Esto se debe a que los mecanismos de discriminación y los grupos discriminados obedecen a racionalidades distintas, aunque existan algunos rasgos constantes que se pueden extrapolar, como la discriminación de personas transgénero por parte de la plataforma Uber. Por eso, es necesario crear mecanismos jurídicos y contrapesos en el diseño e implementación de herramientas de IA en el C5 que respeten los derechos humanos, particularmente los de privacidad, acceso a la información y derecho a la no discriminación; además de la necesidad de que respondan a problemas de responsabilidad y explicabilidad en la IA (imagen 11)¹².

¹² El objeto de contorno que se propone es la IA porque mantiene su contenido en todas sus disciplinas, aunque requiera de traducciones. Además, es el objeto central de esta tesis. Entre estas traducciones se encuentran su aplicación a distintas áreas de la vida cotidiana como el transporte, pero en especial, las traducciones relacionadas para aplicar y concebir normativas éticas.



(Imagen 11, fuente de Microsoft)

B. Japón: Muéstrame tu cara y paga

En Japón, específicamente en el área metropolitana de Tokio, los sistemas de reconocimiento facial se utilizan en la Terminal 2 del Aeropuerto de Haneda como mecanismo para abordar vuelos, según la Tokyo International Air Terminal Corporation (2023). En el Aeropuerto de Narita, se utilizan herramientas de reconocimiento facial tanto en migración como para los pases de abordaje en colaboración con las dos grandes aerolíneas nacionales, ANA y JAL, de acuerdo

con International Airport Review (2024). Es importante destacar que Japón cuenta con tecnologías de desarrollo propio en el ámbito de los sistemas de reconocimiento facial, lideradas por empresas como NEC (Advancing Face Recognition, septiembre de 2022, Highlighting Japan). El país también está implementando una prueba piloto para un sistema de pagos con reconocimiento facial en ciudades como Nagano y Kumamoto para el transporte público, según Marubeni Corporation (2023), un sistema que posteriormente podría aplicarse en ciudades más grandes como Tokio.

Hasta 2021, se han instalado aproximadamente 8,500 cámaras de seguridad en las principales estaciones de tren de Tokio, muchas de ellas gestionadas por la reconocida compañía JR East, pionera en operar el primer tren bala (Shinkansen) del mundo. Aunque inicialmente se consideró que estos sistemas identificarían a individuos buscados por la policía o liberados por delitos graves, esta función fue descartada antes de su implementación. Actualmente, las cámaras se utilizan para detectar comportamientos sospechosos y personas buscadas por la policía; sin embargo, no se han revelado detalles sobre cuántas de las 8,500 cámaras

cuentan con sistemas de inteligencia artificial (Mainichi Shinbun, 2021).



NEC Corporation ha implementado un sistema de reconocimiento facial en seis de los principales aeropuertos de Japón, incluidos Narita y Haneda, con el objetivo de agilizar el flujo de pasajeros y reducir los tiempos de espera. Este sistema, basado en inteligencia artificial, combina el motor “NeoFace” y la plataforma biométrica “Bio-Idiom” para capturar y verificar imágenes faciales en terminales de declaración electrónica y puertas de salida. Según la experiencia del autor y de sus familiares en estos aeropuertos (2023 y 2024), estos sistemas han facilitado el acceso y la salida del país, integrándose con otros mecanismos digitales como los códigos QR electrónicos para optimizar los procesos de migración (Nec México, 2020).

En Narita (fotografía 3), además, se han implementado sistemas biométricos en los filtros de abordaje de vuelos internacionales. Los pasajeros escanean documentos y datos faciales antes de continuar con el control de seguridad y

abordar sin intervención humana (Amadeus, 2022; AirportIndustry-News). Este uso de la tecnología responde no solo a la necesidad de eficiencia en un país con escasez de mano de obra, sino también a cambios acelerados por la pandemia de COVID-19. De manera similar, en diciembre de 2024, Marubeni Corporation y Keisei Electric Railway Co., Ltd. anunciaron un sistema de reconocimiento facial para el tren expreso Skyliner, que conecta Narita con Tokio. A partir del 24 de enero de 2025, los pasajeros podrán registrar su rostro tras la compra de su boleto y abordar sin necesidad de documentos físicos. Desde junio de 2024, la línea Yamaman Yukarigaoka en Chiba implementó un sistema similar, integrando reconocimiento facial con pagos electrónicos (Karakaya, 2023; Bovenizer, 2024; Higa, 2024; Marubeni, 2025; The Yomiuri 新聞, 2025).

Si bien estos avances buscan agilizar los desplazamientos, su implementación no es uniforme en todo el transporte japonés. Servicios como Skyliner y Yamaman Yukarigaoka tienen menor afluencia comparados con las líneas principales como la Yamanote o la Chūō. No obstante, la adopción de tecnologías inteligentes es generalizada en el transporte público, con el uso extendido de tarjetas como Suica y PASMO, disponibles también en versiones digitales para Apple Pay (captura de pantalla 1) y Google Pay. Más allá de la conveniencia, el despliegue de estas tecnologías plantea dilemas éticos sobre la privacidad y el control de datos biométricos. Un caso relevante es el de East Japan Railway Co. (JR East), que ha instalado cámaras con reconocimiento facial en 110 estaciones del área metropolitana de Tokio para identificar personas sospechosas. Aunque la empresa lo justifica como una medida de seguridad, la falta de regulación ha generado preocupación por posibles abusos y violaciones a la privacidad (The Asahi 新聞, 2021).

Originalmente, JR East diseñó su sistema para identificar a ciertos exconvictos, pero tras la polémica generada por un informe de Yomiuri Shimbun (ídem), suspendió su implementación argumentando la ausencia de consenso social. No obstante, la compañía sigue almacenando datos faciales de personas consideradas sospechosas, lo que permite a su personal de seguridad interrogarlas o revisar sus pertenencias sin su consentimiento. En Japón, la

recolección y uso de datos biométricos no requiere aprobación de los individuos ni cuenta con regulaciones claras, a diferencia de la Unión Europea, donde el reconocimiento facial está prohibido en principio, o de Estados Unidos, donde existen restricciones a nivel estatal. Como señala Asahi Shimbun (ídem), la falta de supervisión y los riesgos de discriminación por errores de identificación refuerzan la urgencia de normativas claras y de un organismo independiente que regule su uso. La creciente preocupación pública deja en manos de los legisladores la responsabilidad de proteger la privacidad de los ciudadanos ante estos avances tecnológicos.



(Captura de pantalla 1, tomada por el autor).



(Fotografía 3, tomada por el autor).

En conclusión, los sistemas de reconocimiento facial metropolitanos en Japón, en especial Tokio, han sido utilizados tanto para la prevención de delitos como para la optimización del tránsito en aeropuertos y transporte público. Sin embargo, su implementación plantea dilemas éticos y legales, especialmente en lo que respecta a la privacidad y el uso de datos biométricos sin consentimiento. Mientras que las autoridades destacan su eficiencia y seguridad, la falta de regulaciones claras deja a los ciudadanos expuestos a posibles abusos y errores en la identificación. Asimismo, esto refuerza la necesidad de un debate público y de mecanismos de supervisión que equilibren los beneficios tecnológicos con la protección de los derechos individuales.

Si bien los sistemas de reconocimiento facial agilizan el tránsito en aeropuertos y reemplazan boletos físicos en servicios como el tren Skyliner, su implementación

plantea preocupaciones sobre la privacidad. La falta de una regulación específica en Japón deja abierta la posibilidad de un manejo opaco de los datos biométricos. Por ejemplo, aunque se ha confirmado que NEC (2025) almacena estos datos en servidores dentro del país, no se sabe con certeza durante cuánto tiempo ni qué medidas de seguridad los protegen.

Japón, una sociedad que históricamente adopta tecnologías solo cuando han sido rigurosamente probadas, enfrenta el reto de equilibrar la eficiencia tecnológica con la protección de los derechos individuales. Sin un marco legal claro que regule la recopilación y uso de estos datos, el riesgo de filtraciones o abusos es una preocupación legítima. Esto subraya la urgencia de establecer normativas que garanticen la transparencia y la seguridad en la gestión de la información biométrica desde una perspectiva cultural.

C. México: El C5 y la central de abastos

En la Ciudad de México, la IA se utiliza un sistema de reconocimiento facial en el Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano (C5). Sin embargo, según información proporcionada por el Instituto Nacional de Acceso a la Información (INAI) de México el 24 de octubre de 2022, el ámbito de operación de este sistema de reconocimiento facial se limita al polígono de la Central de Abastos, unas 327 hectáreas. Esta área es 50 veces mayor que la plaza central de la Ciudad de México (Zócalo) (Martínez, 2012), pero está lejos de tener la intensa cobertura de sistemas de reconocimiento facial y prevención del delito que se encuentran en algunas ciudades de China.

En la respuesta a la solicitud de información sobre la implementación de sistemas de reconocimiento facial, se ilustra claramente lo mencionado en el párrafo anterior:

“...conforme a la competencia de este Sujeto Obligado se informa que el reconocimiento facial constituye una funcionalidad de analíticos, nativa e integrada del Sistema de Gestión de Video, dicho Sistema a través del cual se hace la

visualización de todo el Sistema de Videovigilancia del sistema implementado en la Central de Abasto” (Sic. 24 de octubre, 2022).

Esa respuesta también revela una falta de transparencia en torno a su funcionamiento, almacenamiento de datos y mecanismos de control, lo que plantea riesgos significativos para los derechos de privacidad, acceso a la información y no discriminación. Al señalar que esta tecnología es una funcionalidad “nativa e integrada” dentro del Sistema de Gestión de Video, se reconoce su existencia sin aclarar quién tiene acceso a los datos biométricos, cuánto tiempo se conservan o con qué fines específicos se utilizan. Esta opacidad impide que ciudadanos y usuarios puedan ejercer su derecho a la información sobre cómo se procesan sus datos personales, dificultando cualquier posibilidad de rendición de cuentas.

Asimismo, en una segunda solicitud el Instituto Nacional de Acceso a la Información (INAI), en la respuesta a la solicitud de Acceso a la Información Pública ingresada al Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la Ciudad de México (C5) en la fecha 09 de enero de 2023, con folio 090168322000339 (C5/CG/UT/016/2023). En esta solicitud se preguntó: 1. ¿Qué tecnología de IA ocupa la dependencia para el cumplimiento de sus funciones? Entendiéndose por tales tecnologías sistemas de reconocimiento facial, gestión de datos, ciberseguridad y análisis[...] 5.- ¿El uso de tecnologías de IA hace algún tipo de identificación respecto a un grupo o personas específicas? 6.-Explicar (de manera general) cómo funciona la tecnología de IA que ocupa la dependencia (Sic). Y en respuesta a las preguntas 1, 5 y 6 la dependencia respondió:

“Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la Ciudad de México, (C5) cuenta con 50 cámaras ubicadas en el Centro de Comando y Control de la Central de Abasto, C2 CEDA con un aplicativo de reconocimiento facial denominado *Interfaz de reconocimiento de rostros* de Seguros, cabe destacar que este aplicativo es utilizado exclusivamente en la demarcación que abarca la Central de Abasto.

La respuesta del C5 también especifica dos tipos de formas en la que usan el sistema de reconocimiento facial:

- i. Rostros capturados: Aquellos rostros que captura el sistema pero que no se encuentran dentro de una lista en la base de datos, es decir, no hubo un resultado de comparación; y
- ii. Rostros Reconocidos: Aquellos rostros que tienen una coincidencia con un rostro dentro de una lista en la base de datos, ya sea que se conoce su identidad o se dio de alta con un alias, como por ejemplo "*persona extraviada* (idem).

Ahora bien, que también con base en las respuestas de las preguntas 1, 5 y seis que se le hicieron al C5 a través del INAI el organismo argumentó lo siguiente con respecto al uso y almacenamiento de los datos que se recolectan de los sistemas de videovigilancia con reconocimiento facial:

Las bases de datos con las que se cotejan los rostros y/o rastros faciales de ciudadanos captados a través del sistema de videovigilancia y reconocimiento facial del C2 CEDA para verificación y/o identificación, son determinadas por los usuarios de dichos sistemas y no por esta Dirección, en este caso, los usuarios de los referidos sistemas son el personal policial que presta sus servicios dentro de las instalaciones del C2 CEDA, el cual está adscrito a la Secretaría de Seguridad Ciudadana de la Ciudad de México (idem).

Una vez más, la respuesta presenta un alto grado de opacidad, a pesar de haber sido diferida con base en los otrora instrumentos del INAI. No obstante, deja entrever que el almacenamiento de datos es, al menos, descentralizado. Al mismo tiempo, evidencia una doble problemática: por un lado, la negativa del C2 CEDA a proporcionar información sobre cómo las personas en la Central de Abasto pueden acceder a sus datos y, por otro, la falta de entendibilidad de los sistemas de IA que operan en estos mecanismos de vigilancia. Esta falta de claridad no solo obstaculiza el ejercicio de derechos fundamentales como la privacidad y el acceso a la información, sino que también abre la puerta a posibles prácticas discriminatorias sin supervisión adecuada.

Así pues, el uso de sistemas de reconocimiento facial en la central de abastos de la Ciudad de México puede tener un impacto significativo en la privacidad de las

personas, ya que estos sistemas recopilan y almacenan información personal, como imágenes faciales, de personas que entran y salen de la central de abastos. Esto puede generar preocupaciones de seguridad y privacidad, ya que los datos recopilados pueden ser utilizados de manera inapropiada o compartidos con terceros no autorizados. Además, que los datos recopilados los determinen los usuarios de los sistemas de videovigilancia, en este caso el personal policial, y no la Dirección del C5, genera un problema de responsabilidad en torno a la privacidad.

Al momento de redactar este párrafo se le envió una solicitud de revisión con esas preguntas a la secretaría de Seguridad de Seguridad Ciudadana. No obstante, después de que se interpuso el recurso de acceso a la información con el INAI, esta institución simplemente le pasó la responsabilidad a otras instituciones como el Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la Ciudad de México. Aun así, después de redactar nuevamente una solicitud de acceso al C5 con preguntas redactadas con asistencia de los modelos de Chat GTP 3 y Bing Assistant (Chat GTP 4), se obtuvieron respuestas más satisfactorias y concretas.

Con esto, hay que tomar en cuenta que el C5 a través de la Dirección General de Administración y Finanzas, expresó que en 2019 se realizó un contrato en torno a la “Implementación del Centro de Comando y Control C2 de la Central de Abasto de la Ciudad de México el cual incluye la implementación de cámaras de videovigilancia con un aplicativo de reconocimiento facial, el cual se adjunta al presente”. Pero no fue hasta la segunda solicitud de información cuando se expresó puntualmente cuál fue la empresa contratada. Así pues, fue hasta que se realizó la segunda solicitud de información donde se le pidió al C5 información más puntual en torno a los datos personales que recopilan.

Las respuestas de las preguntas 2, 3 y 4 (2 ¿La tecnología de IA empleada por la dependencia, recabar algún dato personal? En caso positivo, señalar que dato puede recabar; 3 ¿Se informa en su caso al titular de los datos personales, la recolección de estos mediante tecnologías de IA?; 4 ¿Los datos recolectados son

resguardados? En caso positivo, indicar como se hace el resguardo respectivo) trasladan también la responsabilidad a la Secretaría de Seguridad Ciudadana. Particularmente se mencionó: “se informa que el resguardo y recolección de datos o bases de datos que utiliza la Secretaría de Seguridad Ciudadana para la operación del reconocimiento facial...” (ídem): No obstante, por lo menos se aclara que los datos recolectados por los sistemas de videovigilancia almacenarán alrededor de 60 días con base en “lo dispuesto en los artículos 1, 22, 23, 24, 25, 26, 29, 30 y 31 de la Ley que Regula el Uso de la tecnología para la Seguridad Pública del Distrito Federal” (ídem).

La respuesta del C5 a través del INAI sobre el uso de tecnologías de IA en la central de abastos de la Ciudad de México con respecto a las preguntas 2, 3 y 4 resulta insuficiente porque, aunque se menciona que la Secretaría de Seguridad Ciudadana es la encargada del resguardo y recolección de los datos recopilados por los sistemas de videovigilancia. Y esto no exime al C5 de su responsabilidad en garantizar que los derechos de privacidad de las personas sean respetados. Además, que los datos se almacenen por solo 60 días no significa que se protejan los derechos de privacidad de las personas, ya que estos datos pueden utilizarse de manera inapropiada durante ese periodo.

Por ende, es preocupante que el C5 no informa a los titulares de los datos personales sobre la recolección de estos mediante tecnologías de IA, y no se especifica si los datos recolectados son debidamente resguardados. Es necesaria una evaluación cuidadosa y una regulación adecuada para asegurar que el uso de tecnologías de IA en la central de abastos no viola los derechos de privacidad de las personas. Asimismo, hay que considerar que la falta de evaluación directa los datos por parte del C5 también generan problemas en torno a qué datos se usan para entrenar al sistema, ya que esto puede conducir a sesgos que lleven a prácticas de discriminación.

En cuanto a las herramientas de prevención del delito con IA, el C5 explicó que hasta el momento no se han adquirido herramientas ni sistemas de esta tecnología. Esto se debe a diversas regulaciones, como la de la Unión Europea en

materia de protección de datos personales, así como propuestas regulatorias en torno a la IA, lo que distancia a esta ciudad del uso intensivo de esta tecnología como se hace en China. En respuesta a la solicitud sobre este tema, el C5 se expresó de la siguiente manera:

“–En el ámbito de competencia de ésta Dirección General, en respuesta al requerimiento: –(...) Así como los contratos de adquisición de estas plataformas.” se informa que después de una búsqueda exhaustiva en los archivos físicos y electrónicos que obran en ésta Dirección General, a la fecha, éste Centro no ha celebrado contratos por concepto de adquisición de plataformas y/o programas computacionales para la prevención del delito, por lo tanto no es posible dar respuesta favorable a su solicitud” (Sic. Octubre, 2022).

Así, el mecanismo utilizado por el C5 se basa más en la investigación que en la prevención del delito, ya que incluso tecnologías como el reconocimiento facial se usan solo en áreas específicas de la ciudad. Aunque esto plantea la necesidad de vigilar cómo se utilizan estas tecnologías. Además, el C5 argumenta que el uso de datos obtenidos a través de la videovigilancia es útil para investigar delitos y que todo se hace en cumplimiento de las leyes existentes; de esta manera, los cuerpos policiales:

“...tienen la atribución de hacer del conocimiento del Ministerio Público los hechos delictivos de que tenga noticia, así como poner a su disposición los datos y elementos que hubiere recabado y que obren en su poder; es un asunto que se encuentra de igual manera en el ámbito de competencia de dicho Sujeto Obligado...” (ídem)

Quizás la organización más importante que tiene México tanto para proteger los datos de su población como para acceder a la información del gobierno es el INAI. Instituciones como esta podrían tener un papel importante para que la ciudadanía audite a los gobiernos en aras de que usen IA de manera más responsable, pero también para tener un marco de cómo usan esta tecnología. Para efectos de este trabajo, la plataforma del Instituto Nacional de Transparencia y Acceso a la Información ha sido fundamental para conocer qué modelos de reconocimiento facial operan en la Ciudad de México, cómo operan, y la forma en la que se gestionan los datos a pesar de que se tuvieron aplicar recursos de revisión en algunos casos.

Ahora bien, El INAI es el organismo autónomo encargado de garantizar el derecho de acceso a la información pública y la protección de los datos personales en posesión de los sujetos obligados. Su función es promover la cultura de la transparencia y la rendición de cuentas, así como resolver las solicitudes y recursos de los ciudadanos que buscan ejercer estos derechos. El INAI también tiene facultades para supervisar y sancionar el cumplimiento de la normatividad en materia de datos personales y acceso a la información (INAI, 2021).

Por otro lado, hay que recordar que el reconocimiento facial es una aplicación de la IA que consiste en identificar o verificar la identidad de una persona a partir de una imagen o un video de su rostro. Esta tecnología se ha implementado en diversos ámbitos, como la seguridad pública, el control migratorio, el comercio o el entretenimiento. Sin embargo, el reconocimiento facial también plantea serios problemas de privacidad, discriminación, vigilancia y libertades civiles. Por ello, es fundamental que los ciudadanos conozcan cómo funciona esta tecnología, quién la usa y con qué fines, y qué derechos tienen para acceder, rectificar o cancelar sus datos personales. (UNESCO, 2019).

Es crucial entender quién y cómo accede a la información recopilada por las cámaras de reconocimiento facial en la CDMX. En marzo de 2024, se realizó una solicitud de información al C5 (Folio Público 090168324000068) donde se indagó sobre las medidas de supervisión y regulación implementadas para prevenir abusos. En respuesta, el C5, a través del INAI, informó lo siguiente:

- Accesos controlados a las áreas de despacho de los C2's y el C5, través de lectores de datos biométricos.
- Los puertos USB, el lector de disco y quemador de disco de las computadoras operativas
- se encuentran deshabilitados.
- EL CPU de cada computadora operativa está ubicada dentro de un gabinete cerrado con llave y únicamente puede ingresar personal autorizado, por alguna falla técnica del aparato.

- En las computadoras no se almacena la información generada por la atención a emergencias, ya que esta se va guardando directamente en el sistema, los servidores y equipos de almacenamiento.
- Administración de los accesos a través de políticas de dominio, directorio activo, políticas de antivirus que permite restringir el uso de dispositivos, instalación de programas y accesos a los diferentes niveles de la red de datos.
- Igualmente, se han implementado medidas de seguridad técnicas para controlar los accesos al sistema, tales como:
- Los accesos son a través de usuarios y contraseñas, que son otorgados al personal que la Secretaría de Seguridad Ciudadana de la Ciudad de México solicita.
- Existen diferentes niveles de acceso a la información dependiendo los requerimientos de
- la referida dependencia, a través de perfiles de usuario, conforme a los requerimientos de ésta para el ejercicio de sus funciones.
- Se realiza el mantenimiento constante al software y hardware de los sistemas, a fin de garantizar la seguridad y funcionalidad de los mismos.
- El almacenamiento de la información está en servidores de los inmuebles, cuyo acceso también se restringe con lectores de datos biométricos y un teclado numérico para ingresar las contraseñas correspondientes.
- Administración de los usuarios a través de políticas de dominio y directorio activo que permite configurar perfiles de acceso dependiendo de la solicitud de creación de usuarios.

Con base en esta información, sí se han implementado diversas medidas para proteger los derechos humanos en relación con el uso de cámaras de reconocimiento facial en la Ciudad de México (CDMX). Además, a dif es necesario señalar que hasta abril de 2024 el uso de estas herramientas sigue siendo restringido en el espacio público de la Central de Abastos. Aun así, es importante seguir supervisando y regulando su aplicación para evitar posibles abusos y garantizar el respeto a los derechos humanos.

Finalmente es importante resaltar que la plataforma del INAI es una herramienta que permite a los ciudadanos solicitar información pública a los sujetos obligados, como dependencias gubernamentales, organismos públicos autónomos o partidos políticos. A través de esta plataforma, se pueden consultar las respuestas a las solicitudes de otros ciudadanos, e interponer recursos de revisión si no se cumple con la respuesta recibida.

Y la plataforma del INAI ha sido fundamental para conocer qué modelos de reconocimiento facial operan en la Ciudad de México, cómo operan y cómo se gestionan los datos personales captados por esta tecnología; sin embargo, también se han presentado casos, como el de acceso a la información de este trabajo con respecto a los sistemas de reconocimiento facial en la Central de abastos, en los que se ha negado o limitado el acceso a esta información, lo que evidencia la necesidad de fortalecer el marco legal y las capacidades institucionales para garantizar el derecho a la información y la protección de datos personales en el contexto de la IA. (R3D: Red en Defensa de los Derechos Digitales, 2020)

Es de suma importancia recordar que se interpuso un nuevo recurso al C5 con base en las herramientas Chat GTP 3 y Bing Assistant (Chat GTP 4), además de la asistencia de la licenciada Sofía, en aras de obtener respuestas más profundas y claras en torno a las problemáticas del tema de investigación. Con base en estas herramientas, y las instrucciones que se le dieron a estos sistemas para redactar un recurso que tuvieran más posibilidades de respuesta, en abril de 2023 se formularon las siguientes preguntas que fueron asignadas al Acceso a la Información y Protección de Datos Personales para obtener una respuesta más profunda del C5, y estas preguntas quedaron de la siguiente manera:

1. "¿Cuántas cámaras con reconocimiento social se instalaron en la central de abastos y en qué áreas específicas están ubicadas?
2. ¿Qué tipo de datos se recopilan a través de estas cámaras y cómo se utilizan?

3. ¿Se han presentado denuncias o quejas sobre la invasión de la privacidad de las personas que
4. transitan por la central de abastos debido a la implementación de estas cámaras?
5. ¿Existen protocolos o medidas de seguridad para protegerla información recopilada a través de estas cámaras?
6. ¿Ha habido alguna violación de la protección de datos personales de las personas que han sido monitoreadas a través de estas cámaras?
7. ¿Cómo se han tratado los problemas relacionados con estas cámaras y qué medidas se han tomado para solucionarlos?
8. ¿Qué medidas se han tomado para garantizar que el uso de estas cámaras sea conforme a la ley y respete los derechos humanos de las personas?
9. ¿Cuáles son las políticas y regulaciones que rigen el uso de cámaras con reconocimiento social en México?
10. ¿Quiénes son los contratistas que colocaron las cámaras con reconocimiento facial en la central de abastos?
11. ¿Qué tipo de tecnología con IA usan las cámaras con reconocimiento facial en la central de abastos?
12. ¿En qué servidores se almacenan los datos de las cámaras con reconocimiento facial en la central de abastos?
13. ¿Quién es el proveedor del sistema operativo SecurOs?. "(Sic, recurso de abril 2023 que se respondió el 2 de mayo de 2023).

Con respecto a la primera pregunta de la solicitud de abril del 2023, el C5 respondió:

La ubicación de las cámaras de videovigilancia que cuentan con la tecnología de reconocimiento facial instaladas en la Central De Abasto de la Ciudad De México reviste el carácter de información reservada, por lo que esta Dirección General propone la clasificación como de acceso RESTRINGIDO en su modalidad de RESERVADA, en términos de lo dispuesto en los artículos, 24 fracción VI, 43

párrafo tercero, 44 fracción II y IV, 45 fracción II y VI, 100, 103, 104, 106, 108 párrafo tercero, 113 fracción I, 114 y 137 de la Ley General de Transparencia y Acceso a la Información Pública; artículos, 6 fracciones XXI, XXV, XXXVI y XLIII, 8, 90 fracciones

11, IV, VIII, XI Y XII, 93 fracciones III, VI y X, 171 segundo párrafo, 173, 174, 176, 177, 179, 178, 183 fracción I y IX, 184 y 216 de la Ley de Transparencia, Acceso a la Información Pública y Rendición de Cuentas de la Ciudad de México; así como los artículos 2 y 23, fracciones I y II de la Ley que Regula el Uso de Tecnología para la Seguridad Pública del Distrito Federal y Capítulos, I numeral segundo, fracción XI, Capítulo II, numeral Cuarto, Quinto, Séptimo y Décimo Séptimo, Capítulo IV, numeral Décimo Octavo y Trigésimo Tercero de los Lineamientos Generales en materia de clasificación y desclasificación de la información, así como la elaboración de las versiones públicas; por lo que se solicita que por su amable conducto se convoque a los integrantes del Comité de Transparencia de este Centro, a fin de someter a consideración de los mismos la presente clasificación de información como reservada, a efecto de que se confirme la clasificación propuesta para los cuales manifiestan las siguientes (sic).

En este sentido, según el C5, la información solicitada se reserva porque su difusión peligraría la seguridad pública y la prevención del delito, al facilitar el acceso a los sistemas de vigilancia y su posible manipulación o sabotaje por parte de personas o grupos delictivos. La clasificación de la información como reservada se sustenta en el artículo 14 fracción III de la Ley General de Transparencia y Acceso a la Información Pública y el artículo 15 fracción III de la Ley de Transparencia, Acceso a la Información Pública y Rendición de Cuentas de la Ciudad de México, que establecen que se podrá reservar la información que afecte la seguridad pública o nacional. Por ello, la reserva de la información se mantendrá por un plazo de cinco años, contados a partir de su generación, conforme al artículo 110 fracción I de la Ley General de Transparencia y Acceso a la Información Pública y el artículo 182 fracción I de la Ley de Transparencia, Acceso a la Información Pública y Rendición de Cuentas de la Ciudad de México.

Con respecto a la pregunta dos, el C5 respondió lo siguiente a través del INAI:

...se informa que el reconocimiento facial constituye una funcionalidad de analíticos, nativa e integrada del Sistema de Gestión de Video, dicho Sistema través del cual se hace la visualización de todo el Sistema de Videovigilancia del sistema implementado en la Central de Abasto.

Dentro de lo plataformas se tienen 2 tipos de resultados:

1) Rostros capturados: Aquellos rostros que captura el sistema pero que no se encuentran dentro de una lista en la base de datos, es decir, no hubo un resultado de comparación; y

2) Rostros Reconocidos: Aquellos rostros que tienen una coincidencia con un rostro dentro de una lista en la base de datos, ya sea que se conoce su identidad o se dio de alta con un alias, como por ejemplo "persona extraviada" (sic).

En torno a esta respuesta se debe señalar y recordar que todo sistema de reconocimiento facial implica la recolección y el tratamiento de datos personales sensibles, como son los rasgos faciales de las personas. Estos datos pueden revelar información sobre la identidad, el origen étnico, la edad, el género, la salud o las preferencias de las personas. Por ello, el uso del sistema de reconocimiento facial debe respetar el derecho a la privacidad y el derecho a la protección de datos personales de las personas capturadas y reconocidas por el sistema. Asimismo, el uso del sistema de reconocimiento facial debe evitar cualquier forma de discriminación o estigmatización basada en los rasgos faciales de las personas; por ello, se debe garantizar que el sistema sea preciso, confiable y transparente, y que se cuente con el consentimiento informado y la participación de las personas afectadas por el sistema, y en este sentido el C5 argumenta que se está empleando medidas de seguridad física para el tratamiento y protección de datos:

- Accesos controlados a las áreas de despacho de los C2's y el C5, a través de lectores de datos biométricos.
- Los puertos USB, el lector de disco y el quemador de disco de las computadoras operativas se encuentran deshabilitados.
- El CPU de cada computadora operativa está ubicada dentro de un gabinete cerrado con llave y únicamente puede ingresar personal autorizado, por alguna falla técnica del aparato.
- En las computadoras no se almacena la información generada por la atención a emergencias, ya que esta se va guardando directamente en el sistema, los servidores y equipos de almacenamiento.

- Administración de accesos mediante políticas de dominio, directorio activo, políticas de antivirus que restringen el uso de dispositivos, instalación de programas y accesos a los niveles de la red de datos.
- Igualmente, se han implementado medidas de seguridad técnicas para controlar los accesos al sistema, tales como:
- Los accesos son a través de usuarios y contraseñas, que son otorgados al personal que el gobierno de la ciudad de México la Secretaría de Seguridad Ciudadana de la Ciudad de México solicita.
- Existen diferentes niveles de acceso a la información dependiendo los requerimientos de la referida dependencia, a través de perfiles de usuario, conforme a los requerimientos de ésta para el ejercicio de sus funciones.
- Se mantiene constante el software y hardware de los sistemas para garantizar la seguridad y funcionalidad de estos.
- El almacenamiento de la información está en servidores de los sitios de los inmuebles, cuyo acceso también se restringe con lectores de datos biométricos y un teclado numérico para ingresar las contraseñas correspondientes.
- Accesos controlados a las áreas de despacho de los C2's y el C5, a través de lectores de datos biométricos.
- Los puertos USB, el lector de disco y el quemador de disco de las computadoras operativas se encuentran deshabilitados.
- El CPU de cada computadora operativa está ubicada dentro de un gabinete cerrado con llave y únicamente puede ingresar personal autorizado, por alguna falla técnica del aparato.
- En las computadoras no se almacena la información generada por la atención a emergencias, ya que esta se va guardando directamente en el sistema, los servidores y equipos de almacenamiento.

- Administración de accesos mediante políticas de dominio, directorio activo, políticas de antivirus que restringen el uso de dispositivos, instalación de programas y accesos a los niveles de la red de datos.
- Igualmente, se han implementado medidas de seguridad técnicas para controlar los accesos al sistema, tales como:
- Los accesos son a través de usuarios y contraseñas, que son otorgados al personal que la Secretaría de Seguridad Ciudadana de la Ciudad De México solicita.
- Existen diferentes niveles de acceso a la información dependiendo los requerimientos de la referida dependencia, a través de perfiles de usuario, conforme a los requerimientos de ésta para el ejercicio de sus funciones.
- Se mantiene constante el software y hardware de los sistemas para garantizar la seguridad y funcionalidad de estos.
- El Almacenamiento de la información se encuentra en servidores ubicados en los sitios de los inmuebles, cuyo acceso también está restringido a través de lectores de datos biométricos y un teclado numérico para ingresar las contraseñas correspondientes.

Ahora bien, en torno a la pregunta cinco, el C5 respondió:

“Me permito informarle que como se precisó con anterioridad la operatividad y / o monitoreo de las cámaras de videovigilancia con el aplicativo de reconocimiento facial, así como el resguardo y uso de la información obtenida a través de este lo realiza el personal policial que presta sus servicios en las instalaciones del C2 CEDA designado por la Secretaría de Seguridad Ciudadana de la Ciudad de México, no este Centro, por lo que, corresponde a dicha Secretaría conocer la información de su interés.

Sin embargo, dadas las atribuciones de este órgano desconcentrado, se informa que a nivel sistema no ha existido ninguna afectación que pudiera exponer o vulnerar información (sic).

La respuesta de la pregunta seis fue la siguiente:

“me permito informarle que en el marco de las funciones y atribuciones de este Centro se informa que no existen denuncias o quejas sobre el tema que indica, aclarando que este Centro únicamente fue quien en ejercicio de sus funciones

implementó el C2 CEDA y el sistema de vigilancia, sin embargo, la operación de las cámaras de videovigilancia, el monitoreo y análisis de la información captada corresponde a personal policial adscrito a la Secretaría de Seguridad Ciudadana”.

Además, el C5 respondió así a la pregunta siete:

“por lo que hace al mantenimiento correctivo, este se implementa cuando se genera un reporte de falla en alguno de los elementos del STV (cámara, botón de auxilio y altavoz) para reparar dicha falla; de manera paralela para llevar un seguimiento del funcionamiento de cada STV y por lo que hace al mantenimiento preventivo, se lleva a cabo a través de un programa establecido, siendo en esta actividad en la cual se hace una revisión periódica en sitio en cada uno de los STVs, generando una prueba de cámara, botón y de audio en los altavoces en cada visita que se le haga al ST”.

En torno a la pregunta ocho, el C5 respondió:

“me permito informarle que la operación de las cámaras de videovigilancia con el aplicativo de reconocimiento facial se encuentra a cargo de la Secretaría de Seguridad Ciudadana, en el Centro de Comando y Control C2 CEDA, en cumplimiento de sus funciones cuyo objetivo es la prevención del delito y la seguridad ciudadana;

así como la operación de cualquier sistema tecnológico de videovigilancia en la Ciudad de México de conformidad con lo dispuesto en los artículos 1, 20 y 24 de la referida Ley que Regula el Uso de Tecnología Para la Seguridad Pública del Distrito Federal, por lo que corresponde a la Secretaría de Seguridad Ciudadana de la Ciudad de México, la operación de las cámaras de video vigilancia, así como la clasificación, análisis, custodia, y remisión de la información obtenida con los equipos y sistemas tecnológicos”.

Es relevante mencionar que en la respuesta de la pregunta diez de esta solicitud de información, el C5 por fin reveló que el proveedor del sistema operativo con el que operan las cámaras con reconocimiento facial es Teléfonos de México. Pero con respecto a los contratistas instalaron las cámaras de videovigilancia y con reconocimiento facial, el organismo de nueva cuenta remitió esta pregunta a la secretaria de Seguridad Ciudadana. Aun así, este tercer recurso de acceso a la información arrojó respuestas más profundas con respecto al uso y empleo de las cámaras con reconocimiento facial por parte del C5.

El C5 se deslindó de la responsabilidad de la operación y el uso de la información de las cámaras de videovigilancia con reconocimiento facial, que recae en la Secretaría de Seguridad Ciudadana, y que podría implicar una violación a la

privacidad y la identidad de las personas. Además, negó que haya habido alguna falla en el sistema o alguna denuncia o queja al respecto. Por último, describió el mantenimiento de los elementos del STV como una actividad rutinaria y sin mayor relevancia.

Si bien, no se pudo reconocer quiénes son los proveedores de las 50 cámaras con sistemas de reconocimiento facial mediante los recursos interpuestos al INAI, otros medios como Contra Línea plantean que “los proveedores de esta tecnología son: Telmex, Hanwa y Micofocus, en el caso de la Central de Abastos” (Flores, 2021). Esta información es confirmada por una investigación realizada por la organización Al Sur (Venturini y Vladimir Garay, 2021), la cual plantea:

La empresa mexicana Teléfonos de México S.A.B. de C.V. fue adjudicada en una licitación pública nacional en el valor total de 197.564.863,80 MXN (alrededor de 9.702.029 USD) para la implementación del Centro de Comando y Control C2 de la Central de Abastos de la Ciudad de México de la Central de Abastos de la Ciudad de México.

Hay que resaltar que uno de los elementos más importantes que se descubrieron, a pesar de los problemas en torno para conseguir información en torno al C5 (lo cual, incluso requirió un recurso de impugnación) fue saber cuál es el sistema operativo con el se operan las cámaras de vigilancia: el sistema operativo Secure Os. De acuerdo con la respuesta que brindó el C5 (09 de 2023):

Este Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la Ciudad de México, (C5) cuenta con 50 cámaras ubicadas en el Centro de Comando y Control de la Central de Abasto, C2 CEDA con un aplicativo de reconocimiento

facial denominado "Interfaz de reconocimiento de rostros" de SecurOS, cabe destacar que este aplicativo es utilizado exclusivamente en la demarcación que abarca la Central de Abasto. En este sentido, la compañía Syscom Internet Security Systems (SYSCOM: ISS, s. f.) expone que SecurOS:

“...contempla un amplio abanico de soluciones diseñadas para adaptarse a cualquier proyecto y complejidad (...) Uno de los diferenciadores de la analítica de

imágenes, ya sea de reconocimiento facial, Reconocimiento de placas (LPR), u otras soluciones, es la naturaleza de los algoritmos de ISS. Las Soluciones de video analíticos de ISS son reconocidas por su precisión y su capacidad de captura de múltiples dígitos o rostros. Además, se cuentan con analíticos situacionales y de rastreo como son: cruce de línea, identificación de matrícula de contenedores y trenes, objeto olvidado, detección de multitud o merodeo, punto de venta (POS), etc.; y con la unificación de aplicaciones de terceros, analizadas por ingeniería”.

La implementación de sistemas con IA como SecureOS en la vigilancia y monitoreo de la ciudad puede tener implicaciones en la privacidad y acceso a la información de los ciudadanos, ya que estos sistemas recopilan y procesan grandes cantidades de datos. Además, existe la posibilidad de que estos sistemas perpetúen y amplíen la discriminación hacia ciertos grupos sociales, como ha sucedido en otros lugares del mundo. En particular, porque el sistema alimentado de datos se ha entrenado con datos, desarrollado con fines de lucro, y en especial si se implementa para videovigilancia con herramientas de aprendizaje máquina, tendrá problemas en torno a la transgresión de los derechos privacidad, acceso a la información y no discriminación en sus plataformas con IA.

Desde la perspectiva de la teoría de actor-red, estos sistemas de vigilancia se pueden entender como un conjunto de actores humanos y no humanos que interactúan entre sí para llevar a cabo una función específica. En este caso, el sistema SecureOS, la compañía Syscom Internet Security Systems, los servidores, cámaras y otros componentes tecnológicos involucrados actúan como actores no humanos, mientras que los operadores y supervisores del sistema, las autoridades gubernamentales y la población en general actúan como actores humanos.

¿Qué tan seguro es SecurOS? Aun no se han realizado estudios con respecto al impacto de SecurOs; no obstante, como se ha discutido a lo largo de este trabajo, existen riesgos importantes en torno a la implementación de sistemas de video vigilancia, en especial aquellos con la capacidad de reconocimiento facial. Por ejemplo, un estudio comparativo entre los sistemas de reconocimiento facial de Reino Unido y Estados Unidos realizado por Almeida et al (2021) plantean que:

Aplicar este mismo marco conceptual a los algoritmos de reconocimiento facial, a veces defectuosos, sin asumir la responsabilidad por las consecuencias de este uso no solo podría conducir a una mayor discriminación y victimización de comunidades específicas, sino también a una pérdida de confianza aun mayor entre la población en general y la ley. organismos de aplicación.

Además de los riesgos de entorno a sesgos que pueden, y han generado discriminación, persiste el problema de privacidad y con ello el reto de la responsabilidad y el fenómeno de la caja negra. La respuesta que dio el C5 alrededor de la pregunta de qué se hacen con los datos y quién los guarda, expone que el organismo quien guarda los datos es el C2 o Centro de Comando y Control por una duración de 60 días. La respuesta también plantea grosso modo que los datos son utilizados para fines de seguridad bajo las siguientes premisas: la prevención del delito y la seguridad ciudadana: así como la operación de cualquier sistema tecnológico de videovigilancia en la Ciudad de México. Asimismo, en el ámbito de competencia de este Centro, se informa que el Sistema de reconocimiento facial implementado en el C2 CEDA, a nivel técnico tiene las siguientes funcionalidades.

Dentro de la plataforma se tienen dos tipos de resultados: Rostros capturados: Aquellos rostros que captura el sistema pero que no se encuentran dentro de una lista en la base de datos, es decir no hubo un resultado de comparación y Rostros Reconocidos: Aquellos Rostros que tienen una coincidencia con un rostro dentro de una lista en la base de datos, va sea que se conoce su identidad se dio de alta con un alias, como por ejemplo "*persona extraviada*".

En este sentido se ve un problema en torno al acceso a la información con respecto a qué hace el C2 CEDA con los datos que emplea y recolecta de los sistemas de reconocimiento facial y de videovigilancia en la Central de Abastos. De nueva cuenta, este es un problema ético transversal dentro de los sistemas que emplean esta tecnología. Por ello, como lo explican Almeida et al (2021) en su artículo titulado *The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks. In AI and Ethics:*

El segundo tema se centra en las diferencias estado por estado; ocurre gracias a la ausencia de una legislación federal general sobre privacidad, y la ley estatal a menudo proporciona sólo derechos muy básicos para responsabilizar a los operadores de FRT. El alcance de las leyes de privacidad en la mayoría de los estados se limita a notificar a una persona si sus datos han sido robados en una brecha de seguridad [27], lo que no es un consuelo para alguien que se ha visto afectado por un uso malicioso o sesgado involuntario de FRT.

El problema de derechos humanos relacionado con el acceso a la información en el C2 CEDA puede tener implicaciones similares a las que se presentan en otros sistemas que utilizan tecnología de reconocimiento facial y videovigilancia. Esto en buena parte tiene que ver con la diversidad de datos con los que se ha entrenado algoritmo, no obstante, como planteó Leslie (2020) “los problemas de prejuicio y juicio erróneo en torno a cómo estos datos están etiquetados y anotados también puede conducir a la discriminación”. Así pues, los sistemas de reconocimiento facial como el implementado en la Central de Abastos puede conducir a problemas de discriminación con repercusiones en torno a la vigilancia y sanción de un delito, por lo que es importante conocer cómo y qué bases de datos se ha entrenado al sistema SecurOs con el que estas cámaras operan.

Además, existe una falta de legislación federal general sobre privacidad en México con respecto al impacto en la privacidad que sistemas con IA pueden tener. Esto es debido a que, en la mayoría de los casos, las leyes de privacidad sólo se articulan en torno a la protección de datos personales, que sí bien han sido de utilidad como el caso de Richter contra Uber, aún no están articuladas para enfrentar los problemas de responsabilidad, privacidad y sesgos con respecto a los sistemas de reconocimiento facial que operan con IA. Por ello, se requieren crear herramientas y formatos para poder enfrentar problemas con respecto al daño, intencional o no, que sistemas como SecurOs puedan hacer durante y antes de su implementación.

Así pues, la implementación de sistemas de IA en videovigilancia, como SecureOS, puede tener implicaciones para la privacidad y el acceso a la información. Estos sistemas son capaces perpetuar y amplificar la discriminación contra ciertos grupos sociales y tener problemas de privacidad, acceso a la

información y no discriminación. Particularmente, es necesario examinar aún más el impacto del sistema operativo SecureOS, que se utiliza en la Central de Abastos de la Ciudad de México; sobre todo porque esta herramienta incluye reconocimiento facial y una amplia gama de análisis de video alrededor de la IA.

D. Reflexiones finales sobre los sistemas de reconocimiento facial metropolitanos en México y Japón desde la mirada de la teoría de actor-red

Desde la perspectiva de la teoría de actor-red, estos sistemas de vigilancia se pueden entender como un conjunto de actores humanos y no humanos que interactúan entre sí para llevar a cabo una función específica. No obstante, como lo plantea la misma teoría de actor-red, el impacto y los riesgos en torno a los sistemas de videovigilancia, en este caso el de la Central de abastos, puede ser contenido si se detiene su programa u objetivo. En este sentido, los actores pueden resistir la acción del programa al conocer lo qué hace y cómo funciona: o por medio de otro objeto o legislación, como es el caso de las equivalencias en mundo cívico que ilustran Boltanski y Thevenot.

En torno al derecho de acceso a la información, es importante saber quién puede ser el responsable de los problemas que ocasionen los sistemas de reconocimiento facial en la central de abastos. Al final de cuentas, esas herramientas con IA son números sin espíritu, estadísticas que se alimentan de almas en datos que pueden contener sesgos, y por ello, como bien lo plantea Carissa Velíz, no tienen capacidad de hacer juicios morales. Así pues, necesario contar con herramientas con el INAI que ayuden a rastrear quiénes son los actores humanos que, en el último caso, son quienes son los responsables de empujar un programa dentro de la red: los responsables de la Secretaría de Seguridad Social, quienes desarrollaron el software de SecurOs, y quiénes toman decisiones en torno a los datos y predicciones de los sistemas de reconocimiento facial de la Central de abastos.

Con base en esto se pueden determinar los problemas con respecto al C5 así. El programa es emplear la IA en las cámaras de vigilancia para prevenir y ayudar a combatir delitos, el programa son los las transgresiones presentes y posibles con respecto a los derechos de privacidad, acceso a la información, y el derecho a la no discriminación; el antiprograma surge en el momento de resistencia por parte de los actores humanos que pueden generar un actante capaz de alterar el antiprograma para que el programa original (prevenir y combatir delitos) avance, o simplemente tirar el antiprograma y el programa que lo originó.

Así pues, los actores humanos que se oponen al antiprograma de las violaciones o amenazas a sus derechos, buscan crear o movilizar otros actores no humanos que les ayuden a defender su programa original de prevenir y combatir delitos. Por ejemplo, podrían recurrir a la constitución, a los jueces, a las organizaciones civiles, a los medios de comunicación o a las redes sociales, para generar resistencias o alianzas que les permitan seguir utilizando la IA en las cámaras de vigilancia sin renunciar a sus derechos. De esta manera, el antiprogramase manifiesta en la negociación o el conflicto entre diferentes actores humanos y no humanos que tienen intereses o valores distintos o contrapuestos.

No obstante, para crear un anti antiprograma eficiente, y con ello la capacidad de resistencia, hay que tomar en cuenta que en la red de asociaciones surgen de manera constante nuevos actantes. Por ejemplo, los sistemas con IA en áreas con video vigilancia que el C5 logre implementar en un futuro herramientas de la IA generativa como Generative Pre-trained Transformers (GPT o transformadores generativos pre-entrenados). Y hay que resaltar que la IA generativa es un conjunto de algoritmos que utilizan datos de entrenamiento para generar contenido nuevo (texto, imágenes, audio o video) (CampusMVP, s.f.): Chat GPT, quizás es la aplicación de IA generativa más conocida hasta la fecha, es un chatbot creado en los modelos de lenguaje grande (LLM) GPT-3.5 y GPT-4 de OpenAI (BBC, 2023). Estas herramientas podrían generar nuevos programas o antiprogramas que afecten a los derechos humanos de la IA, como la privacidad, el acceso a la información o la no discriminación.

Aun así, más allá de los nuevos actantes que entren a la red asociaciones, dentro de la red, los responsables del programa son actores humanos. Y con base en los planteamientos de Trejo y Véliz, al final, los responsables de transgresiones en los derechos humanos por parte del C5 y toda herramienta con IA son los actantes humanos. Por lo tanto, es necesario examinar las prácticas de crítica y de justificación que estos actores humanos realizan frente a los problemas que plantea la IA.

Hay que resaltar que los mediadores a favor el caso de los sistemas de reconocimiento facial metropolitano, como el C5, son aquellos que defienden la seguridad, la prevención, la vigilancia y el control de la población y el territorio. Estos mediadores pueden ser autoridades, policías, militares, empresas, medios de comunicación además de otros actores que promueven o participan en el uso y la gestión de estos sistemas, como las cámaras, los sensores, las bases de datos, los centros de monitoreo. No obstante, también resaltan tanto los fabricantes y distribuidores de estos sistemas que los promueven como un mecanismos para reducir el crimen, así como aquellos mediadores a modo de funcionarios públicos que fomentan su instalación y su implementación so pretexto de mayor seguridad sin considerar las transgresiones a nivel de derechos humanos.

Ahora bien, los mediadores en contra de sus sistemas de IA son aquellos que protegen la privacidad, la libertad, la dignidad y los derechos humanos de las personas y las comunidades. Estos mediadores pueden ser ciudadanos, activistas, abogados, académicos, organizaciones de derechos civiles que cuestionan o resisten la implementación y el impacto de estos sistemas, como la discriminación, la violación, el abuso, la represión además de los distintos niveles de transgresión que estos sistemas de reconocimiento facial metropolitano. Precisamente estos mediadores dan voz a distintos actores para generar resistencia en torno a transgresiones en derechos humanos, en asociaciones humanas en torno a la teoría de actor-red, técnico, y sociotécnico.

NIVEL	TRANSGRESIÓN	EJEMPLO
Derechos humanos	Los sistemas de reconocimiento facial violan	En Estados Unidos, se ha documentado que la tecnología de reconocimiento facial refuerza las

	el derecho a la privacidad, a la libertad de expresión, a la no discriminación, a la presunción de inocencia y a la protección de datos personales de las personas que son sometidas a su escrutinio sin su consentimiento ni control (Amnistía internacional, s.f).	prácticas policiales racistas y discriminatorias, al identificar erróneamente a personas negras y latinas como sospechosas de delitos ³ . En China, se ha denunciado que el gobierno usa el reconocimiento facial para vigilar y reprimir a la minoría musulmana uigur (Amnistía internacional, 2022).
Asociaciones humanas	Los sistemas de reconocimiento facial alteran las formas de relacionarse e interactuar de las personas en el espacio público, al generar una sensación de vigilancia constante y de pérdida de anonimato. Esto puede afectar la confianza, la solidaridad, la creatividad y la participación social de las personas Amnistía Internacional, s.f).	En Reino Unido, se ha reportado que el uso de reconocimiento facial por parte de la policía ha provocado que algunas personas eviten ciertas áreas, cambien su apariencia o se cubran el rostro para evitar ser identificadas. En México, se ha criticado que el uso de reconocimiento facial en el transporte público y en las calles viola el derecho a la libre circulación y a la protesta social (Amnistía Internacional, s.f).
Técnico	Los sistemas de reconocimiento facial presentan limitaciones y sesgos técnicos que afectan su precisión, su fiabilidad y su transparencia. Estos sistemas dependen de la calidad de los datos, de los algoritmos y de los dispositivos que los implementan, y pueden ser vulnerables a errores, manipulaciones o hackeos (Amnistía internacional, s.f)	Como lo plantea O'Neil (2016), los sistemas de reconocimiento facial suelen contener o reproducir riesgos que fomentan la discriminación facial. En este sentido, el problema general es que una IA puede producir resultados que tienen efectos diferenciales inaceptables desde el punto de vista ético (y, en algunos casos, legal) en diversas subpoblaciones, y esto puede ocurrir sin que los ingenieros tengan la intención de discriminar (Blackman, 2022). Por ejemplo, un estudio de la Universidad de Essex en el Reino Unido demostró que la tecnología que probaron era precisa sólo en el 19% de los casos (Ahmad, 2023).
Sociotécnico	Los sistemas de reconocimiento facial implican una compleja red de actores humanos y no humanos que interactúan y se influyen mutuamente, generando efectos sociales,	Podrías pensar que la solución fácil es incluir más diversidad en los sistemas de entrenamiento para que funcionen mejor. Pero hay una pregunta más grande aquí: ¿queremos que funcione bien para vigilar a las personas negras? ¿Para vigilar a nadie? Y fundamentalmente, ¿queremos que exista en

	políticos, económicos y culturales que van más allá de la mera aplicación técnica. Estos sistemas están insertos en contextos históricos, institucionales y normativos que determinan sus fines, sus usos y sus consecuencias.	absoluto? (Ahmad, 2023) En este sentido, en Detroit, en 2020, la policía arrestó a un hombre negro llamado Robert Williams basándose en una identificación por reconocimiento facial. Él no había hecho nada malo (Ahmad, 2023). Además de ello está el problema de cómo los sesgos de los sistemas reconocimiento facial se relacionan con los problemas de discriminación en México y en Japón, lo cuáles no han recibido tanta atención como en sus pares europeos o del norte de América.

Desde la perspectiva de la sociología pragmática de la crítica de Boltanski, se podría analizar cómo los actores humanos involucrados en el uso y el desarrollo de la IA en el C5 recurren a diferentes principios o mundos para evaluar y resolver los conflictos que surgen en torno a los derechos humanos. Por ejemplo, en el mundo cívico, el principio de justicia es el derecho; en el mundo industrial, es la eficiencia; y en el mundo del proyecto, es la creatividad. Así, se podría estudiar cómo los actores humanos que participan en el C5 apelan a distintos principios de justicia para defender o cuestionar el uso de la IA en las cámaras de vigilancia, y cómo negocian o confrontan sus puntos de vista con otros actores que tienen principios diferentes o contrarios, y cómo al final estos principios de justicia crean en la red de asociaciones de Latour un actante que surja como elemento fundamental para empujar un anti-anti-programa.

Para finalizar, tanto México como Japón implementan sistemas de reconocimiento facial en sus áreas metropolitanas. Sin embargo, lo hacen de manera diferente: México, en megalópolis como la Ciudad de México, lo implementa bajo el pretexto de prevención del delito, pero únicamente en la Central de Abastos. Por otro lado, Japón utiliza el reconocimiento facial desde una perspectiva de usuario, como mecanismo de abordaje, controles de migración y sistemas de pago en el transporte público, como ya se da en la estación de Osaka (TRAICY Japan Inc, 2023). Además, Japón está probando mecanismos para la prevención del delito, como pruebas de personal de policía equipados con sistemas de reconocimiento facial para proteger a personas importantes, como funcionarios de alto rango

(Nikkei Staff, 2023), y está desarrollando herramientas para retirar dinero de los cajeros mediante esta tecnología (Maeda, 2023).

En este sentido, y con base en la teoría de la red de actores, los mediadores a favor de la implementación de sistemas de reconocimiento facial en México serían las autoridades que buscan fortalecer las medidas de seguridad y los proveedores de tecnología que ven una oportunidad de mercado. Por otro lado, los mediadores en contra podrían incluir a los defensores de la privacidad y los grupos de derechos humanos que temen la invasión a la intimidad y el potencial abuso de poder, pero ¿cuál es la situación en Japón?

En Japón, los mediadores a favor incluyen a las corporaciones tecnológicas y al gobierno, que buscan modernizar y asegurar la infraestructura pública. Por otro lado, los mediadores en contra podrían ser aquellos preocupados por la sobrevigilancia y la erosión de las libertades civiles en una sociedad altamente tecnológica. Finalmente, es relevante destacar que mientras México implementa tecnologías generadas desde el extranjero, Japón utiliza tecnologías locales o desarrolladas por empresas locales. Esto significa que la adopción de tecnologías de reconocimiento facial en Japón no solo refleja un compromiso con la innovación y la seguridad nacional, sino también un impulso hacia la autosuficiencia tecnológica y el fortalecimiento de la economía local; y al confiar en el desarrollo interno, Japón puede tener un mayor control sobre la implementación y la regulación de estas tecnologías, lo que potencialmente podría conducir a un equilibrio más refinado entre la seguridad y la privacidad de los ciudadanos.

Una reflexión final, el uso de los sistemas de reconocimiento facial se expande. Estos sistemas ya no sólo sirven como mecanismos de seguridad sino también como formas de pago y acceso a servicios. No obstante, la aplicación indiscriminada de estos sistemas para ambos casos representa un peligro para la privacidad, acceso a la información, y no discriminación, sobre todo si no se cuentan con actantes o equivalencias que sirvan de contrapeso ante las transgresiones en estos derechos, así como la voluntad organizada de la población afectada.

El papel de los organismos centralizados volvió a ser relevante en 2023 cuando el INAI (2022) ordenó al Aeropuerto Internacional Felipe Ángeles (AIFA), administrado por la Sedena, proporcionar información sobre el sistema E-Gates, diseñado para recopilar datos biométricos de los pasajeros como parte del control de acceso. La instrucción del INAI respondió a la respuesta inicial del AIFA, que declaró inexistente la información sin justificar adecuadamente su búsqueda, lo que llevó a la revocación de dicha respuesta y a la exigencia de una nueva indagación en todas las áreas competentes. En este contexto, el INAI actuó como un actante que buscó estabilizar la red de asociaciones mediante un anti-anti-programa orientado a la protección de los derechos humanos, en particular, el derecho a la privacidad y el acceso a la información.

En suma, en México y Japón, el reconocimiento facial se implementa como un programa para optimizar la seguridad y agilizar procesos en espacios metropolitanos. En México, su uso está orientado principalmente al control policial y de seguridad, como en el C2 y C5, donde las cámaras de vigilancia con reconocimiento facial buscan identificar a personas en espacios públicos, con bases de datos poco transparentes. En Japón, el reconocimiento facial también se integra en sistemas de seguridad, pero su implementación se está inclinando más hacia la eficiencia y la comodidad, como en aeropuertos y estaciones de tren con los E-Gates, o puertas electrónicas, y el Skyliner, donde reemplaza boletos físicos para agilizar el flujo de pasajeros.

Las diferencias en el despliegue de estos sistemas pueden explicarse a través de la relación entre tecnología, sociedad y regulación en cada país. En México, la falta de marcos normativos claros y la opacidad institucional generan un anti-programa en el que la población no tiene acceso a información sobre cómo se almacenan y usan sus datos biométricos, lo que dificulta la posibilidad de resistencia o regulación efectiva que se ha agudizado con la reforma de Simplificación Orgánica (Basham, 2024). En Japón, aunque la tecnología es aceptada con menos oposición, existe un anti-programa basado en la reticencia cultural hacia la adopción apresurada de innovaciones, lo que ha llevado a que el

reconocimiento facial se implemente solo en espacios donde su uso está altamente justificado y alineado con la percepción de seguridad y eficiencia.

El anti-anti-programa en Japón se manifiesta en el equilibrio que las instituciones buscan entre innovación y privacidad, reflejando una tradición animista en la que los objetos tecnológicos no son intrínsecamente problemáticos, sino que deben ser armonizados con el entorno social. En contraste, en México, la tecnología se percibe con mayor desconfianza debido a la historia de vigilancia estatal y la falta de garantías sobre el manejo de datos personales, lo que refuerza una resistencia implícita en la sociedad. A pesar de estas diferencias, en ambos países el reconocimiento facial se enfrenta a desafíos éticos y legales que requieren respuestas institucionales para estabilizar las redes de actores y garantizar el respeto a los derechos fundamentales.

CONCLUSIONES

Esta tesis ha analizado los efectos sociales de la IA de Alphabet-Google, Uber y los sistemas de videovigilancia (C5) en México y Japón, a partir del análisis de los principios éticos que garanticen el respeto y la protección de los derechos humanos en la implementación de la IA. Se ha encontrado que estos sistemas plantean desafíos para la privacidad, el acceso a la información y la no discriminación de las personas, así como para la explicabilidad, la rendición de cuentas y la responsabilidad de los actores involucrados. Tanto México como Japón han desarrollado marcos regulatorios para proteger los datos personales de los usuarios y limitar el poder de las empresas de tecnología, pero que aún existen brechas y áreas de oportunidad que busquen proteger derechos humanos. Es claro que se requiere una mayor participación de la sociedad civil, la academia y los medios de comunicación para generar conciencia y exigir el respeto a los derechos humanos en el contexto de la IA.

Además, el trabajo ilustra cómo la integración de tecnologías de inteligencia artificial puede transgredir derechos humanos fundamentales—como la privacidad, el acceso a la información y la no discriminación—al mismo tiempo que evidencia la utilidad de marcos teóricos como la teoría de actor-red para analizar estas dinámicas. La comparación entre México y Japón aporta a la academia, la ciencia de datos, el derecho y la sociología, ya que revela cómo factores regulatorios, culturales y tecnológicos influyen en el impacto social de la IA en contextos muy diferentes. Además, ofrece un modelo comparativo que puede servir para formular recomendaciones de políticas públicas y mejorar la protección de los derechos fundamentales en un mundo cada vez más digitalizado.

Aunque países como China pueden estar a la vanguardia en el desarrollo tecnológico, comparar México y Japón es crucial para entender las distintas formas en que las sociedades abordan los dilemas éticos de la IA. En México, la ausencia de marcos regulatorios claros y la menor conciencia pública sobre la protección de datos biométricos generan vulnerabilidades, mientras que en Japón

se observa una fuerte integración tecnológica acompañada de una cautela cultural y regulatoria. Este estudio comparativo resalta que la inteligencia artificial y los sistemas de reconocimiento facial no operan en un vacío técnico, sino que están moldeados por el contexto social, político y cultural de cada país, lo que subraya la necesidad de adoptar enfoques contextuales que equilibren la innovación tecnológica con la salvaguarda de los derechos humano.

Los análisis culturales, como el animismo en Japón, ayudan a explicar por qué la aceptación y la resistencia a estas tecnologías varían entre sociedades. En Japón, la visión animista y el Shinto facilitan una integración más orgánica de la IA en la vida cotidiana, mientras que en México, otros factores históricos y normativos influyen en el debate sobre su implementación. Esto no solo enriquece la perspectiva sociológica del estudio, sino que también demuestra que la tecnología no es solo una cuestión de infraestructura y regulación, sino también de valores, creencias y expectativas sociales; ***básicamente revela que cualquier propuesta regulatoria o medidas de mitigación, incluyendo las incluyan actantes con IA, debe de tomar las perspectivas culturales locales: creencias, religión, entre otras.***

En particular, se ha analizado el caso del C5, un sistema de videovigilancia con reconocimiento facial que opera en la Ciudad de México bajo la supervisión de la Secretaría de Seguridad Ciudadana. Se ha evidenciado que el C5 se deslinda de la responsabilidad del uso y la protección de la información captada por las cámaras, y que podría vulnerar los derechos a la privacidad y a la identidad de las personas. Con las respuestas obtenidas en el portal de transparencia, ha quedado de manifiesto que el C5 no ofrece garantías de explicabilidad ni de responsabilidad en caso de fallas o denuncias.

Es importante continuar con investigaciones que analicen el impacto social y la vulnerabilidad de los derechos humanos en el uso de datos biométricos, sobre todo, el del reconocimiento facial en México y Japón, así como en otros países con diferentes contextos culturales, políticos y económicos. Para el caso del C5 es necesario que los gobiernos adopten medidas tecnológicas y generen

regulaciones que fortalezcan la supervisión de esta herramienta tecnológica. Es urgente promover el desarrollo de una IA ética, transparente y humana, que respete la dignidad y la diversidad de las personas.

Esta tesis ha respondido a la pregunta de investigación planteada al inicio: ¿Cómo afectan las plataformas digitales de Alphabet-Google, Uber y los sistemas de videovigilancia (C5) que emplean IA en México y Japón a los derechos humanos relacionados con los principios éticos de transparencia, explicabilidad, rendición de cuentas, responsabilidad, privacidad y protección de datos, en el contexto de las decisiones automatizadas y sus repercusiones sociales a lo largo de todo el ciclo de vida de estos sistemas basados en IA?.

Además, se analizó que estos problemas se deben en gran parte a la falta de regulación, supervisión y participación social en el desarrollo y uso de la IA. De esta forma, se contribuyó al debate académico y social sobre los desafíos éticos y legales de la IA, y se ha propuesto una visión crítica y humanista de la misma. Se espera que esta tesis sirva como un punto de partida para futuras investigaciones y acciones que busquen promover el respeto a los derechos humanos en el contexto de la IA.

También se ha analizado el caso de Uber, una plataforma que utiliza la IA para ofrecer un servicio de transporte urbano innovador y eficiente, pero que genera problemas y conflictos relacionados con los derechos humanos de la IA, especialmente, en temas relativos a la privacidad, la discriminación y la opacidad. Al hacer el comparativo de la situación de Uber en México y Japón, se encontró que la plataforma se enfrenta a diferentes niveles de resistencia y negociación por parte de los actores sociales, como los taxistas tradicionales, las autoridades reguladoras y los usuarios.

Para llevar a cabo este análisis se usó especialmente la teoría del actor-red, en especialmente desde la perspectiva Latour, y en menor medida el modelo de equivalencia de Boltanski y Thévenot para estudiar cómo Uber logra imponer su programa de expansión bajo cualquier costo mediante un proceso de traducción que implica problematizar, interesar, enrolar y movilizar a los actores humanos y

no humanos que participan en su red. También se examinó cómo los actores humanos opuestos al programa de Uber recurren a diferentes principios o mundos de justificación para resistir y criticar la recolección de datos o la generación de sesgos por parte de la IA de Uber.

Asimismo, se analizó el caso de Alphabet Google, una empresa que utiliza la IA para ofrecer diversos servicios digitales a los usuarios, como por ejemplo, el buscador, el correo electrónico, el asistente virtual o el chatbot. De dicho análisis, quedó de manifiesto que Alphabet Google recolecta una gran cantidad de datos personales de los usuarios para vender anuncios personalizados y generar contenido nuevo mediante algoritmos generativos. Así pues, se observó que Alphabet Google podría vulnerar los derechos a la privacidad y al acceso a la información de los usuarios, así como generar sesgos o discriminaciones en sus sistemas y aplicaciones.

Además, se comparó la situación de Alphabet Google en México y Japón, y se encontró que la empresa se enfrenta a diferentes niveles de resistencia y negociación por parte de los actores sociales, como las autoridades reguladoras o las organizaciones en defensa de los derechos humanos. Alphabet Google logra imponer su programa de recolección de datos mediante un proceso de traducción que implica problematizar, interesar, enrolar y movilizar a los actores humanos y no humanos que participan en su red. También se examinó la forma en la que los actores humanos que se oponen al programa de Alphabet Google recurren a diferentes principios o mundos de justificación para resistir y criticar la recolección de datos o la generación de contenido por parte de la IA de Alphabet Google.

Las teorías empleadas y sus herramientas también mostraron la problemática de opacidad y explicabilidad que enfrentan plataformas de video vigilancia con IA como el C5 en la Ciudad de México, y las diferentes cámaras instaladas en los sistemas de metro y tren en Japón. Además, este trabajo de investigación, en torno a las propuestas teóricas metodológicas, demuestra que los actores humanos no son pasivos. Esto también se vio enriquecido con productos de investigación derivados de la tesis; es decir, las bases teóricas y resultados de

este trabajo también sirvieron como fundamento para al menos dos artículos de alcance internacional.

Con respecto al C5 y la teoría empleada, se observa que la IA en el C5 plantea desafíos y oportunidades para la seguridad pública y los derechos humanos. Para comprender mejor estos fenómenos, se propone utilizar las teorías de la sociología de la traducción y la sociología pragmática de la crítica, que permiten analizar las redes de asociaciones entre actores humanos y no humanos, así como las prácticas de crítica y justificación que estos realizan. Así, se podrá identificar los programas, anti-programas y anti-anti-programas que se generan en torno a la IA en el C5, así como los principios o mundos que los sustentan o los cuestionan. Esto podría contribuir a generar propuestas para mejorar el uso ético y responsable de la IA en el ámbito de la seguridad pública.

Y es importante resaltar que en contraste con la noción de Latour, donde el “programa” se refiere a la intención que surge dentro de la red de asociaciones mediante traducciones, y el “anti-programa” es la tensión o resistencia a esa intención, en las plataformas que utilizan sistemas de inteligencia artificial, el anti-programa surge cuando ocurren momentos de decepción o malas prácticas que transgreden la intención original del programa. Este fenómeno lleva a que los usuarios y otros actores dentro de diversas constelaciones (redes relacionadas con las plataformas de IA) generen lo que puede denominarse “anti-anti-programas,” que son respuestas creativas y estratégicas para resistir las fallas o abusos del sistema.

En este contexto, la teoría del actor-red (ANT) aporta un marco clave para comprender cómo las interacciones entre actores humanos y no humanos (en este caso, los algoritmos y sistemas de IA) pueden generar nuevas formas de resistencia y adaptación dentro de redes complejas. Esta dinámica introduce una aportación teórica relevante dentro de ANT, extendiendo su enfoque sobre la agencia distribuida y los procesos de traducción a los ecosistemas de plataformas digitales. Y con respecto a la aportación teórica de este trabajo, esta dinámica se

extiende la idea de resistencia en las redes tecnológicas al proponer nuevas formas de agencia y reacción dentro de los ecosistemas de plataformas digitales.

1. Conclusión primera: ¿Cómo informar a la sociedad?

Uno de los grandes problemas de la IA es la falta de transparencia en torno a su funcionamiento y sus efectos en los derechos humanos, a nivel de asociaciones humanas en torno a la teoría de actor-red, a nivel técnico y sociotécnico. Asimismo, hay una enorme falta de divulgación por parte de los voceros que buscan regular, diseñar, señalar los riesgos, y gestionar mejor las herramientas de los sistemas en torno al público general; en este sentido, los académicos tienen que buscar canales de comunicación, salir de los pasillos de las aulas, y buscar mejores estrategias de comunicación como Latour lo llegó a plantear. Pero, además, se deben promover y proteger organizaciones y organismos autónomos como el INAI que permitan conocer el impacto no sólo de los sistemas de inteligencia, sino también de los algoritmos que los componen.

En este sentido, se debe reconocer el valor de organizaciones como R3D en México o como la Comisión de Protección de Información Personal de Japón que recientemente advirtió que podría tomar medidas relacionadas con la recolección de datos sensibles para el aprendizaje automático, dado que OpenAI, la compañía que creó al famoso ChatGPT, debería reducir los datos que recopila y hacerlo con permiso (IAPP, 2023). Este tipo de organizaciones actúan a modo de contra peso, de voceros que denuncian tanto los problemas en torno a los algoritmos, pero que también generan marcos para un mejor diseño. En este sentido, la pregunta gira en torno a cómo llegar a un mayor público y sobre cómo hacer nuestras estrategias de comunicación más efectivas.

Esto implica responder no sólo a cómo, y con qué lenguaje informar a la sociedad, sino también brindar herramientas y técnicas para protección de datos personales al usar navegadores relativamente seguros como Brave, a utilizar aplicaciones que permitan el cifrado de mensajes como Signal o a configurar las opciones de privacidad de las redes sociales como Facebook o Twitter. Estas medidas pueden

ayudar a prevenir o mitigar los riesgos de la exposición, el robo, la manipulación o el mal uso de los datos personales por parte de los algoritmos de IA.

Sin embargo, estas herramientas y técnicas no son suficientes si no se acompañan de una mayor conciencia y educación sobre la importancia de la protección de datos personales y los derechos humanos en el contexto digital. Por consiguiente, se debe fomentar una cultura de la información y la comunicación que promueva el acceso, la participación, la transparencia, la responsabilidad y la solidaridad entre los diversos actores involucrados en el desarrollo y el uso de la IA. Así pues, se podrá generar un diálogo constructivo y crítico que contribuya a la creación de soluciones innovadoras y éticas para los problemas sociales que plantea la IA.

Los resultados de esta tesis evidencian que en México existe una falta de conciencia sobre los riesgos técnicos y de derechos humanos asociados a la inteligencia artificial, lo que hace urgente el desarrollo de estrategias de comunicación accesibles y adaptadas a distintos sectores de la población. Una primera estrategia es la implementación de campañas de educación pública que expliquen de manera clara y concreta cómo funcionan los sistemas de IA y los riesgos que conllevan, especialmente en temas de privacidad, sesgo algorítmico y vigilancia. Estas campañas deben aprovechar plataformas digitales, redes sociales y medios tradicionales para maximizar su alcance, utilizando ejemplos cercanos a la realidad de los ciudadanos, como el uso de reconocimiento facial en espacios públicos o la toma de decisiones automatizadas en el empleo y los servicios.

Además, es crucial fomentar espacios de diálogo y aprendizaje mediante talleres comunitarios y programas educativos en universidades y centros de trabajo. Estos talleres pueden incluir simulaciones prácticas sobre el uso de datos personales, análisis de casos reales y debates sobre el impacto de la IA en los derechos humanos, permitiendo que los participantes comprendan de manera vivencial los riesgos y beneficios de estas tecnologías. Paralelamente, la publicación de materiales informativos, como guías, estudios de caso y reportes accesibles, contribuiría a consolidar un conocimiento más profundo y crítico en la sociedad.

Además, para garantizar un impacto a largo plazo, es necesario integrar la educación sobre IA y derechos humanos en programas de formación académica y profesional, especialmente en disciplinas como derecho, ciencia de datos y comunicación. La colaboración entre instituciones académicas, organismos gubernamentales y organizaciones de la sociedad civil es clave para generar políticas de alfabetización digital inclusivas y sostenibles. Informar a la sociedad no solo permitirá una mejor comprensión de la IA, sino que también fortalecerá la capacidad de exigir regulaciones y mecanismos de protección adecuados ante el avance tecnológico.

Finalmente, en el caso de México se deben promover la investigación inter y transdisciplinaria sobre qué es la IA, cómo funciona, para qué funciona, cuáles son sus riesgos, cómo implementarla y desarrollarla desde una perspectiva que se alinee con los derechos humanos y una perspectiva con respecto a cómo los seres humanos la pueden aprovechar. En este sentido, es necesario fomentar seminarios, conferencias, publicaciones y espacios de diálogo que reúnan a expertos en derecho, ciencia de datos, sociología, filosofía y otras disciplinas para analizar de manera integral el impacto de la inteligencia artificial en la sociedad. Asimismo, la colaboración entre universidades, el sector privado y organismos gubernamentales puede impulsar el desarrollo de marcos normativos y éticos que orienten el uso responsable de estas tecnologías, ya que solo a través de una aproximación holística y una participación de diversos actores se podrá garantizar que la IA en México no solo sea innovadora, sino también equitativa y respetuosa de los derechos fundamentales.

Para abordar los desafíos que la inteligencia artificial (IA) plantea en términos de derechos humanos y privacidad, es esencial que México promueva la investigación interdisciplinaria que explore qué es la IA, cómo funciona, sus aplicaciones, riesgos y cómo implementarla de manera alineada con los derechos humanos, beneficiando a la sociedad en su conjunto. Esto puede lograrse mediante la organización de seminarios, conferencias y publicaciones que reúnan a expertos en derecho, ciencia de datos, sociología, filosofía y otras disciplinas, permitiendo un análisis integral del impacto de la IA en la sociedad. La

colaboración entre universidades, el sector privado y organismos gubernamentales es crucial para desarrollar marcos normativos y éticos que orienten el uso responsable de estas tecnologías.

En Canadá, la Directiva sobre la Toma de Decisiones Automatizada exige evaluaciones de impacto antes de implementar sistemas de IA en el sector público, asegurando transparencia y equidad en su uso (Treasury Board, 2024). Japón, por su parte, como se mencionó a lo largo de la tesis, ha adoptado una estrategia que combina principios éticos con la promoción de la innovación, reflejada en su Estrategia de IA, que enfatiza la seguridad, inclusión y respeto por los valores culturales del país (Verma, 2024). Finlandia ha lanzado programas educativos como *Elements of AI*, un curso gratuito diseñado para capacitar a la ciudadanía en los fundamentos de la inteligencia artificial (University of Helsinki, 2023).

2. Conclusión segunda: ¿Cómo generar marcos de transparencia y privacidad en torno a los sistemas y algoritmos de IA en México y Japón?

Una pregunta constante, más allá del entorno regulatorio, es saber cómo generamos un mayor marco de transparencia y de privacidad en torno a los sistemas y algoritmos de IA. Para responder esta cuestión, se debe reconocer de nueva cuenta que esta tecnología genera transgresiones a nivel técnico, sociotécnico, de derechos humanos y de asociaciones humanas en torno a la teoría de actor-red. Así pues, los mecanismos de transparencia se deben crear en torno a estos problemas.

Con respecto al nivel técnico es necesario garantizar que los sistemas y algoritmos de IA sean auditables, explicables y verificables, es decir, que se pueda acceder, comprender y evaluar su funcionamiento, su lógica, sus criterios y sus resultados. Para ello, se puede recurrir a estándares, protocolos, certificaciones y herramientas que faciliten la inspección, la documentación y la validación de los sistemas y algoritmos de IA. En esta línea, el AI Act del Parlamento Europeo propone mecanismos regulatorios que obligan a las empresas a validar sus

sistemas de IA mediante evaluaciones previas y continuas, con el fin de verificar su conformidad con criterios de seguridad, transparencia, explicabilidad y respeto a los derechos fundamentales.

También se incentiva el desarrollo de una IA más holgada que no requiera muchos datos para su entrenamiento. Además, se puede emplear técnicas que fomenten la privacidad en la etapa de entrenamiento de los algoritmos como lo es el aprendizaje de máquina federado. Por ejemplo, Qammar et al (2022), explican como un paradigma de aprendizaje automático distribuido que entrena el modelo con datos locales de los dispositivos garantizando la privacidad (McMahan et al. 2017 en Qammer et al); es importante destacar que esta técnica permite que el modelo se entrene directamente en los dispositivos locales de los usuarios, en lugar de recopilar grandes cantidades de datos en un servidor centralizado; así pues, el aprendizaje ocurre de manera distribuida, lo que significa que los datos personales nunca abandonan los dispositivos de origen.

Dentro del aspecto técnico, también se pueden diseñar plataformas contra algunos sistemas sumamente invasivos como lo son el reconocimiento facial. Esto conlleva involucrar a la población para que conozcan estos sistemas, tengan acceso a ellos y sepan cómo usarlos. Por ejemplo, en el caso de los sistemas de reconocimiento facial, resaltar contrapesos técnicos tales como las herramientas de anti reconocimiento facial que entre otras se basan en atacar para interrumpir la recopilación de datos, impedir el raspado de imágenes en línea, evitar la captura de imágenes, ocultar rostros, interrumpir la cámara; atacar para alterar el preprocesamiento facial, modificar directamente imágenes digitales, usar objetos físicos diseñados a medida; Anonimización de Rostros, ataque a bases de datos corruptas, entre otros (Wenger et al, 2022); otra serie de técnicas se dirige a interrumpir el preprocesamiento facial, es decir, las etapas iniciales de análisis que los algoritmos de reconocimiento facial emplean para interpretar y clasificar características faciales lo cual puede lograrse mediante la modificación directa de imágenes digitales o el uso de objetos físicos especialmente diseñados para confundir al sistema, como gafas, maquillaje o ropa que interfiera con la detección de rasgos faciales.

Esto también está relacionado con las mejores prácticas de diseño que tiene un impacto fuerte en torno a los derechos humanos. En este sentido, Toxtli et al. (2021) sugieren que, de manera similar, los algoritmos que facilitan la asignación automática de tareas y las recomendaciones podrían ser útiles para reducir el trabajo invisible al minimizar el tiempo de búsqueda de tareas. Además, se pueden desarrollar estrategias técnicas que ofrezcan contrapesos a los modelos de vigilancia en las plataformas, como proponen Moreno et al. (2024).

A nivel sociotécnico, se debe asegurar que los sistemas y algoritmos de IA sean responsables, éticos y participativos. Que se ajusten a los principios, valores y normas que rigen la convivencia social, que respeten los derechos humanos e involucren a los actores interesados en su diseño, implementación y uso. Para ello, se puede recurrir a marcos, códigos, guías y mecanismos que orienten, regulen y supervisen el desarrollo y el uso de los sistemas y algoritmos de IA; en este caso, vale la pena destacar que incluso países con fama de autoritarios están intentando regular algunos sistemas como los de reconocimiento facial (Costigan, 2023); por ejemplo, China, conocida por su amplio uso de sistemas de vigilancia, ha implementado regulaciones que exigen la transparencia y el consentimiento en la recopilación de datos faciales en ciertas situaciones, como en la industria privada o en entornos no gubernamentales; además, se puede enseñar a la población a implementar ataques adversarios no técnicos con el objetivo de proteger sus datos biométricos de los sistemas de reconocimiento facial (Domingo, 2022; Contreras y Soto, 2023).

A nivel de derechos humanos, se debe proteger que los sistemas y algoritmos de IA sean equitativos, inclusivos y no discriminatorios. Esto significa que se deben diseñar y gestionar sistemas que no reproduzcan ni generen sesgos, desigualdades o exclusiones basados en el género, la raza, la etnia, la edad, la discapacidad, la orientación sexual, la religión o cualquier otra condición. Para ello, se puede recurrir a políticas, leyes, acciones y recursos que prevengan, detecten y corrijan las violaciones o amenazas a los derechos humanos por parte de los sistemas y algoritmos de IA; y entre estas políticas se puede promover un derecho

a la desindexación, para que como el caso jurídico en torno a Alphabet-Google, una persona puede optar por desaparecer de los sistemas digitales.

Con respecto a las leyes hay considerar iniciativas puntuales como la Sociedad 5.0 de Japón busca integrar tecnologías avanzadas, como la inteligencia artificial y la robótica, en todos los aspectos de la vida diaria para crear una sociedad más eficiente, inclusiva y centrada en el bienestar humano; es importante resaltar que esta propuesta se enfoca en equilibrar la innovación tecnológica con regulaciones que aseguren la protección de los derechos fundamentales, como la privacidad y la equidad en el acceso a estas tecnologías. Pero actualmente en México, más allá del papel importante que tiene mecanismos como el INAI, también existen propuestas de iniciativas y acciones como la Alianza Nacional de la IA (ANAI) que busca estudiar los desafíos, ventajas e implicaciones del empleo de la IA para la elaboración de políticas públicas, normativa y modelo de gobernabilidad. (ANAI, 2023). Esta asociación, además, cuenta con una gran variedad de políticos, legisladores, y académicos investigadores que estudian y fomentan una IA confiable.

En el plano de asociaciones humanas en torno a la teoría del actor-red, se debe promover que los sistemas y algoritmos de IA sean colaborativos, transparentes y solidarios, o sea, que contribuyan a crear redes de cooperación, comunicación y apoyo entre actores humanos y no humanos que interactúan con esta tecnología. En este sentido, se pueden recurrir a iniciativas, proyectos, plataformas y redes que fomenten el intercambio, la difusión y la generación de conocimiento sobre los sistemas y algoritmos de IA. Además, se puede involucrar al público en torno medidas educativas sobre cómo funcionan los sistemas, qué pueden hacer para protegerse; pero también en torno a medidas de la gestión y diseño de estos sistemas.

Asimismo, es necesario resaltar que la teoría de actor-red demuestra que las relaciones entre la IA y los humanos no son pasivas; es decir, los humanos suelen resistir o aprovecharse de las acciones de los algoritmos de forma activa y consciente, en lugar de simplemente someterse a ellos; por ejemplo, los actores

humanos pueden usar atajos o trucos para mostrar resistencia o beneficio frente al programa (Morton, 2023). Así pues, en el caso de México y Japón, se pueden observar diferentes formas de apropiación, uso y regulación de la IA, que reflejan sus respectivas culturas, valores y contextos como el hecho de que México tenga una legislación menos avanzada y que sólo esté en propuestas; mientras que Japón no sólo tiene propuestas legislativas robustas, sino también un mapa con respecto a la IA en torno a sus valores y problemas sociales. No obstante, ambos países se pueden beneficiar de estas propuestas de resistencia en torno a las transgresiones a nivel de derechos humanos, técnico, sociotécnico, y de asociaciones humanas en torno a las teorías de actor-red al tener en cuenta que los actores humanos no son pasivos y sólo requieren de conocimientos, educación y herramientas para mostrar resistencia.

El primer paso para México en la protección de la privacidad y el derecho a la información en la era de la inteligencia artificial debe ser la restitución y fortalecimiento de organismos autónomos como el INAI. No solo es necesario recuperar su función original, sino ampliar sus facultades para incluir la regulación de datos biométricos, la detección y mitigación de sesgos algorítmicos, y la supervisión de sistemas de IA que puedan afectar derechos fundamentales. Para garantizar un uso responsable de estas tecnologías, este organismo debe contar con plena autonomía y autoridad para regular, evaluar y sancionar prácticas que vulneren la privacidad y la equidad en la toma de decisiones automatizadas.

Finalmente, y considerando las características de cada país, los marcos regulatorios, y transparencia se pueden organizar de esta manera; para el caso de México:

- Fortalecimiento de organismos autónomos: Como se mencionó más arriba, es imperativo revitalizar instituciones como el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI) o en un futuro crear otro organismo autónomo que supla las funciones del INAI. Este organismo debería tener competencias específicas para regular la protección de datos biométricos y abordar los desafíos

únicos que presenta la IA, garantizando así la privacidad y los derechos de los ciudadanos.

- Implementación de leyes y regulaciones específicas: Desarrollar y promulgar una legislación integral que aborde el uso ético y transparente de la IA. Esta normativa debe establecer directrices claras sobre la recopilación, almacenamiento y procesamiento de datos, asegurando que las prácticas de IA respeten los derechos humanos y la privacidad
- Creación de códigos de ética y mecanismos de supervisión: Elaborar códigos de conducta que orienten a las empresas y entidades gubernamentales en el desarrollo y aplicación responsable de la IA. Además, establecer organismos de supervisión independientes encargados de monitorear el cumplimiento de estas directrices y de sancionar posibles infracciones.

Japón:

- Actualización y fortalecimiento de la Ley de Protección de la Información Personal (APPI): Aunque Japón cuenta con la APPI, es esencial revisarla periódicamente para adaptarla a los avances tecnológicos y a los nuevos desafíos que plantea la IA. Esto incluye la regulación específica de datos biométricos y la implementación de salvaguardas contra sesgos algorítmicos.
- Fomento de la auto-regulación y creación de códigos de ética sectoriales: Dada la cultura corporativa japonesa, promover que las industrias desarrollen sus propios códigos de ética en relación con la IA puede ser efectivo. Estos códigos deben alinearse con los principios nacionales y enfatizar la transparencia, la equidad y el respeto por la privacidad.
- Establecimiento de mecanismos de supervisión y auditoría: Crear organismos especializados que supervisen el uso de la IA en diferentes

sectores, asegurando el cumplimiento de las normativas y evaluando continuamente el impacto de la IA en la sociedad.

Finalmente, es necesario valorar las consideraciones culturales y contextuales de cada país:

México: La diversidad cultural y la desigualdad socioeconómica requieren campañas de concienciación pública que informen a la ciudadanía sobre sus derechos en relación con la IA. Además, es crucial garantizar que las regulaciones sean inclusivas y consideren las particularidades de las comunidades indígenas y otros grupos vulnerables.

Japón: La sociedad japonesa valora la armonía y la confianza en las instituciones. Por lo tanto, es vital que las políticas relacionadas con la IA sean transparentes y cuenten con la participación de la comunidad, reforzando la confianza pública en estas tecnologías.

3. Conclusión tercera: ¿Cómo armonizar la implementación y desarrollo de la IA más allá de un modelo para prevenir transgresiones?

Es fundamental armonizar la implementación y el desarrollo de la inteligencia artificial más allá de un enfoque meramente preventivo de transgresiones. Esto requiere una perspectiva teórica, social y jurídica que permita gestionar a todos los actantes dentro de la red de asociaciones, considerando sus necesidades y resistencias. De esta manera, es posible agilizar un proceso de interpretación regulatoria flexible, donde la tecnología se integre como una parte de la sociedad y no como un elemento que la supere o imponga sus propias dinámicas.

En este sentido, armonizar implica alinear propuestas, modelos de desarrollo e investigación en IA con marcos jurídicos a nivel internacional, regional y cultural. A nivel regulatorio, esto significa coordinar normativas, leyes y estándares para que sean compatibles entre distintos sistemas jurídicos, sectores o países. Esto puede traducirse en la adopción de marcos regulatorios comunes, la homologación de

criterios técnicos o la adaptación de legislaciones nacionales a estándares internacionales, evitando conflictos normativos y garantizando una aplicación coherente de las reglas.

Desde una perspectiva cultural, armonizar implica integrar valores, percepciones y prácticas sociales en la adopción de nuevas normativas o tecnologías. En el caso de la IA, esto supone diseñar regulaciones que sean comprendidas y aceptadas dentro del marco cultural de cada sociedad, considerando factores como la percepción del riesgo, la confianza en las instituciones y las creencias sobre la relación entre humanos y tecnología. Por ejemplo, en Japón, la integración de la IA puede verse favorecida por el animismo tecnológico del Shinto, mientras que en México, las regulaciones deben responder a preocupaciones sobre desigualdad y transparencia institucional, asegurando que la tecnología no profundice brechas existentes.

La IA no es un ente aislado; forma parte de una red de asociaciones en la que cada implementación introduce un programa con expectativas sobre su desempeño. Sin embargo, debido a problemas técnicos como sesgos en los datos, errores humanos y efectos no previstos, la IA ha generado transgresiones en torno a derechos fundamentales como la transparencia, la equidad, la responsabilidad y la privacidad. A raíz de estas problemáticas, la mayoría de los debates regulatorios han estado dominados por un enfoque correctivo, centrado en prevenir daños desde una ética mayoritariamente occidental (Jecker & Nakazawa, 2022).

Hacia un marco de armonización en la regulación de la IA: Como se expuso en el primer capítulo, es crucial desarrollar un marco regulatorio complementario que no solo se enfoque en las transgresiones, sino que también incorpore valores positivos que guíen el desarrollo de la IA. McStay (2021) propone que la combinación de filosofías japonesas con enfoques modernos puede aportar una sensibilidad ética y política valiosa en la relación entre humanos y sistemas de IA. Esto permitiría no solo comparar modelos regulatorios, sino también diseñar marcos de implementación basados en una visión de armonía.

En este sentido, Contreras y Soto (2023) identifican tres modelos regulatorios clave en Japón que pueden servir de referencia:

1. Autorregulación, donde las empresas establecen sus propias normas y principios éticos.
2. Regulación adaptada a normativas preexistentes, como las leyes de protección de datos personales.
3. Regulación específica para sistemas de IA, que atiende sus impactos y riesgos concretos.

Además, la Estrategia de IA 2022 del gobierno japonés enfatiza la seguridad, la armonía con la naturaleza y el equilibrio entre innovación y bienestar social (CAO, 2022). Aunque estas ideas pueden parecer ajenas a un manual técnico sobre IA, reflejan un enfoque regulatorio basado en la convivencia y la integración tecnológica dentro de un ecosistema social y natural.

Lecciones para México. Hacia una regulación con armonía: ¿Cómo integrar estas perspectivas en un contexto como el mexicano?

- Incorporar valores de armonización junto con los de prevención. Como sugieren Jecker y Nakazawa (2022), además de valores como transparencia y justicia para mitigar riesgos, es clave promover principios como beneficencia, autonomía, confianza, sostenibilidad, dignidad y solidaridad en la regulación de la IA.
- Adaptar estos valores a las realidades locales. Por ejemplo, el derecho a la transparencia puede ejercerse a través del Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI), pero esta institución también podría asumir un papel más proactivo en fomentar un uso responsable de la IA. Asimismo, iniciativas como la Alianza Nacional de Inteligencia Artificial en México (AINA) pueden contribuir a consolidar estrategias nacionales.
- Garantizar el acceso a la información y fortalecer organismos de supervisión. Un ejemplo concreto de los desafíos actuales es la dificultad

para obtener información sobre el uso de IA en cámaras de vigilancia del C5, donde las respuestas han sido confusas y poco claras. La desaparición del INAI implicaría un retroceso en la protección de derechos fundamentales, lo que subraya la necesidad de organismos que aseguren la transparencia en el desarrollo e implementación de IA.

Hacia una IA que beneficie a la sociedad mexicana: El objetivo final debe ser que la IA mejore la calidad de vida, la educación, la salud y la seguridad, respetando los derechos, necesidades y aspiraciones de los ciudadanos. También es clave fomentar la IA como herramienta de expresión artística, diversidad e inclusión, reconociendo la pluralidad de voces, lenguas y tradiciones del país. Además, debe fortalecerse el uso de la IA para potenciar sectores estratégicos e impulsar la competitividad industrial de México, como ha planteado la Secretaría de Economía y Tecnología (CAO, 2022).

Finalmente, esta tesis ha demostrado que las relaciones de poder se manifiestan a través de asociaciones (Morton, 2023; 2024). Esto significa que la red de IA no es neutral: sus modelos y sistemas tienen intenciones y consecuencias que pueden transgredir derechos humanos fundamentales. Sin embargo, con herramientas teóricas y metodológicas adecuadas, es posible desarrollar políticas que hagan de la IA una tecnología ética, equitativa y en sintonía con las necesidades locales y globales.

En definitiva, armonizar la IA no es solo cuestión de regulación, sino de integración cultural, social y ética. México tiene la oportunidad de aprender de modelos internacionales y, al mismo tiempo, construir un marco propio que garantice que la IA funcione como un aliado del bienestar colectivo y no como un factor de desigualdad o exclusión. Esto tomando en cuenta la red asociaciones, resistencias, y perspectivas regulatorias internacionales y culturales locales.

Finalmente, sin importar cuántas propuestas regulatorias se presenten, es la sociedad civil la que debe luchar activamente por sus derechos. Como se observó en los casos de estudio, las organizaciones solo reaccionan cuando los individuos o la sociedad presionan por el reconocimiento y la defensa de sus derechos.

Desde nuestro campo, resulta fundamental no solo promover la divulgación y el debate público, sino también formular propuestas regulatorias sólidas que contribuyan a la protección de los derechos y al desarrollo de marcos normativos más justos y eficaces.

Uso de la IA generativa en la tesis

En esta tesis, se emplearon modelos de lenguaje de gran escala (LLMs), como ChatGPT y Qwen, exclusivamente para correcciones de estilo, revisión gramatical y verificación de referencias; así como modelos generativos, como DALL·E 3, para la creación de la portada. En ningún caso se utilizaron estas herramientas para la redacción o generación de párrafos completos, ni para la creación de referencias. No obstante, con el fin de garantizar el rigor y la integridad académica, es importante reconocer tanto el uso de estas herramientas como la forma en que fueron aplicadas.

FUENTES DE REFERENCIA

1. Bibliografía

ABITIA, E. A. (2020, 23 junio). *Valores de los mexicanos*. Este País.
<https://estepais.com/impreso/valores-de-los-mexicanos/>

ADAME GODDARD, J. (1983). *Historia Del Derecho Mexicano*. Instituto de Investigaciones Jurídicas, UNAM.
<https://archivos.juridicas.unam.mx/www/bjv/libros/1/320/3.pdf>

AGAR, J. (2020). "What is science for? The Lighthill report on artificial intelligence reinterpreted". *The British Journal for the History of Science*, 53(3), 289-310.
doi:10.1017/S0007087420000230

ALBRIEU, R., RAPETTI, M., et. al. (2018, diciembre). "Inteligencia artificial y crecimiento económico. Oportunidades y desafíos para México" CIPPEC, (nº 1).
<https://news.microsoft.com/uploads/prod/sites/41/2018/11/IA-y-Crecimiento-MEXICO.pdf>

ALMEIDA, D., SHMARKO, K., & LOMAS, E. (2021). "The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks". In *AI and Ethics* (Vol. 2, Issue 3, pp. 377–387). Springer Science and Business Media LLC. <https://doi.org/10.1007/s43681-021-00077-w>

Álvarez Ledesma, M. I. (2023). *Derechos humanos: Una visión multidimensional* (Primera edición). McGraw Hill Interamericana Editores.

Alvarez-Melis, D., & Jaakkola, T. S. (2018). Towards Robust Interpretability with Self-Explaining Neural Networks. Recuperado de <https://arxiv.org/abs/1806.07538>

ALLISON, A. (2006). *Millennial Monsters*. Amsterdam University Press.

Amnistía Internacional. (2022). *Surveillance Giants: How the Business Model of Google and Facebook Threatens Human Rights*. Amnistía Internacional.

- ASAMBLEA GENERAL DE LAS NACIONES UNIDAS. (1948, 10 diciembre). *La Declaración Universal de Derechos Humanos*, ONU. Recuperado 23 de enero de 2022, de <https://www.un.org/es/about-us/universal-declaration-of-human-rights>
- Asís, R. de (2020, abril 1–4). “Inteligencia artificial y derechos humanos”, *Seminario Permanente Gregorio Peces-Barba [Presentación]*. Madrid, España.
- ATKINSON, W. (2019). “Luc Boltanski’s pragmatic sociology: A Bourdieusian critique”. *European Journal of Social Theory*, 23(3), 310-327. <https://doi.org/10.1177/1368431019855004>
- BAILEY, B. & ANDERSON, T. (2005). *Taxonomies for the Development and Verification of Digital Systems* (2005 ed.). Springer.
- BAKATOR, M. & RADOSAV, D. (2018). “Deep Learning and Medical Diagnosis: A Review of Literature”. *Multimodal Technologies and Interaction*, 2(3), 47-55. <https://doi.org/10.3390/mti2030047>
- BALDWIN, C. Y., & WOODARD, C. J. (2008). “The Architecture of Platforms: A Unified View”. *SSRN Electronic Journal*, 24(3), 373–390. <https://doi.org/10.2139/ssrn.1265155>
- Baños-Gonzalez, V. (2024). La inteligencia artificial, estudio de su evolución y aplicación en México. *Pädi*, 12(Esp. 4), 250-260.
- Bareebe, R. (2022). The Cambridge Analytica scandal and its impact on Meta. <https://doi.org/10.13140/RG.2.2.19583.69285>
- Bansal, C., Pandey, K. K., Goel, R., Sharma, A., & Jangirala, S. (2023). Artificial intelligence (AI) bias impacts: Classification framework for effective mitigation. *Issues in Information Systems*, 24(4), 367–389. https://doi.org/10.48009/4_iis_2023_127
- BATHAE, Y. (2018). “The artificial intelligence black box and the failure of intent and causation”. *Harvard Journal of Law & Technology*, 31(2), 889–938.
- BECK, U. (2002). *La Sociedad del Riesgo Global*. Paidós.

- BELL, J. A., KOBAC, B., et. al. (2018b). "Unseen Connections: The Materiality of Cell Phones". *Anthropological Quarterly*, 91(2), 465–484.
<https://doi.org/10.1353/anq.2018.0023>
- BEYEBACH, M., NEIPP, M., et. al. (2021). „Bibliometric differences between WEIRD and Non-WEIRD countries in the outcome research on Solution-Focused brief Therapy". *Frontiers in Psychology*, 12.
<https://doi.org/10.3389/fpsyg.2021.754885>
- BILBENY, N. (2012). *Ética* (1a ed). Editorial Ariel.
- Bhumarkar, M., Nimgade, T., Shaikh, K., Rezaie, M., & Dongardive, J. (2024). The impacts of the use of AI in surveillance and the privacy rights of people. *International Journal of Creative Research Thoughts (IJCRT)*, 12(7), Article IJCRT2407640. <https://www.ijcrt.org>
- BLACKMAN, R. (2022). *Ethical machines: Your Concise Guide to Totally Unbiased, Transparent, and Respectful AI*. Harvard Business Press.
- Boden, M. A. (2018). *Artificial Intelligence: A Very Short Introduction*. Oxford University Press.
- BOLTANSKI, L. (2012). *Énigmes et complots: Une enquête à propos d'enquêtes*. Gallimard.
- BOLTANSKI, L., & THÉVENOT, L. (1999). "The Sociology of Critical Capacity". *European Journal of Social Theory*, 2(3), 359–377.
<https://doi.org/10.1177/136843199002003010>
- BOUDIA, S., & JAS, N. (2007). "Introduction: Risk and 'Risk Society' in Historical Perspective". *History and Technology*, 23(4), 317–331.
<https://doi.org/10.1080/07341510701527393>
- BRÖMMELSTROET, M. T, NIKOLAEVA, A., et. al. (2021). "Have a good trip! Expanding our concepts of the quality of everyday travelling with flow theory". *En Applied*

- Mobilities* (pp. 1–22). Informa UK Limited.
<https://doi.org/10.1080/23800127.2021.1912947>
- BROWNE, Kieran. (2019). *Artificially neural networks: deep learning's origin in statistics*. 10.13140/RG.2.2.20066.48325.
- BROWNLEE, J. (2019, 10 November). “14 Different Types of Learning in Machine Learning”. *Machine Learning Mastery*.
<https://machinelearningmastery.com/types-of-learning-in-machine-learning/>
- BRUSSEAU, J. (2022, December 4). “Acceleration AI Ethics, the Debate between Innovation and Safety, and Stability AI’s Diffusion versus OpenAI’s Dall-E”. *arXiv.org*. Retrieved from <http://arxiv.org/abs/2212.01834>
- BUGHIN, J., CHUI, M., & MANYIKA, J. (2018). *Artificial intelligence: The next digital frontier?* McKinsey & Company.
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, 1–15.
- CABINET OFFICE GOVERNMENT OF JAPAN. (2022). *AI strategy 2022 (overview)*. From https://www8.cao.go.jp/cstp/ai/aistratagy2022en_ov.pdf
- CABINET OFFICE GOVERNMENT OF JAPAN. (2022). *Artificial Intelligence Strategy 2022: Toward the Realization of a Human-Centric Society*. From: <https://www8.cao.go.jp/cstp/ai/aistratagy2022en.pdf>
- CALLON, M. (1984). “Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay”. *The Sociological Review*, 32(1_suppl), 196–233. <https://doi.org/10.1111/j.1467-954X.1984.tb00113.x>
- CALLON, M. (1986). “Éléments pour une sociologie de la traduction: La domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc”. *L’Année sociologique*, 36, 169-208.

- CAMARENA, M., & TUNAL, G. (2009). "La religión como una dimensión de la cultura". *Nómadas* (nº 22) Universidad Complutense de Madrid, 1-15.
- CAPONE, L., & BERTOLASO, M. (2022). "Reflections on a Theory of Artificial Intelligence". *Digital Humanism*, 13–29. https://doi.org/10.1007/978-3-030-97054-3_2
- CÁRDENAS GRACIA, J. F. (2010). *Introducción al estudio del derecho*. Instituto de Investigaciones Jurídicas, UNAM.
- CARLUCCI AIELLO, L. (2016). "The multifaceted impact of Ada Lovelace in the digital age". *Artificial Intelligence*, 235, 58–62. <https://doi.org/10.1016/j.artint.2016.02.003>
- CEJUDO, G. (2019). *Diccionario de transparencia y acceso a la información pública*.
- Chesney, R., & Citron, D. K. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, 107(6), 1753–1819.
- Clark, J., Maslej, N., Brynjolfsson, E., Etchemendy, J., Niebles, J. C., Parli, V., Perrault, R., & Fattorini, L. (2023). Artificial Intelligence Index Report 2023 . Stanford University, Human-Centered Artificial Intelligence (HAI). https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf
- Climent, R. C., Haftor, D. M., & Staniewski, M. W. (2024). AI-enabled business models for competitive advantage. *Journal of Innovation & Knowledge*, 9(3), 100532. <https://doi.org/10.1016/j.jik.2024.100532>
- CMIEL, K. (2004). "The Recent History of Human Rights". *The American Historical Review*, 109(1), 117–135. <https://doi.org/10.1086/530153>
- Coalición IA2030Mx. (2020). *Agenda Nacional Mexicana de Inteligencia Artificial*. IA2030Mx. https://wp.oecd.ai/app/uploads/2022/01/Mexico_Agenda_Nacional_Mexicana_de_IA_2030.pdf

COMISIÓN EUROPEA (2020). *Libro Blanco de la Inteligencia Artificial*. Recuperado de https://ec.europa.eu/info/publications/artificial-intelligence-white-paper-fostering-trustworthy-ai_en

COMISIÓN NACIONAL DE DERECHOS HUMANOS. (2021, abril). CDHCM alerta sobre implicaciones de privacidad y datos personales del Padrón Nacional de Usuarios de Telefonía Móvil (N.o 70/2021). www.cdHCM.org.mx. <https://cdHCM.org.mx/2021/04/cdHCM-alerta-sobre-implicaciones-de-privacidad-y-datos-personales-del-padron-nacional-de-usuarios-de-telefonia-movil/>

COMISIÓN NACIONAL DE LOS DERECHOS HUMANOS (CNDH). (2020). *Acceso a la información*. Recuperado de <https://www.cndh.org.mx/acciones-y-programas/derechos-humanos/acceso-a-la-informacion>

CONSTITUCIÓN POLÍTICA DE LOS ESTADOS UNIDOS MEXICANOS. (1917). Recuperado de http://www.diputados.gob.mx/LeyesBiblio/pdf/1_24.pdf

CONTRERAS, P., & SOTO, K. (2023). "Reconocimiento facial, ética e inteligencia artificial". en *Introducción a la ética y el derecho de la inteligencia artificial* (1st ed., Vol. 1).

CURRIE, G., & ROHREN, E. (2021). "Intelligent Imaging in Nuclear Medicine: the Principles of Artificial Intelligence, Machine Learning and Deep Learning". *Seminars in Nuclear Medicine*, 51(2), 102–111. <https://doi.org/10.1053/j.semnuclmed.2020.08.002>

Chau, B. K. (2023). Governing the algorithmic turn: Lyft, Uber, and disparate impact. Social Science Research Network. <https://doi.org/10.2139/ssrn.4605607>

CHARMAN-ANDERSON, S. (2020). "Ada Lovelace: A Simple Solution to a Lengthy Controversy". *Patterns*, 1(7), 100118. <https://doi.org/10.1016/j.patter.2020.100118>

CHEN, J. Y., & QIU, J. L. (2019). "Digital utility: Datafication, regulation, labor, and DiDi's platformization of urban transport in China". *Chinese Journal of* 456

Communication, 12(3), 274–289.
<https://doi.org/10.1080/17544750.2019.1614964>

CHEN, Y., MANCINI, M., et. Al. (2022). *Semi-Supervised and Unsupervised Deep Visual Learning: A Survey*. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Covington (2023). Uber-CRA Report August 2023 [PDF file]. Retrieved from https://s23.q4cdn.com/407969754/files/doc_governance/2023/Uber-CRA-Report-August-2023.pdf

D'HOTMAN, D., & LOH, E. (2020). "AI enabled suicide prediction tools: a qualitative narrative review". *BMJ Health & Care Informatics*, 27(3), e100175. <https://doi.org/10.1136/bmjhci-2020-100175>

DEL POZO, C. M., FERREIRA, R., et. al. (2020). "Ética". *Agenda Nacional Mexicana de Inteligencia Artificial*. 45-76. México: IA2030Mx.

DEL POZO, C. M., GÓMEZ MONT, C., et. al. *Agenda Nacional de IA de México*. México: IA2030Mx.

De los Santos, M. D. L., Do, K., Muller, M., & Savage, S. (2024). Designing sousveillance tools for gig workers. <https://doi.org/10.1145/3613904.3642614>

DENG, Y. (2019). *Deep Learning on Mobile Devices-A Review*. En arXiv:1904.09274v1

DERROW-PINION, A., SHE, J., et. al. (2021). *ETA Prediction with Graph Neural Networks in Google Maps. Proceedings of the 30th ACM International Conference on Information; Knowledge Management*. <https://doi.org/10.1145/3459637.3481916>.

DESCARTES, R., KENNINGTON, R., et. al. (2012). *Discourse on Method* (vol. 1). Van Haren Publishing.

DIJCK, V. J., POELL, T., et. al. (2018). *The Platform Society: Public Values in a Connective World (English Edition)*. Oxford University Press.

DIRKSEN, N., & TAKAHASHI, S. (2020, october). *Artificial intelligence in Japan 2020: Actors, market, opportunities and digital solutions in a newly transformed world*. Netherlands Enterprise Agency is part of the ministry of Economic Affairs and Climate Policy.
<https://www.rvo.nl/sites/default/files/2020/12/Artificial-Intelligence-in-Japan-final-IAN.pdf>.

DITTMAR, R & WU, A (s.f). *El Abogado del Futuro: Legaltech y la Transformación Digital del Derecho*. [MOOC]. Coursera.
<https://www.coursera.org/learn/legaltech/lGillies>.

DKE GERMAN COMMISSION FOR ELECTRICAL, ELECTRONIC & INFORMATION TECHNOLOGIES OF DIN AND VDE, Wahlster, W., & Winterhalter, C. (2020, November). Germanu Standardization Roadmap AI: Launch of the second edition (N.o 1). DKE.
<https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsroadmap-en-data.pdf>

DOMINGO, C. (2022). “Aplicación del sistema de reconocimiento facial para prevenir la violencia asociada al deporte en los encuentros calificados de alto riesgo”. *Inteligencia artificial legal y administración de justicia* (1st ed., Vol. 1).

DURAND-LOSE, J., & MARGENSTERN, M. (2007). *Machines, Computations, and Universality: 5th International Conference*, MCU 2007 Orleans, France, September 10–13, 2007 Proceedings: 4664 (2007 ed.). Springer.

E Silva, N. (2024). The Artificial Intelligence Act: Critical Overview.
<https://doi.org/10.2139/ssrn.4937150>

EPSTEIN, B. (2015). *The Ant Trap: Rebuilding the Foundations of the Social Sciences (Illustrated)*. Oxford University Press, USA.

ERTEL, W. (2019). *Artificial Intelligence and Society*. 15. 20-25.

ETICAS FOUNDATION. (2021, 14 septiembre). *Sistema de crédito social chino*. Eticas Foundation. Recuperado 9 de marzo de 2022, de <https://eticasfoundation.org/sistema-de-credito-social-chino/>

- ESTILOGRAFICA. (2024). La inteligencia artificial como herramienta impulsora de la economía de la comunidad indígena de Guasave, Sinaloa, México: el diagnóstico. CIECEM 2024.
https://ciecem.org/ponencia/la-inteligencia-artificial-como-herramienta-impulsora-de-la-economia-de-la-comunidad-indigena-de-guasave-sinaloa-mexico-el-diagnostico/?utm_source=chatgpt.com
- Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.
- FELDHOFF, B., STOCKMANN, N., et. al. (2019). "Bridging Theories and Practices: Boundary Objects and Constellation Analysis as Vehicles for Interdisciplinary Knowledge Integration". *Sustainability*, 11(19), 53–57.
<https://doi.org/10.3390/su11195357>
- FERNÁNDEZ MATEO, J. (2020). "Los fundamentos epistemológicos de la transformación digital y sus efectos sobre la Agenda 2030 y los derechos humanos". *Revista Icade. Revista de las Facultades de Derecho y Ciencias Económicas y Empresariales*, 108.
<https://doi.org/10.14422/icade.i108.y2019.004>
- FERRUCCI, D., BROWN, E., et. al. (2010). "Building Watson: An Overview of the DeepQA Project". *AI Magazine*, 31(3), 59.
<https://doi.org/10.1609/aimag.v31i3.2303>
- FIGUEROA RODRÍGUEZ, K. A., FIGUEROA SANDOVAL, B., et. al. (2012). "Análisis de los valores que construyen la identidad del mexicano". *Culturales*, VIII (16), 7-32.
- FLORIDI, L., COWLS, J., et. al. (2020). "How to design AI for social Good: Seven essential factors". *Science and Engineering Ethics*, 26 (3), 1771-1796.
<https://doi.org/10.1007/s11948-020-00213-5>
- Fox, N. J. (2011). "Boundary Objects, Social Meanings and the Success of New Technologies". *Sociology*, 45(1), 70–85.
<https://doi.org/10.1177/0038038510387196>

- FREUND, Y., & SCHAPIRE, R. E. (1999). "Large Margin Classification Using the Perceptron Algorithm". *Machine Learning*, 37 (3), 277–296. <https://doi.org/10.1023/a:1007662407062>
- FUJIMOTO, T. (2018). "The role of technological innovation in Japan's economic success". En T. Fujimoto & R. R. Nelson (Eds.), *Technology, institutions, and government policies: Japan's comparative advantage* (pp. 3-34). Oxford University Press.
- FUKUYAMA, M. (2018, 1 agosto). Society 5.0: Aiming for a New Human-Centered Society. JEF 一般財団法人 国際経済交流財団 July/August 2018 Issue. <https://www.jef.or.jp/jspotlight/backnumber/detail/220.html>
- FUMIKO, K., ARAI, H., et. Al. (2020). *Ethical Issues Regarding the Use of AI Profiling Services for Recruiting: The Japanese Rikunabi Data Scandal (Version 1)*. arXiv. <https://doi.org/10.48550/ARXIV.2005.08663>
- FUMIKO, K., et al. (2000). *Ethical Issues Regarding the Use of AI Profiling Services for Recruiting: The Japanese Rikunabi Data Scandal*. <https://doi.org/10.48550/ARXIV.2005.08663>.
- FURNHAM, A. (2019). *Artificial intelligence: A beginner's guide*. Routledge.
- GAL, D. (2020). *Perspectives and Approaches in AI Ethics*. The Oxford Handbook of Ethics of AI, 606-624. <https://doi.org/10.1093/oxfordhb/9780190067397.013.39>
- GARCÍA SAN JOSÉ, D. (2021). "Implicaciones jurídicas y bioéticas de la inteligencia artificial". Especial consideración al marco normativo internacional. Cuadernos De Derecho Transnacional, 13(1), 255. <https://doi.org/10.20318/cdt.2021.5959>
- GARCÍA-HERRERO, M., & LÓPEZ-TARRUELLA, A. (2020). *Inteligencia artificial y derechos humanos: desafíos y oportunidades*. Tirant lo Blanch.

- GHOTBI, N., HO, M. T., et. al. (2021). "Attitude of college students towards ethical issues of artificial intelligence in an international university in Japan. *AI & Society*, 37(1), 283–290. <https://doi.org/10.1007/s00146-021-01168-2>
- GILLES, M. (S.F). *Artificial intelligence* [curso en línea]. Londres (Video), Reino Unido, Universidad de Londres (En Coursera) <https://www.coursera.org/learn/uol-machine-learning-for-all/home/welcome>.
- GOBIERNO DE JAPÓN. (2020). Social Principles of Human-centric AI. <https://www8.cao.go.jp/cstp/english/humancentricai.pdf>
- GOBIERNO DE JAPÓN. (s. f.). Realizing Society 5.0. Japan.go. Recuperado 18 de marzo de 2022, https://www.japan.go.jp/abenomics/_userdata/abenomics/pdf/society_5.0.pdf
- GOBIERNO DE MÉXICO. (2020). Estrategia Nacional de Inteligencia Artificial. https://www.gob.mx/cms/uploads/attachment/file/580987/Estrategia_Nacional_de_Inteligencia_Artificial.pdf
- GÓMEZ MONT, C., DEL POZO, C. M., et. al. (2020). *La inteligencia artificial al servicio del bien social en América Latina y el Caribe: Panorámica regional e instantáneas de doce países*. Banco Interamericano de Desarrollo, 1–138. <https://doi.org/10.18235/0002393>.
- GONZALO-BRIZUELA, R. y GARRIDO-MERCHAN, E. *ChatGPT is not all you need. A State of the Art Review of large Generative AI models*. arXiv, 11 de enero de 2023. arXiv.org, <http://arxiv.org/abs/2301.04655>.
- GOOGLE (S.f). *Perspectives on Issues in AI Governance*. Recuperado 12/01/2023. <https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf>
- GRÖNLUND, Å & HORAN, T. (2004). *Introducing e-Gov: History, Definitions, and Issues. Communications of the Association for Information Systems*. 15. 713-729. 10.17705/1CAIS.01539.

- GUGGENHEIM, M., & POTTHAST, J. (2012). "Symmetrical twins: On the relationship between Actor-Network theory and the sociology of critical capacities". *European Journal of Social Theory*, 15(2), 157-178. <https://doi.org/10.1177/1368431011423601>
- HARDING, H. (2003). "Japan's long road to modernization: Modernization, culture, and identity". En *H. Harding (Ed.), Japan's modern histories* (pp. 1-27). Routledge.
- HAYES, P. J., & MORGENSTERN, L. (2007). "On John McCarthy's 80th Birthday, in Honor of His Contributions". *AI Magazine*, 28 (4), 93. <https://doi.org/10.1609/aimag.v28i4.2063>
- HAZELWOOD, K., BIRD, S., et. al. (2018). "Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective". *IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 1 (1), 620–629. <https://doi.org/10.1109/hpca.2018.00059>
- HENRICH, J. (2020). *The weirdest people in the world: How the West Became Psychologically Peculiar and Particularly Prosperous*. Penguin UK.
- HENRY, E., & BARRAGÁN, T. (2013). "Las lecciones del modelo económico de China en la reforma del Estado en México y América Latina". En *Las relaciones económicas y geopolíticas de América Latina con China, Estados Unidos y Rusia (1977–2010)*. 121–139. Universidad Autónoma Metropolitana, Unidad Xochimilco, División de Ciencias Sociales y Humanidades.
- Hill, K. (2020). The Secretive Company That Might End Privacy as We Know It. The New York Times.
- HOFSTEDE, G. (1984). *CULTURE'S CONSEQUENCES: INTERNATIONAL DIFFERENCES IN WORK-RELATED VALUES*. Sage.
- Hofstede, G. (1985). "The interaction between national and organizational value systems 1". *Journal of Management Studies*, 22(4), 347-357.

HÖLLDOBLER, S. & MÖHLE, S. (2017). *Lessons Learned from AlphaGo*.

HONDA, K. (2019). "Use of Robots in Healthcare: the Japanese Experience and the Relevance of Culture." *Robo Ethics: Humans, Machine and Health, Use of Robots in Healthcare: the Japanese Experience and the Relevance of Culture*.

HONJO, E. (1928). "Changes of social classes during the tokugawa period". En *Kyoto University Economic Review* (nº 1; Vol. 3, pp. 56–74). Department of Economics in the Imperial University of Kyoto.
<https://doi.org/10.11179/ker1926.3.56>

HUANG, C., ZHANG, Z., MAO, B., & Yao, X. (2023). An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence*, 4(4), 799-819.
<https://doi.org/10.1109/TAI.2022.3194503>

IBARRA, B. L. (2020). *Diccionario de Protección de Datos Personales. Conceptos Fundamentales*. Ciudad de México: Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI). Recuperado de https://home.inai.org.mx/wp-content/documentos/Publicaciones/Documentos/DICTIONARIO_PDP_digital.pdf

Iniciativa que reforma y adiciona diversas disposiciones de la ley de ciencia y tecnología, suscrita por el diputado Mario Alberto Rodríguez Carrillo (LXIV Legislatura 07.01.2021). México.

INMAN, D., REJEWSKI (2000). *Patterns*, 1 (1),
<https://doi.org/10.1016/j.patter.2020.100011>.

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA (INEGI) (2018). *Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares* (ENDUTIH) 2018. (179/19). Recuperado de https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2019/OtrTemEcon/ENDUTIH_2018.pdf

Instituto Nacional de Estadística y Geografía (INEGI). (2024, 13 de junio). Comunicado de prensa número 372/24: Encuesta Nacional sobre Disponibilidad

y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2023 [PDF]. Recuperado de https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2024/ENDUTIH/ENDUTIH_23.pdf

INSTITUTO NACIONAL DE TRANSPARENCIA, ACCESO A LA INFORMACIÓN Y PROTECCIÓN DE DATOS PERSONALES (INAI) (2021). Acuerdo INAI Interponer acción de inconstitucionalidad ante SCJN en contra del Decreto que crea Padrón Nacional de usuarios de telefonía móvil. Recuperado de <https://home.inai.org.mx/wp-content/documentos/SalaDePrensa/Comunicados/Comunicado%20INAI-139-21.pdf>

INSTITUTO NACIONAL DE TRANSPARENCIA, ACCESO A LA INFORMACIÓN Y PROTECCIÓN DE DATOS PERSONALES (INAI). (2014). Expediente PPD.0094/14: *Google México*, Recuperado de <https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2017/06/Mexico-RTBF-INAI-DerechoOlvidoCorrupcion.pdf>

Ioku, T., Song, J., & Watamura, E. (2024). Trade-offs in AI assistant choice: Do consumers prioritize transparency and sustainability over AI assistant performance? *Big Data & Society*, 11(4). <https://doi.org/10.1177/20539517241290217>

IPUMS. (2019, 9 septiembre). What is IPUMS? <https://ipums.org/what-is-ipums>

JAPANESE GOVERNMENT (2020). *Artificial Intelligence Promotion Law*. Recuperado de https://www.soumu.go.jp/main_content/000687877.pdf

JAPANESE GOVERNMENT (2021). *AI Governance in Japan Ver. 1.0*. Retrieved from <https://www.meti.go.jp/press/2020/01/20210115003/20210115003-3.pdf> en el 12 diciembre de 2022..

JAPANESE GOVERNMENT (2021). *AI Governance in Japan Ver. 1.1*. Retrieved from https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_8.pdf en el 22 de diciembre de 2022.

- JECKER, N. S., & NAKAZAWA, E. (2022). "Bridging East-West differences in ethics guidance for AI and robotics". *AI*, 3(3), 764–777. <https://doi.org/10.3390/ai3030045>
- JENSEN, C. B., & BLOK, A. (2013). "Techno-animism in Japan: Shinto Cosmograms, Actor-network Theory, and the Enabling Powers of Non-human Agencies". *Theory, Culture & Society*, 30(2), 84-115. <https://doi.org/10.1177/0263276412456564>
- JEONG HOON, J., PRADIP KUMAR, S., et. al. (2019). "Emerging Technologies for Sustainable Smart City Network Security: Issues, Challenges, and Countermeasures". *Journal of Information Processing Systems*, 15 (4), 765-784. DOI: 10.3745/JIPS.03.0124.
- JINJA, H. *Association of Shinto Shrines*. <https://www.jinjahoncho.or.jp/en/index.html>. Recuperado 13 de enero de 2021, de <https://www.jinjahoncho.or.jp/en/index.html>
- JOHNSON, M. G., SYMONIDES, J. (1998). *The Universal Declaration of Human Rights* (vol. 1). UNESCO Pub.
- Jung, A. (2022). *Machine learning: The basics*. Springer.
- KAWAKAMI, K., & TANAKA, H. (2020). "Japanese regulatory approaches to artificial intelligence". *International Journal of Law and Information Technology*, 28 (3), 282-303. doi:10.1093/ijlit/ea008
- KAWASAKI, K. (2018). "Meiji Ishin and the Meiji Constitution". En *The Meiji State and the Meiji Constitution* (pp. 17-32). Routledge.
- KAWASHIMA, T., & SAKURADA, R. (2007). *Introduction to Japanese Law*. University of Tokyo Press.
- KIM, D. (2019, 25 julio). *Artificial Intelligence Policies in East Asia: An Overview from the Canadian Perspective*. Asia Pacific Foundation of Canada.

<https://www.asiapacific.ca/research-report/artificial-intelligence-policies-east-asia-overview-canadian>

KITANO, N. (2006). "Rinri": An Incitement towards the Existence of Robots in Japanese Society." *International Review of Information Ethics* 6: 78–83.

KITANO, N. (2007). *Animism, Rinri, Modernization; the Base of Japanese Robotics*.

KORNICKI, P. F. (1995). MEIJI. En P. F. KORNICKI & I. J. McMULLEN (Eds.), *The Cambridge encyclopedia of Japan* (pp. 193-195). Cambridge University Press.

La Comisión acoge con satisfacción el acuerdo político sobre la Ley de Inteligencia Artificial | Configurar el futuro digital de Europa. (2023). <https://digital-strategy.ec.europa.eu/es/news/commission-welcomes-political-agreement-artificial-intelligence-act>

LACAN, J. (2005). *O Seminário, livro 10: A Angústia*. Jorge Zahar Editor.

LATOUR, B. (1990). "Technology is Society Made Durable". *The Sociological Review*, 38(1 suppl), 103–131. <https://doi.org/10.1111/j.1467-954x.1990.tb03350.x>

LATOUR, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. En W. E. Bijker y J. Law (Eds.), *Shaping technology/building society: Studies in sociotechnical change* (pp. 225-258). MIT Press.

LATOUR, B. (1993). *La clef de Berlin et autres leçons d'un amateur de sciences*. Paris: La Découverte.

LATOUR, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford Univ Pr (Txt).

LATOUR, B. (2008). "A cautious Prometheus? A few steps toward a philosophy of design (with special attention to Peter Sloterdijk). En F. HACKNE, et. al. (eds.), *Networks of design* (pp. 2-10). Universal Publishers.

- LATOUR, B. (2017). "On Actor-Network Theory. A Few Clarifications, Plus More Than a Few Complications". *Philosophical Literary Journal Logos*, 27 (1), 173-197. <https://doi.org/10.22394/0869-5377-2017-1-173-197>
- LATOUR, B., & AÚZ, T. F. (2001). *La Esperanza de Pandora*. Gedisa.
- LATOUR, B., & PORTER, C. (1993). *We Have Never Been Modern*. Harvard University Press.
- Ledesma, M. I. Á. (2023). Derechos humanos: una visión multidimensional.
- LESLIE, D. (2020). *Understanding bias in facial recognition technologies*.
- LEY FEDERAL DE PROTECCIÓN DE DATOS PERSONALES EN POSESIÓN DE LOS PARTICULARES. DOF: 05.07.2010 <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>
- Luscombe, A. (2018). "Eleanor Roosevelt: A crusading spirit to move human rights forward". *Netherlands Quarterly of Human Rights*, 36 (4), 241–246. <https://doi.org/10.1177/0924051918801610>
- LUTHFI, A. M., KARNA, N., et. al. (2019). *Google Maps API Implementation On IOT Platform For Tracking an Object Using GPS. 2019 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob)*. <https://doi.org/10.1109/apwimob48441.2019.8964139>
- MADIEGA, M. (2021, julio). *Artificial intelligence act* (PE 698.792). EU Parlament. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2021\)69879](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)69879)
- Mamzer, H. (2006). La identidad y sus transgresiones. *Revista de Estudios de Género. La ventana*, (24), 118-149. Universidad de Guadalajara.
- Marini, F. (2022). Artificial neural networks [Presentación]. ChemSchool 2022, University of Rome "La Sapienza", Italy. <https://www.sensorfint.eu/wp-content/uploads/2022/11/ANN.pdf>
- MARCUS, A. (2002). "Dare we define user-interface design?", *Interactions*, 9 (5), 19–24. <https://doi.org/10.1145/566981.566992>

- MARKS, S. (2014). *Human Rights: A Brief Introduction. Working Paper*, Harvard School of Public Health.
- MARTÍNEZ GARCÍA, J. I. (2019). "Inteligencia y derechos humanos en la sociedad digital". Intelligence and Human Rights in Digital Society. *Cuadernos Electrónicos de Filosofía del Derecho*, 40, 168. <https://doi.org/10.7203/cefd.40.13846>
- MARTINHO-TRUSWELL, E., MILLER, H., et. al. (2018, junio). *Towards an ai strategy in Mexico*. Oxford University Press. <https://www.oxfordinsights.com/mexico>
- McSTAY, A. (2021). "Emotional AI, ethics, and Japanese spice: contributing community, wholeness, sincerity, and heart". *Philosophy & Technology*, 34 (4), 1781-1802. <https://doi.org/10.1007/s13347-021-00487-y>
- MORITA, A. (2012). A neo-communitarian approach on human rights as a cosmopolitan imperative in
- MENDOZA-ENRÍQUEZ, O. A. (2021). "El derecho de protección de datos personales en los sistemas de inteligencia artificial". *Revista IUS* 15 (48). <https://doi.org/10.35487/rius.v15i48.2021.743>
- METI. (2021, 6 julio). *Call for Public Comments on «AI Governance Guidelines for Implementation of AI Principles Ver. 1.0»* Opens. Meti.Go.Jp. Recuperado 19 de marzo de 2022, de https://www.meti.go.jp/english/press/2021/0709_004.html
- MINISTERIO DE INNOVACIÓN, CIENCIA Y DESARROLLO ECONÓMICO DE CANADÁ. (2021). *Inteligencia artificial: marco de ética y responsabilidad*. Recuperado de https://www.ic.gc.ca/eic/site/061.nsf/eng/h_00043.html
- Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2024). AI R&D in Japan (Chapter 2). Retrieved from https://www.mext.go.jp/en/content/20241224-mxt_chousei01-000036407-06.pdf
- Mishra, R. K., Reddy, G. Y. S., & Pathak, H. (2021). The understanding of deep learning: A comprehensive review. *Mathematical Problems in Engineering*, 2021, 1-15. <https://doi.org/10.1155/2021/5548884>

- Mohsienuddin, S., & Sabri, M. (2020). Facial recognition technology. SSRN Electronic Journal, 7(176-184)
- MORCK, R and NAKAMURA, M. (2005). "A Frog in a Well Knows Nothing of the Ocean: A History of Corporate Ownership in Japan", p. 367-466 in *A History of Corporate Governance around the World: Family Business Groups to Professional Managers*, National Bureau of Economic Research, Inc.
- MORENO, F. R. (2024). "Generative AI and deepfakes: a human rights approach to tackling harmful content". *International Review of Law Computers & Technology*, 1–30. <https://doi.org/10.1080/13600869.2024.2324540>
- MORGAN, P. (2018). "Machine Learning is Changing the Rules". Culemborg, Países Bajos: Van Duuren Media.
- MORI, H. (2019). *Historia y naturaleza jurídica del derecho japonés*. Instituto de Investigaciones Jurídicas, UNAM. Recuperado de <https://repositorio.unam.mx/contenidos/5012006>
- MORITA, A. (2012 september), "Uma abordagem neo-comunitária nos direitos humanos como um imperativo cosmopolita na Ásia Oriental", *Philosophy of Technology*, 33, 551–569. <https://doi.org/10.1007/s13347-020-00412-9>.
- MORTON, J. L. (2020). *Las plataformas digitales de trabajo en la Ciudad de México: Uber, Didi, la economía gig y nuevas formas de socialización* (Folio 2183804240 / 468). Tesis de Maestría, Universidad Autónoma Metropolitana.
- MORTON, J. L. (2021). *Más allá de una historia: La teoría de actor-red y la narrativa en los videojuegos*. En TecnoEntornos Entornos multidiversos mediados por la innovación tecnológica: pp. 185–197. GKA Ediciones. <https://gkacademics.com/es/book/tecnoentornos-entornos-multidiversos-mediados-por-la-innovacion-tecnologica/>
- Morton, J. L. (2022). Replika y la compañía de la inteligencia artificial emocional. VISUAL REVIEW International Visual Culture Review / Revista Internacional De Cultura Visual, 10(3), 1–13. <https://doi.org/10.37467/revvisual.v9.3606>

MORTON, J. L. (2023). *On Actor-network Theory and Algorithms: ChatGPT and the New Power Relationships in the Age of AI*. AI and Ethics. <https://doi.org/10.1007/s43681-023-00314-4>

Morton, J. L. (2024). On inscription and bias: data, actor network theory, and the social problems of text-to-image AI models. AI and Ethics. <https://doi.org/10.1007/s43681-024-00431-8>

MÜLLER, V. (2020). "Ethics of Artificial Intelligence", en *The Routledge social science handbook of AI* (pp.1-20). RoutledgeProject: "Ethics of AI and Robotics" for Stanford Encyclopedia of Philosophy

NAKASONE, T. (2014). "Diferencias culturales entre México y Japón: desde las perspectivas de los japoneses en Guadalajara". *Revista mexicana de estudios sobre la Cuenca del Pacífico*, 8(3), 29–59. <http://www.portesasiapacifico.com.mx/revistas/epocaiii/numero16/2.pdf>

NEWMAN, J. (2023). *A taxonomy of trustworthiness for artificial intelligence: Connecting properties of trustworthiness with risk management and the AI lifecycle*. Center for Long-Term Cybersecurity. Recuperado el 29 de noviembre de 2023. Asamblea Legislativa de El Salvador. (2023). Ley general para la modernización digital del estado. <https://www.asamblea.gob.sv/sites/default/files/documents/correspondencia/7D94EDC0-62DF-4334-85AA-90935557B9E7.pdf>

Ng, Andrew. (s.f). Geoffrey Hinton interview [Vídeo]. Deep Learning AI (en Coursera). <https://www.coursera.org/lecture/neural-networks-deep-learning/geoffrey-hinton-interview-dcm5r>

Ng, Andrew. (s.f). Neural networks and Deep learning semana 1 [Vídeo]. Deep Learning AI (en Coursera). <https://www.coursera.org/learn/neural-networks-deep-learning/home/week/1>

- NGUYEN, T. & NGUYEN, C. et. al. (2019). *Deep Learning for Deepfakes Creation and Detection: A Survey*.
- NINO, C. S. (1989). *Ética y derechos humanos*. Alianza Editorial.
- NISHIDA, T. (2012). The Best of AI in Japan — Prologue. *AI Magazine*, 33 (2), 108.
<https://doi.org/10.1609/aimag.v33i2.2358>
- NIST. (2021). *Draft -Taxonomy of AI Risk*. Recuperado el 29 de noviembre de 2023, de 4.
- O'NEIL, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group (NY).
- O'REGAN, G. (2012). *A Brief History of Computing*. Springer.
- Ohlhorst, D., & Schön, S. (2015). Constellation Analysis as a means of interdisciplinary innovation research – theory formation from the bottom up.
doi:10.12759/hsr.40.2015.3.258-278
- OHMAE, K. (1986). *The Impact of Artificial Intelligence on Japanese Society*.
- OKAZAKI, Y. (2020). *Japan's Relationship with Artificial Intelligence: A Cultural Perspective*.
- OLAZARAN, M. (1996). "A Sociological Study of the Official History of the Perceptrons Controversy". *Social Studies of Science*, 26 (3), 611–659.
<https://doi.org/10.1177/030631296026003005>
- ONU. (2019). Resolución aprobada por el Consejo de Derechos Humanos el 26 de septiembre de 2019 (Resolución 42o período de sesiones). Organización de las Naciones Unidas.
<https://docstore.ohchr.org/SelfServices/FilesHandler.ashx?enc=dtYoAzPhJ4NMy4Lu1TOebIM8c1X4GZjGEGHV9SBM9XScbOWilnFQtxKb2selGQpTIGfsGsFM7FyA4HVW4dFoJ8HnRxMIE717i7F2cfEG85R03fWJu8IzNpANe0oSyWYR>

- ORGANIZACIÓN ÉTICA DE LA CULTURA DE LA UNIVERSIDAD DE AUBURN. (2022). *Google's Search for Solutions to Privacy Issues*. Recuperado de <https://harbert.auburn.edu/binaries/documents/center-for-ethical-organizational-cultures/cases/google.pdf>
- ORTEGA, R. (2023). *Derecho e inteligencia artificial: Cuatro Estudios*. Ediciones Jurídicas Olejnik.
- ORTIZ DE ZÁRATE ALCARAZO, L. (2022). "Explicabilidad (de la inteligencia artificial)". *Eunomía. Revista en Cultura de la Legalidad*, 22, 328-344. <https://doi.org/10.20318/eunomia.2022.6819>
- Páez, A. (2025). Transparencia, explicabilidad y confianza en los sistemas de aprendizaje automático. En J. D. Gutiérrez & R. Manrique (Eds.), *IA en acción: Desarrollo, impacto e implicaciones éticas en la sociedad*. Ediciones Uniandes.
- PAPACCHINI, A. (1998). "Los derechos humanos a través de la historia". *Revista Colombiana de Psicología*, 7.
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Books.
- PARDO-BOSCH, F. (2019). *La Inteligencia artificial para Impulsar los objetivos de desarrollo sostenible* [Libro electrónico]. En TORRENS, M. (Ed.), *Hacia el gobierno digital en México: conceptos y experiencias* (pp. 41–54). Prometeo Editores. <http://iippg.cucea.udg.mx/sites/default/files/Hacia%20el%20Gobierno%20Digital%20de%20Mexico%20%2819-12-05%29.pdf>
- PARLAMENTO EUROPEO. (2020). *Inteligencia artificial: oportunidades y desafíos*. Recuperado de <https://www.europarl.europa.eu/news/es/headlines/society/20200918STO87404/inteligencia-artificial-oportunidades-y-desafios>
- PARLAMENTO EUROPEO. (2023). *EU AI Act: first regulation on artificial intelligence*. Recuperado de

<https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>

PÉREZ, R., CAMACHO, O. (2016). *Análisis general del gobierno electrónico en México*, 5 (9), 1-12. Recuperado de <http://www.udgvirtual.udg.mx/paakat/index.php/paakat/article/view/253/376>

PERRY, M. C. (1853). *Japan expedition, 1852-1854: The report of Commodore Perry to the Secretary of the Navy*.

Purwono, P., Ma'arif, A., Rahmaniar, W., Fathurrahman, H. I. K., Frisky, A. Z. K., & Haq, Q. M. U. (2023). Understanding of convolutional neural network (Cnn): A review International Journal of Robotics and Control Systems, 2(4), 739-748. <https://doi.org/10.31763/ijrcs.v2i4.888>

QAMMAR, A., KARIM, A., et. al. (2022). "Securing Federated Learning with Blockchain: A Systematic Literature Review". *Artificial Intelligence Review*, 56 (5), 3951-3985. <https://doi.org/10.1007/s10462-022-10271-9>

RAJARAMAN, V. (2014). "JohnMcCarthy — Father of artificial intelligence". *Resonance*, 19(3), 198–207. <https://doi.org/10.1007/s12045-014-0027-9>

RAMÍREZ, H. E., & GUERRERO, J. C. (2011). "La justicia, la crítica y la justificación. Un análisis desde la perspectiva de la sociología pragmática". *Revista Colombiana de Sociología*, 34(1), 25-46.

RASO, F., Hilligoss, H., et. al. (2018). *Artificial Intelligence & Human Rights: Opportunities & Risks*. Berkman Klein Center for Internet & Society Research Publication.

REAL ACADEMIA ESPAÑOLA. (2021). "Transgresión". *En Diccionario de la Lengua Española*. Recuperado el 29 de noviembre de 2023, de 4.

RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE. (2023, 16 mayo). UNESCO. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence?TSPD_101_R0=080713870fab2000ebd131a77e68a1d7571d3dbc35b3e0e89e0

54080964d6b5dd1f162bc3a03c59708ca12ca051430008eaef992189509bb173c
3e23dc6e7573d6f31b3551bb8a3e164ded2cf4fafcaf42a8aa7791c0df4ef5f7f122
302068a9

REILING, A. D. D. (2020). "Courts and Artificial Intelligence". *International Journal for Court Administration*, 11(2). <https://doi.org/10.36745/ijca.343>

REN, K., ZHENG, T., et. al. (2020). *Adversarial Attacks and Defenses in Deep Learning*. *Engineering*, 6 (3), 346–360. <https://doi.org/10.1016/j.eng.2019.12.012>

RENDÁ, A. (2018). "Ethics, algorithms and self-driving cars a CSI of the 'trolley problem'". *Policy Insights: thinking about ahead for Europe*, 2018 (02), 3–14. http://aei.pitt.edu/93153/1/PI_2018-02_Renda_TrolleyProblem_1.pdf

Responsible AI Transparency Report How we build, support our customers, and grow Contents. (2024). <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW1I5BO>

RICO ÁLVAREZ, F., GARZA BANDALA, P. (2011). *Introducción al estudio del derecho civil y personas*. Editorial Porrúa.

RICHARDSON, K. (2016). "Technological Animism: The Uncanny Personhood of Humanoid Machines". *Social Analysis*, 60 (1). <https://doi.org/10.3167/sa.2016.600108>

ROBERTS, H., COWLS, J., et. al. (2021). "The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation". *Philosophical Studies Series*, 1(1), 47–79. https://doi.org/10.1007/978-3-030-81907-1_5

ROMÁN VERGARA, J. (s. f.). No 01. *Protección de Datos*, México Transparente. https://snt.org.mx/revistadigital/?page_id=1113. Recuperado 29 de abril de 2022, de https://snt.org.mx/revistadigital/?page_id=1113

ROSENBERG, M. (2018). *Law enforcement is using how facial recognition technology*. American Civil Liberties Union.

- ROYAL SOCIETY. (2017). *Machine Learning*. Royal Society.
- Ryan, M. (2020). In *AI We Trust: Ethics, Artificial Intelligence, and Reliability. Science and Engineering Ethics*, 26(5), 2749–2767.
<https://doi.org/10.1007/s11948-020-00228-y>
- SAEED & AL-MUSHAYT, O. S. (2019). “Automating E-Government Services With Artificial Intelligence”. *IEEE Access*, 7, 21-29.
<https://doi.org/10.1109/access.2019.2946204>
- SÁNCHEZ RODRÍGUEZ, A., & RODRÍGUEZ, G. (2020). “Gobernanza, Gobierno y Servicios Públicos”. En DEL POZO, C. GÓMEZ MONT, C. (eds.), *Agenda Nacional Mexicana de Inteligencia Artificial* (pp. 85-108). México: IA2030Mx.
- SCHABAS, W. A. (2013). *The Universal Declaration of Human Rights 3 Volume Hardback Set: The Travaux Préparatoires*. Cambridge University Press.
- SCHUCHMANN, S. (2019). *Analyzing the Prospect of an Approaching AI Winter*. 10.13140/RG.2.2.10932.91524.
- SENADO DE LA REPÚBLICA. (2021, enero). *La implementación del TEMEC: una prueba para América del Norte* (No 1).
https://centrogilbertobosques.senado.gob.mx/docs/La_implementacion_del_T-MEC_una_prueba_para_America_Norte.pdf
- SILVER, D., HUANG, et. al. (2016). “Mastering the game of Go with deep neural networks and tree search”. *Nature*, 529 (7587), 484–489.
<https://doi.org/10.1038/nature16961>
- SOLAECHÉ, M. C. (1994). “Lady Ada Byron y el primer programa para computadoras”. *Divulgaciones Matemáticas*, 2 (1), 75–81.
https://www.academia.edu/31864989/Lady_Ada_Byron_y_el_primer_programa_para_computadoras

- SONI, J., & RAYMAKER, J. (2011). "Focus Introduction: Toward Sharing values across cultures and religions". *Journal of Religious Ethics*, 39 (2), 193–203. <https://doi.org/10.1111/j.1467-9795.2011.00472>.
- STAR, S. L., & GRIESEMER, J. R. (1989). "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39". *Social Studies of Science*, 19(3), 387–420. <https://doi.org/10.1177/030631289019003001>
- STEIN, D. (1985). *Ada: A Life and a Legacy*. Mit Pr.
- SULLIVAN, R. J. (1989). Immanuel Kant's moral theory (1.a ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511621116>
- TAN, K. H., & LIM, B. P. (2018). "The artificial intelligence renaissance: deep learning and the road to human-Level machine intelligence". *APSIPA Transactions on Signal and Information Processing*, 7, 1–19. <https://doi.org/10.1017/atsip.2018.6>
- TASHAKKORI, A., & TEDDLIE, C. (1998). *Mixed methodology: combining qualitative and quantitative approaches*. Thousand Oaks, Calif.
- Textos legislativos | ley de inteligencia artificial de la UE (AI Act). (2024). Recuperado 16 de julio de 2024, de <https://artificialintelligenceact.eu/es/el-acto/>
- THÉVENOT, L. (1984). "Rules and implements: investment in forms". *Social Science Information*, 23 (1), 1–45. <https://doi.org/10.1177/053901884023001001>
- THEVENOT, L., & BOLTANSKI, L. (1991). *De la justification: Les économies de la grandeur*. Gallimard.
- TOXTLI, C., SURI, S., & SAVAGE, S. (2021). *Quantifying the Invisible Labor in Crowd Work*. arXiv. <https://doi.org/10.48550/ARXIV.2110.00169>
- Toxtli, C., & Savage, S. (2023). Designing AI Tools to Address Power Imbalances in Digital Labor Platforms. *Human-Computer Interaction Series*, 121–137. https://doi.org/10.1007/978-3-031-31642-5_9

- TREJO, S., MEZA, I., & LÓPEZ-ESCOBEDO, F. (2020). *Hacia los Comités de Ética en Inteligencia Artificial*. <https://doi.org/10.48550/ARXIV.2002.05673>
- TSUNODA, R. (1951). "The samurai tradition in modern Japan". En A. F. Wright (Ed.), *The Confucian Persuasion* (pp. 186-214). Stanford University Press.
- TURING, A. (1950). *Computing machinery and intelligence*. *Mind*, LIX(236), 433–460. <https://doi.org/10.1093/mind/lix.236.433>
- TURPO GEBERA, O. W. (2008). "La netnografía: un método de investigación en Internet". *Revista Iberoamericana de Educación*, 47(2), 1-10. <https://doi.org/10.35362/rie4722372>
- UNESCO. (2022). Recomendación sobre la ética de la inteligencia artificial (SHS/BIO/PI/2021/1). Francia: UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa
- VADILLO BUENO, G. (2020). "Futuros de la inteligencia artificial". *Revista Digital Universitaria*. Vol. 21, nº 1 enero-febrero. <http://doi.org/10.22201/codeic.16076079e.2020.v21n1.a3>
- VALENZUELA, M. E. (2006). *Japan and Mexico: A Strategic Partnership for the 21st Century*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008.
- VELARDE, G. (2019). "Artificial Intelligence and its Impact on the Fourth Industrial Revolution: A Review". *International Journal of Artificial Intelligence & Applications*, 10 (6), 41-48. <https://doi.org/10.5121/ijaia.2019.10604>
- VÉLIZ, C (2021). "La privacidad como instrumento de resistencia colectiva". *Anuario internacional CIDOB*, nº 1, pp. 21-27. https://www.cidob.org/articulos/anuario_internacional_cidob/2021/la_privacidad_como_instrumento_de_resistencia_colectiva

- VÉLIZ, C. (2020). “La privacidad es poder”. *Paradigma*, 1(23), 24–31.
<https://riuma.uma.es/xmlui/bitstream/handle/10630/19524/24.pdf?sequence=1&isAllowed=y>
- VÉLIZ, C. (2021). “Moral zombies: why algorithms are not moral agents”. In *AI & Society* (Vol. 36, Issue 2, pp. 487–497). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00146-021-01189-x>
- Véliz, C. (2022). Privacidad es poder: Datos, vigilancia y libertad en la era digital. Debate
- VENTURINI, J. y GARAY, V. (2021), *Reconocimiento facial en América Latina: tendencias en la implementación de una tecnología perversa*, *Al Sur*, https://estudio.reconocimientofacial.info/reports/ALSUR-Reconocimiento_f...
- VERLOO, N. (2019). “Captured by bureaucracy”. *Planning and knowledge*, 1(1), 75–88. <https://doi.org/10.2307/j.ctvkjb1z8.11>
- VILLORIA, M. & RAMIREZ-ALUJAS, A. (2013). “Los modelos de gobierno electrónico y sus fases de desarrollo: Un análisis desde la teoría política”. *Gestión y Política Pública*. 22. 69-103.
- VVAA (2006), History of AI, University of Washington, History of Computing Course (CSEP 590A), <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf>.
- VVAA. Behind Covid-19 vaccine development. (2021, 18 mayo). MIT News | Massachusetts Institute of Technology. <https://news.mit.edu/2021/behind-covid-19-vaccine-development-0518>
- WALKER, B. L. (2015). *A Concise History of Japan*. Cambridge University Press.
- WALTZ, S. (2002). “Reclaiming and Rebuilding the History of the Universal Declaration of Human Rights”. *Third World Quarterly*, 23(3), 437–448. <http://www.jstor.org/stable/3993535>

- WANG Z., QIN Z., et. al. (2019, 10 agosto). *Artificial Intelligence in Transportation. DiDi Chunxing*. <https://outreach.didichuxing.com/IJCAI2019/tutorial/>
- WANG, H. & RAJ, B., *On the Origin of Deep Learning*, arXiv preprint arXiv:1702.07800.
- WANG, M. & SHAHEER, N. et. al. (2020). *Tiktok's rise to global markets 1*.
- Wang, K., Cai, M., Wang, J., & Chen, E. (2024, May 2). From predictive to Generative - How Michelangelo Accelerates Uber's AI journey | Uber Blog. Uber Blog. <https://www.uber.com/en-GB/blog/from-predictive-to-generative-ai/>
- WEBB, M. (2019). "The Impact of Artificial Intelligence on the Labor Market". *SSRN Electronic Journal*, 108(6), 1143-1171. <https://doi.org/10.2139/ssrn.3482150>
- WEBER, M. (2012). "Los tipos de dominación". En *Economía y Sociedad* (pp. 334-422). México: Fondo de Cultura Económica.
- Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11), 39–52. <https://doi.org/10.22215/timreview/1282>
- Williams-Ceci, S., Jakesch, M., Bhat, A., Kadoma, K., Zalmanson, L., & Naaman, M. (2024). Bias in AI autocomplete suggestions leads to attitude shift on societal issues [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/mhjn6>
- WENGER, E., SHAN, S., et. al. (2023). *SoK: Anti-facial recognition technology. In 2023 IEEE Symposium on Security and Privacy*, pp. 864-881. IEEE. <https://doi.org/10.1109/SP46215.2023.10179445>
- WINGET, M. A. (2008). "A methodology and model for studying boundary objects, annotations, and collaborative practices: Musicians and musical scores". *Proceedings of the American Society for Information Science and Technology*, 44(1), 1–13. <https://doi.org/10.1002/meet.1450440238>

- YASEEN, H. K., & OBAID, A. M. (2020). BIG DATA: DEFINITION, ARCHITECTURE & APPLICATIONS. *JOIV : INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION*, 4(1), 45–51. [HTTPS://DOI.ORG/10.30630/JOIV.4.1.292](https://doi.org/10.30630/JOIV.4.1.292)
- YONEYAMA, S. (2017). “Animism: a grassroots response to socioenvironmental crisis in Japan: New Worlds from Below. In Morris-Susiki and Soh (Ed). Informal life politics and grassroots action in twenty-first-century Northeast Asia”. *Asian Studies Series* (pp. 99-130). ANU Press.
- YONEYAMA, S. (2020). “Rethinking human-nature relationships in the time of coronavirus: Postmodern animism in films by Miyazaki Hayao & Shinkai Makoto”. *The Asia-Pacific Journal | Japan Focus*, 18(6), 1-16. <https://digital.library.adelaide.edu.au/dspace/handle/2440/129453>
- YU, R., & ALI, G. S. (2019). “What’s Inside the Black Box? AI Challenges for Lawyers and Researchers”. *Legal Information Management*, 19(01), 2–13. <https://doi.org/10.1017/s1472669619000021>
- ZHANG, D., MISHRA, S., et. al. (2021). *The AI Index Report – Artificial Intelligence Index. AI Index Steering Committee*, Human-Centered AI Institute, Stanford University, Stanford, CA. Recuperado 11 de marzo de 2022, de <https://aiindex.stanford.edu/report/>
- ZHAO, Z. (2021). “Analysis on the “Douyin (Tiktok) Mania” Phenomenon Based on Recommendation Algorithms”. *E3S Web of Conferences*, 235(1), 03029. <https://doi.org/10.1051/e3sconf/202123503029>
- ZUIDERWIJK, A., CHEN, Y. C., & SALEM, F. (2021). *Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda*. *Government Information Quarterly*, 101577. <https://doi.org/10.1016/j.giq.2021.101577>

2. Hemerografía

ADVANCING FACE RECOGNITION SEPTEMBER 2022 HIGHLIGHTING JAPAN. (n.d.).
https://www.gov-online.go.jp/eng/publicity/book/hlj/html/202209/202209_10_en.html

AFP. (2024, November 20). Japan ramps up tech ambitions with \$65 billion for AI, chips. The Economic Times; Economic Times.
<https://economictimes.indiatimes.com/tech/artificial-intelligence/japan-ramps-up-tech-ambitions-with-65-billion-for-ai-chips/articleshow/115474130.cms?from=mdr>

AGENCIA DE SEGURIDAD NACIONAL DE JAPÓN (2020). "Estado actual y perspectivas del reconocimiento facial". Disponible en:
<https://www.npa.go.jp/cybercrime/report/202007/P0201900.pdf>

AGENCIA DE SEGURIDAD NACIONAL DE JAPÓN (NSO) (2019). Sistema de control y monitoreo de acceso para eventos masivos. Disponible en:
https://www.nso.go.jp/business/business_support/AI/event/index.html (en japonés)

AGUIAR, A. R. (2023). Uber, Cabify y Bolt pueden incumplir varias leyes por sus algoritmos. Business Insider. 3

AHMAD, A. (2023, 8 junio). Racial bias in facial recognition algorithms. Amnesty International Canada.
<https://www.amnesty.ca/surveillance/racial-bias-in-facial-recognition-algorithms/>

AI ACT. (2024, March 1). Shaping Europe's Digital Future.
<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

Airportindustry-News. (2021, April 13). Narita Airport, Amadeus and NEC Introduce Biometric Boarding Process. Airport Industry-News.
<https://airportindustry-news.com/narita-airport-amadeus-and-nec-introduce-biometric-boarding-process/>

- ALBRIEU, R., RAPETTI, M., Brest López, C., Larroulet, P., Sorrentino, A., & Centro de Implementación de Políticas Públicas para la Equidad y el Crecimiento. (2018, 1 noviembre). Inteligencia artificial y crecimiento económico. Oportunidades y desafíos para México Microsoft [Comunicado de prensa]. <https://news.microsoft.com/uploads/prod/sites/41/2018/11/IA-y-Crecimiento-MEXICO.pdf>
- ALDAMA, Z. (2018, 29 marzo). El sistema de crédito social chino salta de la teoría a la práctica. El País. Recuperado 15 de noviembre de 2021, de https://elpais.com/retina/2018/03/27/tendencias/1522145305_569868.html
- ALPHABET-GOOGLE. (2023, febrero). AI Test Kitchen. aitestkitchen.withgoogle.com. Recuperado 9 de febrero de 2023, de <https://aitestkitchen.withgoogle.com>
- AL-SIBAI, N. (2022, 11 marzo). Man Credits Affair With AI Girlfriend For Saving His Marriage. Futurism. Recuperado 14 de marzo de 2022, de <https://futurism.com/ai-girlfriend-wife>
- Amadeus. (2022, May 27). Narita Airport leads the way in Japan with biometrics. YouTube. <https://www.youtube.com/watch?v=6klsSrsRTS4>
- Amazon Web Services. (n.d.). What is Computer Vision? - Image recognition AI/ML Explained - AWS. Amazon Web Services, Inc. Retrieved February 26, 2025, from <https://aws.amazon.com/what-is/computer-vision/>
- AMNISTÍA INTERNACIONAL SECCIÓN ESPAÑOLA. (s. f.). La vigilancia mediante reconocimiento facial retrospectivo oculta abusos contra los derechos humanos. <https://www.es.amnesty.org/en-que-estamos/blog/historia/articulo/la-vigilancia-mediante-reconocimiento-facial-retrospectivo-oculta-abusos-contra-los-derechos-humanos>
- AMNISTÍA INTERNACIONAL SECCIÓN ESPAÑOLA. (s. f.-a). Debe prohibirse la peligrosa tecnología de reconocimiento facial que amplifica la actuación policial racista. <https://www.es.amnesty.org/en-que-estamos/noticias/noticia/articulo/debe-prohib>

irse-la-peligrosa-tecnologia-de-reconocimiento-facial-que-amplifica-la-actuacion-policial-racista/

AMNISTÍA INTERNACIONAL SECCIÓN ESPAÑOLA. (s. f.-c). La vigilancia mediante reconocimiento facial retrospectivo oculta abusos contra los derechos humanos. <https://www.es.amnesty.org/en-que-estamos/blog/historia/articulo/la-vigilancia-mediante-reconocimiento-facial-retrospectivo-oculta-abusos-contra-los-derechos-humanos/>

AMNISTÍA INTERNACIONAL. (2018). Los riesgos de la inteligencia artificial para los derechos humanos. Recuperado de <https://www.idhc.org/es/actualidad/los-riesgos-de-la-inteligencia-artificial-para-los-derechos-humanos.php>

AMNISTÍA INTERNACIONAL. (2022, 16 febrero). Estados Unidos: Nueva investigación revela que la tecnología de reconocimiento facial refuerza la práctica policial racista de dar el alto y registrar en Nueva York. <https://www.amnesty.org/es/latest/news/2022/02/usa-facial-recognition-technology-reinforcing-racist-stop-and-frisk-policing-in-new-york-new-research/>

Amnistía Internacional. (2024, 1 de noviembre). Suecia: Las autoridades deben dejar de usar los sistemas de inteligencia artificial discriminatorios de la agencia de bienestar social. <https://www.amnesty.org/es/latest/news/2024/11/sweden-authorities-must-discontinue-discriminatory-ai-systems-used-by-welfare-agency/>

ANIA. Ania MX. (s. f.). Ania MX. <https://www.ania.org.mx/sobre-ania>

APPLE INC (2021). M1. Recuperado de <https://www.apple.com/mac/m1/>

APPLE INC. (n.d.). Apple A12 Bionic chip. Recuperado de <https://www.apple.com/apple-silicon/>

ARAYA, D. (2019, enero 15). Artificial Intelligence And The End Of Government. Recuperado 20 de mayo de 2020, de

<https://www.forbes.com/sites/danielaraya/2019/01/04/artificial-intelligence-and-the-end-of-government/#57346eb1719b>

Article 19. (2025, February 19). AI Summit: In the rush to embrace AI, do not sacrifice human rights - ARTICLE 19. ARTICLE 19. <https://www.article19.org/resources/ai-summit-in-the-rush-to-embrace-ai-do-not-sacrifice-human-rights/>

ARM LTD. (s. f.). Architecting a Smarter World – Arm. Arm | The Architecture for the Digital World. Recuperado 5 de febrero de 2021, de <https://www.arm.com/>

ASCA TECH LAB. Society 5.0. (s. f.). <https://asca-techlab.com/society-5-0/>

AST, F. (2020, 18 marzo). Las Cortes de Internet de China: Hacia el Tribunal de Justicia del Futuro. Medium. <https://medium.com/astec/las-cortes-de-internet-de-china-hacia-el-tribunal-de-justicia-del-futuro-6521a2834757>

Basham. (2024, December 4). Reforma de simplificación orgánica: Extinción del INAI. - BASHAM. BASHAM. <https://basham.com.mx/reforma-de-simplificacion-organica-extincion-del-inai/>

BBC (2020). "México: ¿Es el sistema de reconocimiento facial de la CDMX una herramienta de seguridad o una invasión de privacidad?". Disponible en: <https://www.bbc.com/mundo/noticias-america-latina-53470987>

BBC. (2023). GPT-4: qué novedades presenta la nueva versión del chat de inteligencia artificial. Recuperado de <https://www.bbc.com/mundo/noticias-64969661>

Bischoff, P. (2020, January 9). Which governments impose SIM-card registration laws to collect data on their citizens? Comparitech. <https://www.comparitech.com/blog/vpn-privacy/sim-card-registration-laws/>

BLOOMBERG NEWS. (2020, 21 mayo). China's Got a New Plan to Overtake the U.S. in Tech. Yahoo News. <https://finance.yahoo.com/news/china-got-plan-seize-world-210000297.html>

Bovenizer, N. (2024, July 8). Japanese rail operator installs facial recognition ticket technology. Railway Technology; Railway Technology. <https://www.railway-technology.com/news/japanese-rail-operator-facial-recognition-technology/>

CAMPUSMVP. (s.f.). Qué es la IA Generativa. Recuperado de <https://www.campusmvp.es/recursos/post/que-es-la-ia-generativa.aspx>

CAUGHILL, P. (2017, 6 noviembre). An Artificial Intelligence Has Officially Been Granted Residency. Futurism. Recuperado 10 de noviembre de 2021, de <https://futurism.com/artificial-intelligence-officially-granted-residency>

CBS NEWS. (2009, 16 septiembre). Facebook «cash flow positive signs 300M users Social Sharing. <https://www.cbc.ca/news/science/facebook-cash-flow-positive-signs-300m-users-1.826223>

CLIFFORD, C. (2018, 14 marzo). Elon Musk: 'Mark my words — A.I. is far more dangerous than nukes'. CNBC. <https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html>

Comunicación Uber México. (2021, September 8). Uber celebra 8 años de generar oportunidades a través del movimiento y la tecnología en México. Uber Newsroom. <https://www.uber.com/es-MX/newsroom/uber-mexico/>

COPELAND, J. (2012, 19 junio). Alan Turing: The codebreaker who saved «millions of lives». BBC News. <https://www.bbc.com/news/technology-18419691>

COSTIGAN, J. (2023, 10 agosto). New Chinese facial recognition regulations could shield citizens from surveillance capitalism. Forbes. <https://www.forbes.com/sites/johannacostigan/2023/08/09/new-chinese-facial-re>

[cognition-regulations-could-shield-citizens-from-surveillance-capitalism/?sh=449aee7c3cc2](https://www.google.com/search?q=cognition-regulations-could-shield-citizens-from-surveillance-capitalism/?sh=449aee7c3cc2)

Coulter, J. (2024, September 10). Google for Mexico 2024: Economic progress, driven by AI. Google. <https://blog.google/around-the-globe/google-latin-america/google-for-mexico-2024/>

CHAS, M. (s. f.). La hechicera de los números. Stony Brook University. Recuperado 1 de junio de 2021, de <http://www.math.stonybrook.edu/~moira/papers/Spanish/AdaLovelace.pdf>

CHEDDAR. (2021, 23 marzo). The Unintended Consequences Of the Facebook Like Button. The Lightbulb Moment [Vídeo]. YouTube. <https://www.youtube.com/watch?v=3LI7Qywbr>

Comisión de Protección de la Información Personal. (s.f.). Act on the Protection of Personal Information. Recuperado de <https://www.ppc.go.jp/en/>

CookieHub ehf. (2021, December 10). ¿Qué es APPI? | Ley de Protección de Datos | CookieHub CMP. CookieHub - Cookie Consent Management Platform (CMP). https://www.cookiehub.com/es/blog/que-es-la-appi-ley-de-proteccion-de-datos-d-e-japon?utm_source=chatgpt.com

Cruz, R. D. L. (2023, November 27). Frank Rosenblatt's Perceptron, The Birth of The Neural Network. Medium. <https://medium.com/@robdelacruz/frank-rosenblatts-perceptron-19fcce9d627f>

Dan-Awoh, D. (2024, December 23). Google faces antitrust battle in Japan over alleged search engine bias. Nairametrics. <https://nairametrics.com/2024/12/23/google-faces-antitrust-battle-in-japan-over-a-lleged-search-engine-bias/>

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters.

DataReportal. (2024). Digital 2024: México. Recuperado de <https://datareportal.com/reports/digital-2024-mexico>

DELUA, J., SME, & IBM Analytics. (2021, 12 marzo). Supervised vs. Unsupervised Learning: What's the Difference? IBM. <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>

Deep Knowledge Group. (2024, February). AI industry in Japan: Landscape overview. Japan Ecosystem. <https://analytics.dkv.global/japan/ai-industry.pdf>

DESMOND, J. (2019, 15 noviembre). Digital Assistants Transforming Public Service. AI Trends. <https://www.aitrends.com/ai-world-government/digital-assistants-transforming-public-service/>

Diario Oficial de la Federación (DOF). (2010, 5 de julio). Decreto por el que se expide la Ley Federal de Protección de Datos Personales en Posesión de los Particulares y se reforman los artículos 3, fracciones II y VII, y 33, así como la denominación del Capítulo II, del Título Segundo, de la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental. Gobierno de México. https://dof.gob.mx/nota_detalle.php?codigo=5150631&fecha=05/07/2010#gsc.tab=0

DIDI CHUXING LAUNCHES GLOBAL DIGITAL INNOVATORS INITIATIVE. (2019, 9 septiembre). BusinessWire. <https://www.businesswire.com/news/home/20190909005476/en/Didi-Chuxing-Launches-Global-Digital-Innovators-Initiative>

Digital Agency. (n.d.). Smartphone Certification Service. Digital Agency of Japan. Retrieved from <https://www.digital.go.jp/en/policies/mynumber/smartphone-certification#Service>

Digital Agency. (n.d.). My Number system. Digital Agency, Government of Japan. Retrieved February 21, 2025, from <https://www.digital.go.jp/en/policies/mynumber>

EDITORIAL BUSINESS WIRE. (2021, 15 diciembre). The UK Leads Europe and Ranks Third Globally in Artificial Intelligence. Business Wire. Recuperado 9 de marzo de 2022, de <https://www.businesswire.com/news/home/20211215005874/en/The-UK-Leads-Europe-and-Ranks-Third-Globally-in-Artificial-Intelligence>

EDUCATION, I. C. (2021, 9 abril). Application Programming Interface (API). IBM. <https://www.ibm.com/cloud/learn/api>

El Consejo de Europa adopta el primer tratado internacional en materia de Inteligencia Artificial—IDHC. (2024). Recuperado 11 de julio de 2024, de <https://www.idhc.org/es/actualidad/el-consejo-de-europa-adopta-el-primer-tratado-internacional-en-materia-de-inteligencia-artificial.php>

El Mercurio S.A.P. (2018, April 3). Criticada, insultada y cuestionada: El caso de la japonesa que denunció públicamente su violación. Emol; Emol.com. https://www.emol.com/noticias/Internacional/2018/04/03/901038/Criticada-insultada-y-cuestionada-El-caso-de-la-japonesa-que-denuncio-publicamente-su-violacion.html?utm_source=chatgpt.com

El panorama de la inteligencia artificial en México. (s. f.). Recuperado 17 de julio de 2024, de <https://www.ikusi.com/mx/blog/inteligencia-artificial-en-mexico/>

EL UNIVERSAL. (2023). \$100 mxn Cupón Uber | Mayo 2023 | El Universal. Recuperado de <https://www.eluniversal.com.mx/descuentos/uber>

El País. (2025, 10 de febrero). DeepSeek te espía y las otras también. <https://elpais.com/opinion/2025-02-10/deepseek-te-espia-y-las-otras-tambien.html>

ENGLER, A. (2022, 28 enero). The EU and U.S. are starting to align on AI regulation. Brookings. Recuperado 7 de marzo de 2022, de <https://www.brookings.edu/blog/techtank/2022/02/01/the-eu-and-u-s-are-starting-to-align-on-ai-regulation/>

Ethics of artificial intelligence | unesco. (s. f.). Recuperado 10 de julio de 2024, de <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

EXPANSIÓN. (2019, 23 octubre). La CDMX le abre la puerta a la IA para digitalizarse. Recuperado 10 de noviembre de 2021, de <https://expansion.mx/tecnologia/2019/10/23/la-cdmx-le-abre-la-puerta-a-la-ia-para-digitalizarse>

EXPANSIÓN. (2021, 26 abril). Kavak y OLX Autos usan la inteligencia artificial para financiar autos usados. <https://expansion.mx/empresas/2021/04/26/kavak-inteligencia-artificial-comprar-autos-usados>

FACE EXPRESS WORKS BETWEEN YOU AND YOUR AIRLINE. Face Express allows "Face Pass" boarding procedures using facial recognition. This service is provided by Tokyo International Air Terminal Corporation (TIAT).

FACEBOOK REVENUE AND USAGE STATISTICS (2021). (2021, 6 julio). Business of Apps. <https://www.businessofapps.com/data/facebook-statistics/>

FEDERAL TRADE COMMISSION. Aiming for truth, fairness, and equity in your company's use of AI. (2021, 19 abril). Recuperado 7 de marzo de 2022, de <https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai>

FERNÁNDEZ, J. (2021, 11 febrero). Tecnologías emergentes: 76% de los mexicanos confían más en los robots que en los humanos cuando se trata de dinero. América Retail. Recuperado 22 de marzo de 2022, de <https://www.america-retail.com/tecnologias-emergentes/tecnologias-emergentes-76-de-los-mexicanos-confian-mas-en-los-robots-que-en-los-humanos-cuando-se-trata-de-dinero/>

Ferus, J. (2023, enero 17). Why falling in love with ai is a dangerous illusion. Medium.

<https://medium.com/@dreamferus/why-falling-in-love-with-ai-is-a-dangerous-illusion-the-limitations-and-harms-of-artificial-5e5b5f04a03c>

Flores, D. (2024, mayo). Conductores de Uber, sin seguridad social, explotados y sumidos en la violencia. Reporte Índigo. <https://www.reporteindigo.com/nacional/Conductores-de-Uber-sin-seguridad-social-explotados-y-sumidos-en-la-violencia-20240507-0031.html>

FLORES, N. (2021, 15 noviembre). En México operan tres sistemas de vigilancia con reconocimiento facial. Contralínea. <https://contralinea.com.mx/interno/semana/en-mexico-operan-tres-sistemas-de-vigilancia-con-reconocimiento-facial/>

FORBES MÉXICO. <https://www.forbes.com.mx/ya-puedes-consultar-info-de-transporte-publico-en-google-maps/>

France 24. (2017, December 17). Faring badly: Uber struggles to make inroads in Japan - France 24. France 24. https://www.france24.com/en/20171217-faring-badly-uber-struggles-make-inroads-japan?utm_source=chatgpt.com

FROM MAPS TO APPS: the power of machine learning and artificial intelligence for regulators. (2017, 22 noviembre). FCA. Recuperado 13 de junio de 2022, de <https://www.fca.org.uk/news/speeches/maps-apps-power-machine-learning-and-artificial-intelligence-regulators>

GALER, S. S. T. J. B. (2020, 5 agosto). TikTok: The story of a social media giant. BBC News. <https://www.bbc.com/news/technology-53640724>

GALINDO, J. S. (2017, 13 agosto). Richter vs. Google va a la Suprema Corte. El Economista. Recuperado 14 de enero de 2023, de <https://www.eleconomista.com.mx/opinion/Richter-vs.-Google-va-a-la-Suprema-Corte-20170813-0001.html>

García, J. (2025, 15 de febrero). Detenido un menor en Palma por crear desnudos de compañeras de clase con inteligencia artificial. El País. <https://elpais.com/sociedad/2025-02-15/detenido-un-menor-en-palma-por-crear-desnudos-de-companeras-de-clase-con-inteligencia-artificial.html>

García, J. (2025, febrero 9). Más vale que borres la app de DeepSeek en tu iPhone: No es nada segura. Cinco Días. <https://cincodias.elpais.com/smartlife/smartphones/2025-02-09/mas-vale-que-borres-la-app-de-deepseek-en-tu-iphone-no-es-nada-segura.html>

GAVILÁN, I. G. R. (2020). Explicabilidad de la Inteligencia Artificial: ¿Qué es lo que queremos explicar? Recuperado de <https://ignaciogavilan.com/explicabilidad-de-la-inteligencia-artificial-que-es-lo-que-queremos-explicar/>

GETAHUN, H. (2023, 16 enero). ChatGPT could be used for good, but like many other AI models, it's rife with racist and discriminatory bias. Insider. <https://www.insider.com/chatgpt-is-like-many-other-ai-models-rife-with-bias-2023>

Ghandi, S. (2024, junio). Case Study: How Uber uses AI to optimize surge pricing. <https://www.linkedin.com/pulse/case-study-how-uber-uses-ai-optimize-surge-pricing-shripal-gandhi-s54mf>

GIATTINO, C. (2020, 4 diciembre). Working Hours. Our World in Data. Recuperado 20 de junio de 2022, de <https://ourworldindata.org/working-hours>

GOEL, V. (2022, 25 enero). Get to know the new Topics API for Privacy Sandbox. Google. <https://blog.google/products/chrome/get-know-new-topics-api-privacy-sandbox/>

Google. (2024). Results about you. Google.com. https://myactivity.google.com/results-about-you?utm_source=HPP&utm_medium=owned&utm_campaign=23hpp&pli=1

- Google. (2023). Google AI Principles. Google AI. <https://ai.google/responsibility/principles/>
- Google. (2024a). Results about you. Google.com. https://myactivity.google.com/results-about-you?utm_source=HPP&utm_medium=owned&utm_campaign=23hpp&pli=1
- Google. (2024b, December 17). Generative AI Prohibited Use Policy. <https://policies.google.com/terms/generative-ai/use-policy>.
- GONZÁLEZ, P. (2021, 28 enero). The AI Market In Japan: Spearheading Industry Innovation - Tokyoesque. Tokyoesque - Globally Aware & Curious. Recuperado 25 de abril de 2022, de <https://tokyoesque.com/ai-market-in-japan/>
- GREIG, J. (2021, 20 abril). How AI is being used for COVID-19 vaccine creation and distribution. TechRepublic. <https://www.techrepublic.com/article/how-ai-is-being-used-for-covid-19-vaccine-creation-and-distribution/>
- GUTIÉRREZ, F. (2022, 12 febrero). México tiene el potencial para 30 unicornios más. El Economista. Recuperado 21 de marzo de 2022, de <https://www.eleconomista.com.mx/sectorfinanciero/Mexico-tiene-el-potencial-para-a-30-unicornios-mas-20220211-0030.html>
- HABUKA, H. (2023). Japan's approach to AI regulation and its impact on the 2023 G7 presidency. Center for Strategic and International Studies. Retrieved from <https://www.csis.org/analysis/japans-approach-ai-regulation-and-its-impact-2023-g7-presidency>
- HENDRIX, J. (2021, 9 diciembre). Document: Senators Propose Platform Accountability and Transparency Act. Tech Policy Press. Recuperado 8 de marzo de 2022, de <https://techpolicy.press/document-senators-propose-platform-accountability-and-transparency-act/>

HERNÁNDEZ, D., & FITCH, A. (2021, 23 febrero). IBM's Retreat From Watson Highlights Broader AI Struggles in Health. WSJ. https://www.wsj.com/articles/ibms-retreat-from-watson-highlights-broader-ai-struggles-in-health-11613839579?mod=article_inline

Higa, E. (2024, June 16). Reconocimiento facial: pasajeros en Chiba pueden subir a trenes sin comprar boletos. International Press - Noticias de Japón En Español. https://internationalpress.jp/2024/06/17/reconocimiento-facial-pasajeros-en-chiba-pueden-subir-a-trenes-sin-comprar-boletos/?utm_source=chatgpt.com

HISTORY OF MACHINE LEARNING. (s. f.). <https://www.doc.ic.ac.uk>. Recuperado 6 de agosto de 2021, de <https://www.doc.ic.ac.uk/%7Ejce317/history-machine-learning.html>

HOFSTEDE INSIGHTS. (2021). Country comparison tool. <https://www.hofstede-insights.com/country-comparison-tool>

Horwitz, J. (2023, November 14). How Facebook went all in on AI. MIT Technology Review. <https://www.technologyreview.com/2023/11/14/1083336/how-facebook-went-all-in-on-ai/>

HOW GOOGLE USES COOKIES. (s. f.). Privacy & Terms at Google. Recuperado 13 de enero de 2023, de <https://policies.google.com/technologies/cookies>

Humble Bunny. (2025). Japan's Top Social Media Platforms for 2025 - 10th Edition. Recuperado de <https://www.humblebunny.com/japan-top-social-media-networks/>

IBM. (s.f.). ¿Qué es el aprendizaje no supervisado? Recuperado de <https://www.ibm.com/mx-es/topics/unsupervised-learning>

IBM. (s.f.). Inteligencia artificial explicable. Recuperado de <https://www.ibm.com/es-es/watson/explainable-ai>

IDHC. (2021). Los riesgos de la inteligencia artificial para los derechos humanos. Recuperado de <https://www.idhc.org/es/actualidad/los-riesgos-de-la-inteligencia-artificial-para-los-derechos-humanos.php>

INAI. (2022, 7 de abril). AIFA debe informar sobre sistema de datos biométricos, resolvió el INAI (Nota informativa No. INAI/038/22). Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales. <https://home.inai.org.mx/wp-content/documentos/SalaDePrensa/Comunicados/Nota%20INAI-038-22.pdf>

INESDI. (s.f.). Aprendizaje no supervisado: qué es y cómo usarlo. Recuperado de <https://www.inesdi.com/blog/aprendizaje-no-supervisado-que-es-y-como-usarlo/>

INTERNATIONAL AIRPORT REVIEW. (2024, March 1). Narita International Airport's facial I.D. tracking software. <https://www.internationalairportreview.com/article/184418/promoting-seamless-travel-at-narita-international-airport-with-face-express-one-id/>

INTERNATIONAL TELECOMMUNICATION UNION. (2022, 31 enero). About. AI for Good. Recuperado 28 de febrero de 2022, de <https://aiforgood.itu.int/about/>

Jiménez, E. C. (2024, November 21). Muere el INAI pero nace un nuevo órgano regulador de la competencia. El País México. <https://elpais.com/mexico/2024-11-21/muere-el-inai-pero-nace-un-nuevo-organo-regulador-de-la-competencia.html>

JUSTICE DEPARTMENT SUES GOOGLE FOR MONOPOLIZING DIGITAL ADVERTISING. (2023, 24 enero). <https://www.justice.gov/opa/pr/justice-department-sues-google-monopolizing-digital-advertising-technologies>

Karakaya, M. (2023). Osaka Station In Japan Pilots Facial Recognition-Based Ticketing System - IPVM Discussions. [ipvm.com](https://www.ipvm.com).

<https://ipvm.com/discussions/osaka-station-in-japan-pilots-facial-recognition-based-ticketing-system>

KLEINMAN, Z., & VALLANCE, C. (2023, 4 de mayo). AI 'godfather' Geoffrey Hinton warns of dangers as he quits Google. BBC News. <https://www.bbc.com/news/world-us-canada-65452940>

KNIGHT, W. (2020, 2 abril). What Marvin Minsky Still Means for AI. MIT Technology Review. <https://www.technologyreview.com/2016/01/26/163622/what-marvin-minsky-still-means-for-ai/>

Kyodo News. (2024). Japanese media say AI search infringes copyright, urge legal reform. Kyodo News. Retrieved from <https://english.kyodonews.net/news/2024/07/63195d98e74e-japanese-media-say-ai-search-infringes-copyright-urge-legal-reform.html>

LANZ, J. A. (2023). AI art wars: Japan says AI model training doesn't violate copyright. Decrypt. Retrieved from <https://decrypt.co/143461/ai-art-wars-japan-says-ai-model-training-doesnt-violate-copyright>

LAUNCH OF A PILOT TEST FOR AN AI FACIAL RECOGNITION Fare Payment System on Kumamoto City Trams: Hands-Free Travel with a Japan-first Technology, Enabling Fare Payment Without Physical Tickets on Trams. (n.d.). <https://www.marubeni.com/en/news/2023/info/00042.html>

LEFKOWITZ, M. (2015, 25 September). Professor's perceptron paved the way for AI – 60 years too soon. Cornell Chronicle. <https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon>

Legends, R. (Ed.). (2024, October 28). Echo Chamber Effect: How AI is Quietly Redefining Human Interaction. Medium.

<https://medium.com/@ronakkabra26/echo-chamber-effect-how-ai-is-quietly-redefining-human-interaction-744930a16320>

LET'S ASK SHOGO. (2020, 2 September). Why You MUST NEVER Stand Out in Japan [Video]. YouTube.

https://www.youtube.com/watch?v=893L3Vd86Pg&list=PLpIWoyf9KNFV_I-719B3pk673xaHZ6fYY&index=31

LISTED, N. (2023, 8 junio). Japan's PIPC warns ChatGPT about data collection, and other AI updates. International Association of Privacy Professionals. <https://iapp.org/news/a/japans-pipc-warns-chatgpt-on-data-collection-and-other-ai-updates/>

Riquelme, R. (2023, 14 julio). Marcelo Ebrard propone usar reconocimiento facial y morfológico en tareas de seguridad. El Economista. <https://www.eleconomista.com.mx/politica/Marcelo-Ebrard-propone-usar-reconocimiento-facial-y-morfologico-en-tareas-de-seguridad-20230710-0057.html>

LIU, L., & HAOHUI. (2020, 3 mayo). Adversarial Attacks in Machine Learning and How to Defend Against Them. Medium. Recuperado 18 de octubre de 2021, de <https://towardsdatascience.com/adversarial-attacks-in-machine-learning-and-how-to-defend-against-them-a2beed95f49c>

MAEDA, T. (2023, May 24). Seven Bank plans Japan's first facial recognition ATM service. Nikkei Asia. <https://asia.nikkei.com/Business/Finance/Seven-Bank-plans-Japan-s-first-facial-recognition-ATM-service>

Mainichi Shinbun. (2021, November 27). Do facial-recognition security cameras in Japan's stations really keep us safe? Mainichi Daily News. <https://mainichi.jp/english/articles/20211127/p2a/00m/0na/011000c>

Marcus, G. (2025, February 16). Elon Musk's terrifying vision for AI. Substack.com; Marcus on AI. <https://garymarcus.substack.com/p/elon-musks-terrifying-vision-for?triedRedirect=true>

Mares, M. A. (2025, February 17). Padrón de telefonía móvil, nueva intentona. El Economista.

<https://www.eleconomista.com.mx/opinion/padron-telefonía-movil-nueva-intentona-20250217-746631.html>

Marubeni. (2025). Launch of an AI Facial Recognition Boarding System for the Keisei Skyliner: Realization of a New System for Seamless, Comfortable, and Congestion-Free Airport Access. Marubeni.com.

<https://www.marubeni.com/en/news/2024/info/00060.html>

MARKETING4Ecommerce. (s.f.). Uber, Didi, Cabify, Beat: cuáles son las apps de movilidad disponibles en México. Recuperado de <https://marketing4ecommerce.mx/uber-didi-cabify-beat-apps-de-movilidad-disponibles-en-mexico/>

MARTÍNEZ, A. R. (2012, 16 octubre). El mercado más grande del mundo. chilango. <https://www.chilango.com/ciudad/la-ceda-el-mercado-mas-grande-del-mundo/>

Martinovs, D. (2019). 10 organizations leading the way in ethical AI. Sage Research Methods Community. <https://researchmethodscommunity.sagepub.com/blog/10-organizations-leading-the-way-in-ethical-ai>

Meta. (2024, April 18). Meet Your New Assistant: Meta AI, Built With Llama 3. Meta. <https://about.fb.com/news/2024/04/meta-ai-assistant-built-with-llama-3/>

McCURRY, J. (2021, 11 marzo). Japan marks 10 years since triple disaster killed 18,500 people. The Guardian. Recuperado 16 de marzo de 2022, de <https://www.theguardian.com/world/2021/mar/11/japan-marks-ten-years-since-triple-disaster-killed-18500-people>

Miller, A. (2023, August 4). 5 Things You Should Never Share with ChatGPT. AgileBlue.

<https://agileblue.com/5-things-that-you-should-never-share-with-chat-gpt/>

MOZUR, P., & The New York Times. (2016, 28 julio). Didi Chuxing and Uber, Popular in China, Are Now Legal, Too. Recuperado 18 febrero, 2020, de <https://www.nytimes.com/2016/07/29/business/international/china-uber-didi-chuxing.html>

MBN Staff. (2024, September 27). Mexico Ranks 7th in AI Adoption Among LATAM Countries: ECLAC. Mexico Business. <https://mexicobusiness.news/cloudanddata/news/mexico-ranks-7th-ai-adoption-among-latam-countries-eclac>

My Privacy Blog. (2024, October 2). The Privacy Implications of Meta AI: User Data and AI Integration Across Platforms. My Privacy Blog. <https://www.myprivacy.blog/the-privacy-implications-of-meta-ai-user-data-and-ai-integration-across-platforms/>

NATURE EDITORIAL. (2020, 2 diciembre). The race to the top among the world's leaders in artificial intelligence. Nature. Recuperado 3 de marzo de 2022, de https://www.nature.com/articles/d41586-020-03409-8?error=cookies_not_supported&code=6f75546d-f0bc-4f97-9fe5-5d2addf5fe79

NEC México. (2020, January). NEC habilita el reconocimiento facial en aeropuertos de Japón y el mundo. NEC. https://mex.nec.com/es_MX/press/PR/20200131005535_8100.html?utm_source=chatgpt.com

NEC. (2025). NEC introduce sistema de reconocimiento facial para pagos en tiendas y control de admisión en la Expo 2025 Osaka, en Japón. NEC. https://cl.nec.com/es_CL/press/PR/20240611215029_30320.html?utm_source=chatgpt.com

NETFLIX RESEARCH. (s. f.). Research Netflix. Recuperado 28 de septiembre de 2021, de <https://research.netflix.com/research-area/machine-learning>

NIGHTCAFE CREATOR. (s. f.). AI Generated Art Prints. Recuperado 23 de septiembre de 2021, de <https://nightcafe.studio>

Nippon.com. (2024, April 19). Japón destinará 72.500 millones de yenes a supercomputadoras para el desarrollo de inteligencia artificial. Nippon.com. https://www.nippon.com/es/news/yjj2024041900320/?utm_source=chatgpt.com

OFFICE FOR ARTIFICIAL INTELLIGENCE. (2021, 22 septiembre). National AI Strategy. GOV.UK. Recuperado 10 de marzo de 2022, de <https://www.gov.uk/government/publications/national-ai-strategy>

OLIVARES, E. (2021, 23 enero). Google Maps ya muestra rutas de microbuses en CDMX y Edomex. Código Espagueti. <https://codigoespagueti.com/noticias/internet/google-maps-rutas-de-microbuses-cdmx-y-edomex-android-ios/>

ONU NOTICIAS. Sesgos, racismo y mentiras: afrontar las consecuencias no deseadas de. (2021, 11 octubre).. Recuperado 18 de octubre de 2021, de <https://news.un.org/es/story/2021/01/1485802>

OPINNO. (2020, enero). LA INTELIGENCIA ARTIFICIAL EN LAS COMPAÑÍAS LATINOAMERICANAS Visión panorámica de la adopción y tendencias en la Región (N.o 1). MIT y Everis. https://www.anuarioseguros.lat/admin/storage/files/EVERIS_3.pdf

ORTEGA, A. (2021, 11 noviembre). Sociedad 5.0: el concepto japonés para una sociedad superinteligente. Real Instituto Elcano. Recuperado 31 de enero de 2022, de <https://www.realinstitutoelcano.org/analisis/sociedad-5-0-el-concepto-japones-para-una-sociedad-superinteligente/>

Oto, W. (2024, July 9). Why You Can't Disable the Shutter Sound on Japanese Phones. Tokyo Weekender. <https://www.tokyoweekender.com/japan-life/news-and-opinion/why-you-cant-disable-the-shutter-sound-on-japanese-phones/>

PALAFox, L. (2021, 14 diciembre). La batalla por la Inteligencia Artificial en México. El Financiero. Recuperado 21 de marzo de 2022, de

<https://www.elfinanciero.com.mx/opinion/leon-palafox/2021/12/14/la-batalla-por-la-inteligencia-artificial-en-mexico/>

Palafox, L. (2025, January 7). Los retos que tiene la IA en México para este 2025. El Financiero. <https://www.elfinanciero.com.mx/opinion/leon-palafox1/2025/01/07/los-retos-que-tiene-la-ia-en-mexico-para-este-2025/>

Pawlowicz, J. (2024, March 12). Social Media in Japan 2025 - everything you need to know. Digital Marketing for Asia. <https://www.digitalmarketingforasia.com/2025-social-media-in-japan/>

PHILLIPS, A. (2022, 6 marzo). fell in love with my AI girlfriend - and it saved my marriage. Sky News. Recuperado 22 de marzo de 2022, de <https://news.sky.com/story/i-fell-in-love-with-my-ai-girlfriend-and-it-saved-my-marriage-12548082>

PICHAU, S. (2023, 6 febrero). An important next step on our AI journey. Google. <https://blog.google/technology/ai/bard-google-ai-search-updates/>

PIEDRAS, E. (2021, 20 mayo). Internautas y Hábitos de Conectividad en 2020. The CIU. <https://www.theciu.com/publicaciones-2/2021/5/19/8y8yv7pogr3b6x8frdgeiibd23496u>

Presidencia de la Republica. (n.d.). Presidenta Claudia Sheinbaum firma iniciativa de reforma para asegurar derechos a trabajadores de plataformas digitales. Gob.mx. <https://www.gob.mx/presidencia/prensa/presidenta-claudia-sheinbaum-firma-iniciativa-de-reforma-para-asegurar-derechos-a-trabajadores-de-plataformas-digitales>

R3D: RED EN DEFENSA DE LOS DERECHOS DIGITALES. (2022, 26 abril). Pleno de la SCJN declara inconstitucional al PANAUT. Recuperado 29 de abril de 2022, de

https://r3d.mx/2022/04/25/pleno-de-la-scjn-declara-inconstitucional-al-panaut/?utm_campaign=meetedgar&utm_medium=social&utm_source=meetedgar.com

RAMOS, J. (2021, 24 julio). Mexicanos confían en inteligencia artificial para hallar el amor. El Sol de México| Noticias, Deportes, Gossip, Columnas. Recuperado 22 de marzo de 2022, de <https://www.elsoldemexico.com.mx/doble-via/mexicanos-confian-en-inteligencia-artificial-para-hallar-el-amor-7000237.html>

RAMOS, M. (2021, 23 febrero). Uso de redes sociales en México: 100 millones acceden al social media, la mayoría a Facebook (2021). Marketing 4 Ecommerce - Tu revista de marketing online para e-commerce. <https://marketing4ecommerce.mx/uso-de-redes-sociales-en-mexico/>

Ravichandran, H. (2024). Can My Identity Be Stolen With My Name and Date of Birth? Identityguard.com. <https://www.identityguard.com/news/can-my-identity-be-stolen-with-my-name-and-date-of-birth>

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de inteligencia artificial). (2024). Diario Oficial de la Unión Europea. <http://data.europa.eu/eli/reg/2024/1689/oj/spa>

Reuters. (2024, October 28). Toyota, NTT to make \$3.3 bln R&D investment for AI self-driving, Nikkei reports. Reuters. <https://www.reuters.com/technology/artificial-intelligence/toyota-ntt-make-33-bln-rd-investment-ai-self-driving-nikkei-reports-2024-10-28/>

Reuters. (2024, October 28). Toyota, NTT to make \$3.3 bln R&D investment for AI self-driving, Nikkei reports. Reuters.

<https://www.reuters.com/technology/artificial-intelligence/toyota-ntt-make-33-blnd-investment-ai-self-driving-nikkei-reports-2024-10-28/>

Riquelme, R. (s. f.). Recuperado 17 de julio de 2024, de <https://www.eleconomista.com.mx/tecnologia/Mercado-de-IA-en-Mexico-alcanza-ra-un-valor-de-3700-millones-de-dolares-en-2024-20240322-0048.html>

ROMÁN-VERGARA, J. (s. f.). No 01. Protección de Datos – México Transparente. La Revista Digital del SNT “México Transparente”. Recuperado 28 de abril de 2022, de https://snt.org.mx/revistadigital/?page_id=1113

ROMO, V. (2021, 5 octubre). NPR Cookie Consent and Choices. NPR. Recuperado 21 de octubre de 2021, de <https://choice.npr.org/index.html?origin=https://www.npr.org/2021/10/05/1043194385/whistleblowers-testimony-facebook-instagram>

SADOWSKI, J. (2021, 6 octubre). Facebook is a harmful presence in our lives. It's not too late to pull the plug on it. The Guardian. Recuperado 15 de octubre de 2021, de <https://www.theguardian.com/commentisfree/2021/oct/06/facebook-scandals-social-media>

Salgado Borge, J. El Congreso mexicano ante la inteligencia artificial. <https://www.proceso.com.mx/opinion/2024/4/18/el-congreso-mexicano-ante-la-inteligencia-artificial-327472.html>. Accedido 17 de julio de 2024.

Se, K. (2024, May 16). The Recipe for an AI Revolution: How ImageNet, AlexNet and GPUs Changed AI Forever And the unbending faith! A behind-the-scenes look at the unlikely ingredients that fueled the 2000s AI boom. TuringPost. <https://www.turingpost.com/p/cvhistory6%23geoffrey-hintons-lab-and-alex-nets->

SENADO DE LA REPÚBLICA. (2021, abril 20). Impulsan política de Estado para la investigación y desarrollo de la Inteligencia Artificial. Comunicación Senado. Recuperado 22 de marzo de 2022, de <http://comunicacion.senado.gob.mx/index.php/informacion/boletines/50757-impu>

[Isan-politica-de-estado-para-la-investigacion-y-desarrollo-de-la-inteligencia-artificial.html](#)

SINGER, P. (2018). The age of AI: How artificial intelligence is transforming our lives. TED.

SMITH, C. S. (2017, 23 junio). The Man Who Helped Turn Toronto Into a High-Tech Hotbed. The New York Times. <https://www.nytimes.com/2017/06/23/world/canada/the-man-who-helped-turn-toronto-into-a-high-tech-hotbed.html>

Sobre ania. (s. f.). Ania MX. Recuperado 10 de julio de 2024, de <https://www.ania.org.mx/sobre-ania>

SoftBank Group. (2025, February 6). OpenAI and SoftBank Group Partner to Develop and Market Advanced Enterprise AI | SoftBank Group Corp. SoftBank Group Corp. https://group.softbank/en/news/press/20250203_0

SPIKE, J. (Director). (2013). Her [Película]. Warner Bros. Pictures

Steffen, S. (2024, November 5). AI-driven bot network tried to help Trump win US election. Dw.com; Deutsche Welle. <https://www.dw.com/en/2024-us-election-ai-driven-social-media-bots-donald-trump-supporters/a-70695526>

STAFF, F. (2015, 14 mayo). Ya puedes consultar info de transporte público en Google Maps

STAFF, F. (2021, 11 noviembre). Récord Guinness para la CDMX por tener el mayor número de puntos wifi gratis. Forbes México. Recuperado 11 de noviembre de 2021, de <https://www.forbes.com.mx/record-guinness-para-la-cdmx-por-tener-el-mayor-numero-de-puntos-wifi-gratis/>

STATISTA. (2021). Tasa de adopción de internet en el mundo desde el año 2000 hasta el 2021. Recuperado de

<https://www.statista.com/estadistica/724276/tasa-de-adopcion-de-internet-en-el-mundo/>

STATISTA. (2021, 12 julio). Internet penetration rate Japan 2000–2019. Recuperado 15 de noviembre de 2021, de <https://www.statista.com/statistics/255857/internet-penetration-in-japan/>

Statista. (s.f.). Internet usage in Japan - statistics & facts . Recuperado de <https://www.statista.com/topics/2361/internet-usage-in-japan/#topicOverview>

Statista. (2023). Número de usuarios de redes sociales en Japón 2013-2028. Recuperado de <https://www.statista.com/statistics/278994/number-of-social-network-users-in-japan/>

Statista. (2023). Porcentaje de usuarios de redes sociales en México en 2023, por plataforma. Recuperado de <https://es.statista.com/estadisticas/1035031/mexico-porcentaje-de-usuarios-por-red-social/>

Steta, S. (2024, mayo 29). El Consejo de Europa adopta el primer tratado internacional en materia de Inteligencia Artificial. Santamarina + Steta. <https://www.santamarinasteta.mx/es/el-consejo-de-europa-adopta-el-primer-tratado-internacional-en-materia-de-inteligencia-artificial/>

Suzuki, I. (2024, September 22). Uber in Japan. East Asian Studies . <https://www.eastasianstudies.org/post/uber-in-japan>

Syscom: ISS. (s. f.). <https://www.syscom.mx/principal/listadopormarca/ISS>

Takenaka, K. (2025, March). AI robots may hold key to nursing Japan's aging population. The Japan Times. https://www.japantimes.co.jp/news/2025/03/01/japan/society/robots-japan-aging-society/?utm_medium=social&utm_source=bluesky#Echobox=1740814461

Tassi, P. (2024, May 29). Every way Twitter has gotten worse since it became 'X.' Forbes.

<https://www.forbes.com/sites/paultassi/2024/05/29/every-way-twitter-has-gotten-worse-since-it-became-x/>

TechAltar. (2025, February 16). China's city of the future: utopia or dystopia? YouTube. <https://www.youtube.com/watch?v=VltBlmIWE4U>

TECHCRUNCH is part of the Yahoo family of brands. (2023, 5 enero). [https://techcrunch.com/2023/01/05/apple-app-store-ad-targeting-eprivacy-breac h/](https://techcrunch.com/2023/01/05/apple-app-store-ad-targeting-eprivacy-breach/)

THE BUSSINESS SCHOOL. (2019, 27 junio). The TikTok Strategy: Using AI Platforms to Take Over the World. INSEAD Knowledge. [https://knowledge.insead.edu/entrepreneurship/the-tiktok-strategy-using-ai-platfo rms-to-take-over-the-world-11776](https://knowledge.insead.edu/entrepreneurship/the-tiktok-strategy-using-ai-platforms-to-take-over-the-world-11776)

THE GLOBAL AI INDEX. TORTOISE. (n.d.). Tortoise. <https://www.tortoisemedia.com/intelligence/global-ai/>

The Yomiuri 新聞. (2025, January 28). Facial Recognition Service Launched on Keisei Skyliner to Get Tickets; Railway Operator Aims to Ease... Yomiuri.co.jp; The Japan News. <https://japannews.yomiuri.co.jp/business/companies/20250128-235625/>

THIS PERSON DOES NOT EXIST. <https://www.thispersondoesnotexist.com>. (2021, 23 septiembre). <https://www.thispersondoesnotexist.com>

TIBCO SOFTWARE. (s.f.). ¿Qué es el aprendizaje no supervisado? Recuperado de <https://www.tibco.com/es/reference-center/what-is-unsupervised-learning>

TIMES, T. N. Y. (1958, 8 julio). New Navy Device Learns By Doing; Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser. The New York Times. <https://www.nytimes.com/1958/07/08/archives/new-navy-device-learns-by-doing-psychologist-shows-embryo-of.html>

Torres, M. (2024, mayo 20). Mexico's three big ai opportunities. Mexico Business.
<https://mexicobusiness.news/tech/news/mexicos-three-big-ai-opportunities>

TRAICY JAPAN INC. (2023, November 27). Osaka Metro to Install Face Recognition Gates at Namba Station, Expected to be Available to the Public by End of FY2024. TRAICY Global.
<https://en.traicy.com/posts/202311271887/#:~:text=Osaka%20Municipal%20Subway%2C%20also%20known,by%20the%20end%20of%20FY2024.>

Treasury Board. (2024). Algorithmic Impact Assessment tool - Canada.ca. Canada.ca.
https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html?utm_source=chatgpt.com

UBER MÉXICO. (s.f.). Aviso de privacidad [Archivo de internet]. Recuperado de <https://www.uber.com/legal/es/document/?name=privacy-notice&country=mexico&lang=es>

UBER. (s.f.). Uber México | Pide un viaje o regístrate como socio conductor. Recuperado de <https://www.uber.com/mx/es/>

Uber. (2021, 26 de agosto). How Data Shapes the Uber Rider App. Uber Blog. Recuperado de <https://www.uber.com/en-MX/blog/how-data-shapes-the-uber-rider-app/>

UNESCO. (2021). Ética de la inteligencia artificial. Recuperado de <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>

University of Helsinki. (2023, May 23). Elements of AI has introduced one million people to the basics of artificial intelligence | University of Helsinki. University of Helsinki.
<https://www.helsinki.fi/en/news/artificial-intelligence/elements-ai-has-introduced-one-million-people-basics-artificial-intelligence>

- UNIR REVISTA. (2022). ¿Qué es un sistema experto? Usos y aplicaciones en la IA. Recuperado de <https://www.unir.net/ingenieria/revista/sistema-experto/>
- UNIVERSIDAD UNADE. (s.f.). ¿Qué es la inteligencia artificial? Recuperado de <https://unade.edu.mx/que-es-la-inteligencia-artificial/>
- Velázquez, J. (2023, October 26). Detienen a estudiante del IPN que usó IA para “desnudar” fotografías de sus compañeras. Reporte Indigo. <https://www.reporteindigo.com/reporte/detienen-a-estudiante-del-ipn-que-uso-ia-para-desnudar-fotografias-de-sus-companeras/>
- VÉLIZ, C. [UNAM]. (2022, 25 abril). Conferencia con Carissa Véliz [Vídeo]. YouTube. <https://www.youtube.com/watch?v=dS6haod98Jc>
- Verma, H. (2024, January 28). Japan’s Approach to AI Regulation and Its Impact on the 2023 G7 Presidency - AI Transparency Institute. AI Transparency Institute. https://aitransparencyinstitute.com/japans-approach-to-ai-regulation-and-its-impact-on-the-2023-g7-presidency/?utm_source=chatgpt.com
- Vu, S. (2021, 16 julio). 5 Ways AI Will Transform the 2021 Tokyo Olympic Games. unlabeled. Recuperado 10 de noviembre de 2021, de <https://resources.unbabel.com/blog/5-ways-ai-will-transform-the-2021-tokyo-olympic-games>
- VVAA (2022, mayo 19). La brecha digital: El horizonte de las desigualdades. Gaceta UNAM. <https://www.gaceta.unam.mx/la-brecha-digital-el-horizonte-de-las-desigualdades/>
- Wa / The Government of Japan - JapanGov -. (2014). The Government of Japan - JapanGov -. <https://www.japan.go.jp/tomodachi/2014/summer2014/wa.html>
- Walker, K. (2024, June 26). 7 principles for getting AI regulation right. Google; Google.

<https://blog.google/outreach-initiatives/public-policy/7-principles-for-getting-ai-regulation-right/>

WIKIPEDIA CONTRIBUTORS. (2021, 26 mayo). Difference engine. Wikipedia.
https://en.wikipedia.org/wiki/Difference_engine#cite_note-:1-8

WRITER, S. (2023, July 7). Japan police to test AI-equipped cameras in protecting VIPs. Nikkei Asia.
<https://asia.nikkei.com/Politics/Japan-police-to-test-AI-equipped-cameras-in-protecting-VIPs>

XALTUS TECH. (2021, 8 junio). Countries leading the way in AI. Xaltius. Recuperado 3 de marzo de 2022, de <https://xaltius.tech/countries-leading-the-way-in-ai/>

YAMAMURA, A. & MCKINSEY & COMPANY. (2021, 14 julio). A more personalized world: An interview with the CEO of Tokyo Metro. McKinsey & Company. Recuperado 10 de noviembre de 2021, de <https://www.mckinsey.com/business-functions/operations/our-insights/a-more-personalized-world-an-interview-with-the-ceo-of-tokyo-metro>