



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN FILOSOFÍA DE LA CIENCIA
FACULTAD DE FILOSOFÍA Y LETRAS

**¿Cajas negras en cuartos blancos?
El papel de la transparencia en el uso de sistemas de Inteligencia Artificial
para la atención clínica en México**

TESIS
QUE PARA OPTAR POR EL GRADO DE
DOCTOR EN FILOSOFÍA DE LA CIENCIA

PRESENTA
DIEGO VÁZQUEZ DÍAZ

TUTORAS PRINCIPALES
DRA. LAURA GONZÁLEZ FLORES
IIE, UNAM
DRA. GISELA MATEOS GONZÁLEZ
CEIICH, UNAM

TUTORES
DR. JESÚS FAVELA VARA
CICESE, CONAHCYT

LECTORES
DR. JUAN PABLO GUTIÉRREZ
FACULTAD DE MEDICINA, UNAM
DR. JORGE ARTURO CERDIO HERRÁN
ITAM

CIUDAD UNIVERSITARIA, CDMX, ENERO 2024



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



**PROPUESTA UNIVERSITARIA DE INTEGRIDAD Y
HONESTIDAD ACADÉMICA Y PROFESIONAL
(Graduación con trabajo escrito)**

De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción 1, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la Institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado

que presenté para obtener el grado de _____ es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi programa de posgrado, citando las fuentes de ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia, acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad de los actos de carácter académico administrativo del proceso de graduación.

Atentamente

(Nombre, firma y Número de cuenta de la persona alumna)

ÍNDICE

INTRODUCCIÓN	5
---------------------------	----------

CAPÍTULO 1. Gobernanza digital internacional y marcos regulatorios en México	12
---	-----------

1.1 Aplicaciones de la Inteligencia Artificial en el área de la salud.....	14
---	-----------

1.2 La transparencia como directriz en la regulación de tecnologías inteligentes	19
---	-----------

1.2.1 Perú: <i>Ley 2775</i>	21
-----------------------------------	----

1.2.2 Ecuador: <i>Diagnóstico sobre la inteligencia artificial en el Ecuador</i>	22
--	----

1.2.3 Chile: <i>Política Nacional de Inteligencia Artificial</i>	23
--	----

1.2.4 Brasil: <i>Proyecto de Ley 2338</i>	23
---	----

1.2.5 OCDE: <i>Uso estratégico y responsable de la inteligencia artificial en el sector público</i>	24
---	----

1.2.6 UNESCO: <i>Recomendación sobre la inteligencia artificial</i>	25
---	----

1.2.7 OPS: <i>Sinopsis de políticas</i>	26
---	----

1.2.8 OMS: <i>Ética y Gobernanza de la Inteligencia Artificial para la Salud</i>	26
--	----

1.2.9 Estados Unidos: <i>Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan</i>	28
---	----

1.3 Actualidad en la gobernanza digital y políticas públicas en México	31
---	-----------

1.3.1 Agenda digital educativa	31
--------------------------------------	----

1.3.2 <i>Proyecto de decreto para expedir la Ley para la Regulación Ética de la Inteligencia Artificial para los Estados Unidos Mexicanos</i>	38
---	----

1.3.3 NOM-241.....	40
--------------------	----

CAPÍTULO 2. Perspectivas frente a la transparencia en el uso de sistemas de Aprendizaje Profundo	45
---	-----------

2.1 Transparencia técnica y estrategias para la promoción de la explicabilidad	48
---	-----------

2.1.1 Métodos de verificación y validación.....	50
---	----

2.1.2 Análisis de robustez	54
----------------------------------	----

2.1.3 Historia de las implementaciones exitosas/no exitosas	56
---	----

2.2 Limitantes y problemas en la búsqueda de transparencia técnica	58
2.3 Abrir las cajas negras: conocimiento tecnológico contra autoridad epistémica.....	63
2.3.1 Transparencia técnica y transparencia tecnológica	65
2.3.2 La IA como sistema tecnológico	70
2.3.3 La autoridad epistémica como límite para la transparencia tecnológica.	75
CAPÍTULO 3. Materiales para una regulación de sistemas de IA/AP en México	81
3.1 El principio precautorio como herramienta de previsión estatal	84
3.2 La gestión de riesgos como agencia de responsabilidad social.....	92
3.3 Limitaciones en el tratamiento del acceso a la información para una democratización digital ética	99
3.3.1 Daños a la privacidad	101
3.3.2 Daños a la propiedad intelectual	102
3.3.3 Creación de mayor opacidad epistémica	103
3.3.4 Otorgamiento de un exceso de agencia al paciente	103
CONCLUSIONES	106
POSFACIO	112
REFERENCIAS BIBLIOGRÁFICAS	115

No hay mejor abono para los riesgos que negarlos.

-Ulrich Beck

INTRODUCCIÓN

Este estudio está animado por la intención de suscitar una toma de conciencia del uso de la Inteligencia Artificial (IA) en la atención médica. Este mismo propósito ha sido una de las guías más importantes en las políticas públicas y marcos normativos que buscan regular el uso de las tecnologías inteligentes en las prácticas clínicas, en la investigación biomédica y en el manejo de los pacientes. El derecho al conocimiento y la búsqueda de transparencia como directrices para el uso ético de la IA han sido rectoras para estos esfuerzos, al mismo tiempo que se ha fomentado la idea de que la educación digital es el principal conducto para empoderar a los usuarios tecnológicos en la gestión de su propia salud. En *¿Cajas negras en cuartos blancos?* pretendo demostrar que estos esfuerzos corren el riesgo de resultar infructuosos en tanto parten de un equívoco que evidenciaré a lo largo del producto de mi investigación; a saber, que la transparencia en el uso de tecnologías de IA es sinónimo de conocimiento técnico.

Por un lado, podemos definir a los sistemas de Inteligencia Artificial como sistemas que pueden emular, aumentar o competir con el desempeño de humanos inteligentes en tareas específicas (Vallor y Bekey, 2017, p. 339). Complementariamente, podemos hablar de Aprendizaje Profundo (AP) cuando las tecnologías de aprendizaje automático son capaces de procesar información mediante transformaciones no lineales múltiples y redes neuronales multicapa para hacer más efectivo el entrenamiento y aprendizaje del sistema mediante el uso de grandes volúmenes de información (Ekmekci y Arda, 2020, p. 25). Estos últimos sistemas pertenecen al conjunto de tecnologías de Aprendizaje de Máquinas (AM), que se caracterizan por operar a través de modelos

lógicos, geométricos y probabilísticos, y que tienen como fin hacer que una computadora optimice sus procesos y mejore sus resultados mediante la experiencia, o, en otros términos, mediante el aprendizaje. Los árboles de decisión, las redes bayesianas y los algoritmos de agrupamiento son algunos ejemplos de técnicas utilizadas para cumplir con este propósito. Sin embargo, hoy en día son las redes neuronales las que dominan en el campo de la Inteligencia Artificial. Actualmente, estas tecnologías son el paradigma al hablar de Aprendizaje de Máquinas y su desarrollo ha permitido el surgimiento de algoritmos aún más complejos.

Como consecuencia, podemos asumir que esta diferencia de grado en la complejidad de los sistemas de Inteligencia Artificial se traduce, a su vez, en una mayor opacidad respecto a sus funcionamientos operativos. Mientras que algunos algoritmos, como los sistemas expertos o los árboles de decisión, son lo suficientemente simples para ser comprendidos a cabalidad por algunos agentes expertos, existen algunos otros, como los de Aprendizaje Profundo, que resultan opacos para todos sus usuarios. Debido a su uso extensivo y la capacidad que tienen para resolver tareas complejas, los agentes reguladores, los desarrolladores y la iniciativa privada se han visto conducidos a promover estrategias para que los usuarios finales de las tecnologías sepan utilizarlas y sean conscientes sobre cómo funcionan. Esta búsqueda ha sido enmarcada dentro de la aspiración por transparencia en el uso de la IA. Pero, naturalmente, esta búsqueda de transparencia ha encontrado como limitante las capacidades cognitivas de los agentes humanos para seguir la operatividad de las tecnologías, que procesan grandes volúmenes de información a un ritmo acelerado.

En paralelo a esta búsqueda de transparencia, al menos desde el año 2001, las prácticas médicas centradas en el paciente han formado parte de las agendas prioritarias de atención en los Estados Unidos gracias al trabajo del entonces Instituto de Medicina (ahora Academia Nacional de Medicina de los Estados Unidos) (Bau, et. al., 2019). Este enfoque afirma que todas las personas tenemos el derecho a formar parte de las

decisiones que afectan nuestra salud, lo que requiere que las personas estén educadas y tengan el soporte necesario para tomar decisiones y participar en la atención que le brindan las instituciones de salud (OMS, 2015). Esta perspectiva se vuelve especialmente importante al situarnos en un contexto de creciente necesidad por proteger el derecho a la información y cuando existe una tendencia por afirmar la búsqueda de prácticas médicas en las que las personas sean capaces de tomar decisiones informadas respecto a su salud. En materia de IA, esto tiene que ver con que el paciente está interesado en comprender cómo las tecnologías inteligentes afectan las decisiones clínicas en su caso específico (Ossa, L, et. al., 1022).

En el contexto latinoamericano, la discusión sobre la importancia de promover servicios centrados en las personas ha derivado en la adopción de este mismo principio como un objetivo básico de los sistemas de atención clínica, poniendo al empoderamiento de las personas en el núcleo de las prácticas médicas (Yudkin, 2022). Esta tendencia por ofrecer atención con énfasis en la protección al paciente ha formado también parte de las agendas en salud pública en México como principio de la alianza entre la Organización Panamericana de la Salud (OPS) y la Secretaría de Salud de México. Sin embargo, con el paso de los años el surgimiento de tecnologías complejas de Inteligencia Artificial, así como su implementación en las prácticas médicas, ha arrojado la pregunta sobre si realmente es posible que las personas tomemos decisiones informadas respecto a nuestra salud cuando esta se encuentra mediada por tecnologías inteligentes.

La implementación de sistemas de IA/AP en el área médica implica una intervención directa en aspectos fundamentales de la vida de los agentes humanos al comprometer la seguridad de los individuos y de la sociedad en su conjunto. El problema es que los agentes humanos involucrados raramente tienen la información necesaria para poder justificar la creencia de que los sistemas cumplen con sus propósitos.

En tanto la atención, los diagnósticos y el tratamiento dependen de tecnologías que superan nuestras capacidades cognitivas de comprensión, parecería que la exigencia por un entendimiento de las tecnologías para tomar decisiones sobre nuestra propia salud está limitada. A simple vista, esto podría sugerir que la aspiración por crear estrategias de atención centradas en las personas entra en conflicto con el desarrollo técnico de dispositivos médicos inteligentes. En este tenor, Bjerring y Busch (2020) han afirmado que una práctica médica centrada en el paciente es incompatible con los usos de tecnologías de IA/AP en el área de salud.

Concretamente, en *¿Cajas negras en cuartos blancos?* pretendo demostrar que cuando hablamos del uso de sistemas inteligentes para la administración de la salud —y, especialmente, de Aprendizaje Profundo— existe, en efecto, una limitante para que las personas puedan tomar decisiones informadas en los niveles de atención clínica debido a que estas tecnologías no son transparentes para sus usuarios. Pero defenderé que este factor de opacidad no es el desconocimiento técnico, sino una limitante paralela que designaré como desconocimiento tecnológico. Afirmaré, además, que las políticas públicas y las agendas digitales actuales no resuelven el problema que representa esta opacidad y que, por el contrario, necesitamos marcos normativos que logren articular la búsqueda de transparencia en el uso de sistemas inteligentes para la salud con un enfoque de atención médica centrada en los pacientes.

Para defender este argumento, en el primer capítulo, titulado *Gobernanza digital internacional y marcos regulatorios en México*, exploraré la actualidad de los marcos normativos y agendas digitales en el país en comparación con los acercamientos que los sectores líderes de la industria, gobiernos y academia a nivel mundial han propuesto como medios para contrarrestar el problema de la falta de transparencia en el uso de las tecnologías de IA/AP y frente a la tendencia por asegurar prácticas médicas centradas en los pacientes.

En el segundo capítulo, titulado *Perspectivas frente a la transparencia en el uso de sistemas de Aprendizaje Profundo* cuestionaré la viabilidad de

la búsqueda de transparencia técnica por parte de las regulaciones internacionales tanto por las limitaciones cognitivas de los usuarios como por la falta de infraestructura en el país. En contraposición a un enfoque de autoridad epistémica basado en el conocimiento experto, propondré a la sensibilización tecnológica como medio para asegurar que los usuarios cuentan con información suficiente para tomar decisiones informadas respecto a su salud. Esta propuesta será el resultado de la diferenciación entre los problemas técnicos y los problemas tecnológicos en el uso de la Inteligencia Artificial entendida como un sistema tecnológico.

Por último, en el tercer capítulo, titulado Vías para una regulación de sistemas de Aprendizaje Profundo en México, caracterizaré las necesidades de los usuarios tecnológicos en el país para presentar algunas consideraciones legales, éticas y epistemológicas de relevancia social que una normativa debería tomar en cuenta para la protección del derecho a la información en el territorio. En este apartado exploraré cómo los marcos normativos mexicanos pueden adaptarse a las nuevas tecnologías y qué recursos pueden ser implementados para su eficiente evaluación, manejo y uso.

A lo largo de esta investigación defenderé que la falta de educación tecnológica y de políticas que defiendan el acceso a la información son las principales barreras epistémicas que impiden a los usuarios de las tecnologías llevar a cabo una toma de decisión informada respecto a su salud. Este acercamiento querrá sacar a la luz la existencia de riesgos en el uso de las tecnologías inteligentes para la salud cuando los pacientes no cuentan con información suficiente para tomar decisiones informadas sobre su propia salud. Así como afirma Ulrich Beck (2000), defenderé que estos riesgos son la síntesis de conocimiento e inconsciencia, así como los productos de un tipo de opacidad producida por incertidumbres o riesgos manufacturados (14) por el funcionamiento mismo de las tecnologías, así como por el entramado social en el que surgen.

A partir de un enfoque tecnocrítico en el que examinaré qué elementos, partes y agentes forman parte del entramado de la Inteligencia Artificial

en el área de la salud, exploraré las causas que están detrás del problema de esta falta de transparencia y propondré la creación de marcos normativos e iniciativas regulatorias como el mejor medio para salvaguardar la integridad de los pacientes y con el fin de promover la toma de decisión informada.

Cabe mencionar que en este estudio centraré mi atención en los problemas específicos del uso de sistemas de Aprendizaje Profundo, aunque hablaré de las tecnologías bajo la denominación de Inteligencia Artificial. Esto se debe a que la terminología usada en los marcos regulatorios y la industria gira en torno a tecnologías inteligentes, sin diferenciar al Aprendizaje Profundo de la Inteligencia Artificial o de algunas otras tecnologías. Además, es importante mencionar las dificultades de comprensión que el uso de múltiples terminologías podría tener en los usuarios finales de las tecnologías, quienes usualmente designan a estos sistemas simplemente como “Inteligencia Artificial”. Al final del día, como afirman Michiel Fierens, Stephanie Rossello y Ellen Wauters, este término es utilizado para designar a una gran variedad de tecnologías, aunque actualmente solemos enfocarnos en implementaciones de Aprendizaje de Máquinas, Aprendizaje Profundo y Redes Neuronales al hablar de IA (2023, 53).

La pertinencia de esta discusión radica en la actualidad del problema del acceso a las tecnologías en México y de la falta de normativas específicas para la promoción del bienestar de nuestros contextos sociales. Además, la investigación busca figurar como un estudio pionero de las implicaciones sociotécnicas que conlleva la emergencia de nuevas tecnologías en el área de la salud. Es importante notar que, hasta ahora, no se tiene constancia de algún estudio pormenorizado sobre la pobreza digital y el analfabetismo computacional que ponga en relación a las comunidades tecnológicas del país con los sistemas de Inteligencia Artificial. Si bien ha habido importantes trabajos realizados en torno a la universalización de las tecnologías en telecomunicaciones (Mariscal, 2007) y respecto a las brechas informáticas en el país (Marsical, 2005,

2011; Toudert, 2014), hasta ahora han sido menos exploradas las diferentes dimensiones de inaccesibilidad a las tecnologías de Inteligencia Artificial en México.

Este estudio pretende contribuir, así, a la apertura de espacios que tengan como fin discutir los usos tecnológicos en nuestro circuito social para su regulación local, regional y global. En última instancia, la investigación tiene como objetivo contribuir, a través de una discusión filosófica y social sobre el problema de la transparencia, a la creación de normativas concretas que respondan a las necesidades regulatorias del contexto mexicano.

CAPÍTULO 1. Gobernanza digital internacional y marcos regulatorios en México

Hoy en día, las tecnologías de Inteligencia Artificial y de Aprendizaje Profundo están en todos lados. Estos sistemas permiten el funcionamiento de electrodomésticos y de teléfonos celulares, así como de los equipos de cómputo y asistentes virtuales que nos acompañan en nuestras tareas cotidianas. Además, permiten el efectivo funcionamiento de grandes empresas mercantiles y de distribución, de sistemas económicos y de servicios gubernamentales e, incluso, de herramientas y recursos utilizados en las áreas clínica y biomédica para la gestión de la enfermedad y la salud.

En este último rubro, el uso de chatbots para monitoreo y servicio, sensores inteligentes para la prevención, expedientes clínicos digitales y tecnologías para el diagnóstico asistido basado en imágenes, así como el desarrollo automatizado de fármacos y la asistencia a pacientes por medio de robots son algunas de las iniciativas que han impulsado la inclusión de diversos sistemas de Inteligencia Artificial en la práctica clínica, farmacéutica e, incluso, quirúrgica.

Estas tecnologías se encuentran disponibles en el mercado y es común encontrarlas en aplicaciones móviles, centros de salud privados y laboratorios. Pero a pesar de su gran utilidad para cumplir con funciones históricamente relegadas a los profesionales médicos, estos desarrollos en Inteligencia Artificial se han enfrentado constantemente al problema del desconocimiento sobre el modo en el que operan y en el que son utilizadas las tecnologías, así como el de las posibles consecuencias éticas, políticas y culturales que un desmedido o irresponsable uso de las mismas podría tener debido a la opacidad, o falta de transparencia, de su funcionamiento.

Las funcionalidades de estos algoritmos, sumadas al desarrollo de computadores con cada vez mayores capacidades para procesar más

datos y de forma más rápida, han superado las capacidades de comprensión de los agentes humanos, obstaculizando el conocimiento en el uso de las tecnologías y limitando el derecho de acceso al mismo. Debido en parte al analfabetismo digital, a las políticas de privacidad de las empresas tecnológicas y a las capacidades cognitivas de las mismas personas, actualmente nos encontramos rodeados de tecnologías que al estar condicionadas por cada vez más opacos y crípticos principios y mecanismos impiden que, en muchos casos, sus usuarios las comprendamos a cabalidad.

Es por ello que eliminar esa opacidad ha formado parte de las iniciativas para hacer un uso seguro de las tecnologías. En este tenor, “transparencia”, “explicabilidad” e “inteligibilidad” son términos que han figurado para designar a factores relevantes para la construcción de sistemas de IA confiables (Wortman y Wallach, 2020). Sin embargo, no existe un consenso sobre lo que estos términos significan ni sobre los requisitos que las tecnologías deberían cumplir para ser transparentes, explicables o inteligibles a las partes interesadas en el problema tecnológico del uso de sistemas de Inteligencia Artificial para la salud.

En este primer capítulo desarrollaré los conceptos relevantes al problema de la transparencia con el fin de establecer los parámetros que diversos marcos regulatorios internacionales y códigos de ética han propuesto para asegurar la transparencia en el uso de las tecnologías. Esto servirá como punto de partida para evaluar la actualidad de la gobernanza digital y las políticas públicas en México. Mi objetivo en este primer apartado es examinar si nuestros marcos normativos se ajustan a estas exigencias, cuáles son sus alcances y en qué situación se encuentra la gestión tecnológica y de infraestructura en el país.

Si bien esta investigación centra su atención en el uso de sistemas de Aprendizaje Profundo, una condicionante radica en que las leyes, así como los marcos conceptuales actuales, han adoptado como terminología a la Inteligencia Artificial para englobar a estos sistemas. Esto ocurre incluso cuando el tratamiento gira en torno al Aprendizaje de Máquinas y, con

mayor precisión, a los sistemas de Aprendizaje Profundo. Por ello resulta necesario delimitar a qué tecnologías haré referencia en esta investigación, aun cuando el objetivo sea que esta discusión se pueda extender a otras herramientas informáticas.

1.1 Aplicaciones de la Inteligencia Artificial en el área de la salud

En lo inmediato, podemos afirmar que las tecnologías de IA/AP pueden aumentar el valor de la industria de la salud al automatizar procesos y elevar la calidad del servicio a las personas. Esto se debe a las contribuciones que las tecnologías tienen para aumentar la velocidad de ejecución de las tareas asignadas, la reducción de costos y de complejidad y el fortalecimiento de la confianza (Ammanath, 2021).

En el área médica, estas contribuciones se hacen patentes, al menos, en el ámbito asistencial (diagnóstico, pronóstico, tratamiento, etcétera), en la salud pública (vigilancia epidemiológica y promoción de la salud), en la administración de las instituciones (para la optimización de recursos y gestión administrativa) y en la investigación biomédica (farmacología y ensayos clínicos, entre otros) (Miralles, 2021). Pero estos dispositivos involucran un amplio espectro de herramientas e instrumentos que trascienden las paredes de las instituciones de salud y que permean en nuestras vidas cotidianas, como son los relojes inteligentes, asistentes digitales y prótesis motorizadas, entre otros.

Un primer problema que representa el uso de las herramientas inteligentes y su aplicación en el área de la salud radica en que, en muchas ocasiones, los pacientes y usuarios finales no tienen conocimiento de si los dispositivos que utilizan cuentan o no con arquitecturas de IA/AP o si, en inicio, han sido utilizadas para el diagnóstico de sus enfermedades, en su tratamiento, o en el seguimiento de sus biométricos.

Hoy en día muchos de los recursos tecnológicos utilizados en estas áreas de atención funcionan a partir de algoritmos computacionales y de software, aunque no es el mismo caso cuando hablamos de sistemas de

Aprendizaje Profundo o de Inteligencia Artificial. Por ejemplo, el uso de computadoras es la norma para el manejo de expedientes médicos, para el procesamiento de imágenes radiológicas o incluso para la interacción con el paciente, pero solo en algunos casos estas tareas se ven optimizadas por el uso de sistemas inteligentes. Es por ello que resulta tan importante definir a qué tecnologías referiré en esta investigación.

Frente a este mismo reto, un equipo dirigido por Emilio Gómez-González, de la Universidad de Sevilla, llevó a cabo un estudio sobre los Niveles de Disponibilidad Tecnológica de la Inteligencia Artificial (NDT o TAL, por sus siglas en inglés) que describe, por un lado, el rango de disponibilidad al año 2020 de tecnologías de IA en el área de la salud reportadas en artículos de investigación, a la vez que, haciendo un estudio de las publicaciones en la materia, resumía el índice de certidumbre e impacto social de las tecnologías estudiadas [Tabla 1].

Gómez-González y su equipo aplicarían una numeración del 0 al 9 para definir la disponibilidad de las tecnologías de IA reportadas en su uso en el área médica y de la salud de acuerdo con los siguientes parámetros:

- NDT 0. Estado desconocido. No se considera factible.
- NDT 1. Estado desconocido. Se considera factible de acuerdo con referencias relacionadas o indirectas.
- NDT 2. Idea general o básica propuesta públicamente.
- NDT 3. Convocatorias abiertas para financiación pública de investigación, desarrollo e innovación (I+D+i).
- NDT 4. Resultados de proyectos parciales o académicos divulgados.
- NDT 5. Primeros diseños del producto divulgados.
- NDT 6. Prototipo operacional o 'primer caso' divulgado.
- NDT 7. Producto divulgado, pero no disponible.
- NDT 8. Disponible para ciertos usuarios (p.ej., profesionales).
- NDT 9. Disponible para el público.

Tecnologías de IA y mediadas por IA	Implementaciones específicas.	NDT	IS
Algoritmos para diagnóstico asistido por ordenador.	SW para soporte de decisiones en (la mayoría) de áreas clínicas.	8, 9	Positivo
Informes estructurados, eSalud.	SW para mejorar el flujo de trabajo, la eficiencia.	8, 9	
RA / RV, herramientas de imagen avanzadas.	Herramientas para visualización de información y navegación.	6, 7, 9	
	Cirugía guiada por imagen. Teleoperación	4, 6, 9	
Patología digital, "virtopsia".	SW para análisis automatizado, extensivo.	4-9	
Medicina de precisión, personalizada.	Tratamientos a medida. Predicción de respuesta.	4-9	
	Modelado y prueba <i>in-silico</i> . El "gemelo digital".	4-8	
	Diseño de fármacos.	4, 8	
<i>Apps, chatbots, dashboards</i> , plataformas en línea.	El "médico digital" (asistencia para profesionales y pacientes).	8, 9	
Robots sociales y de acompañamiento.	Para personas hospitalizadas, niños y ancianos.	4-9	
Recopilación y análisis de Big Data.	Epidemiología, prevención y seguimiento de brotes de enfermedades.	2-9	
	Detección de fraude. Control de calidad, seguimiento de médicos y tratamientos.	4-9	
IoT, <i>wearables</i> , mHealth.	Vigilancia clínica / sanitaria automatizada en cualquier entorno / institución.	7, 8	
	Monitorización, administración automatizada de medicamentos.	7-9	
Edición genética.	Prevención y tratamiento de enfermedades.	7, 8	
Fusión de datos médicos y sociales. Ingeniería social.	Prevención de episodios con relevancia clínica (p. ej. intentos de suicidio).	6, 8	
	Marketing a medida (p. ej., relacionado con los ciclos femeninos).	6, 8	
Lectura y decodificación de señales cerebrales. Interacción con procesos neuronales.	Tratamiento de enfermedades. Restauración de funciones dañadas.	3-8	
	Interfaces cerebro-máquina.	5-8	
	Control de prótesis, exoesqueletos, <i>cyborgs</i> .	2-7	
	Neuroestimulación. Neuromodulación	4-8	
	Neuroprótesis (para el sistema nervioso central).	2-5	
	"Lectura" y "manipulación" mental.	1-3	
Pruebas genéticas Cribado poblacional.	Test de enfermedades. Tests directos para el consumidor.	4-9	
Medicina de precisión, personalizada.	Selección individualizada. Moléculas personalizadas [para tratamiento] a precios "imposibles".	3-8	
Edición genética.	Humanos "diseñados"	2, 6	
	"Superhumanos" mejorados.	2	
	Medicina de autoexperimentación. <i>Biohacking</i> .	2, 6	
Sistemas de IA totalmente autónomos.	El "médico digital".	2-5	
	El "robot cirujano".	2, 4	
Embriones humano-animal.	Órganos para trasplantes.	2, 4, 5	
	Seres híbridos ("quimeras").	2, 4	
Búsqueda de la inmortalidad.	Emulación de todo el cerebro / "trasplante de cabeza".	1, 2	
Búsqueda de formas de vida artificiales.	"Máquinas vivas" ("robots biológicos", "biobots")	4, 6	
	Militares.	2, 3	
<i>Biohacking</i> malicioso.	Dirigido a individuos o grupos específicos.	1, 2	
[Conversión en] Armamento	Desde "pequeños laboratorios" hasta laboratorios militares.	1, 2	
Bioterrorismo.	En "pequeños laboratorios".	1, 2	
			Negativo

Tabla 1. Propuesta de clasificación de las aplicaciones de la inteligencia artificial (IA) y mediadas por la IA en Medicina y Salud de acuerdo con su impacto social (IS). SW = software, RA = realidad aumentada, RV = realidad virtual, IoT = internet de las cosas.

El estudio de Gómez-González deja ver que al hablar de Inteligencia Artificial o de Aprendizaje Profundo se abre un abanico muy amplio de tecnologías que pueden ser utilizadas en el área médica y de la salud, transitando desde la investigación biomédica hasta el acompañamiento de los usuarios sin intermediación de médicos expertos, como son el uso de wearables o robots sociales. Este amplio panorama arroja la pregunta sobre si las normativas deberían tener una flexibilidad tan alta para cubrir tan variados usos tecnológicos.

Complementariamente, de acuerdo con los datos presentados por el equipo financiado por el Centro de Estudios Avanzados del Joint Research Centre, podemos afirmar que el uso de sistemas inteligentes en el área de la salud no necesariamente nos compromete con un efectivo beneficio para los individuos que interactúan con estas tecnologías. Esto queda constatado en que el uso de las herramientas con fines bioterroristas o de armamento siguen desempeñando un papel como tecnologías con incidencia en la salud. Por ello resulta tan importante definir a qué tecnologías hacen referencias las normativas y con qué especificidad se puede promover la creación de estas regulaciones.

Una limitante para cumplir con esta tarea radica en que la *movilidad* del concepto de Inteligencia Artificial y la constante innovación tecnológica imposibilitan tener un recuento definitivo de lo que podemos considerar dentro del rubro de las tecnologías inteligentes. Además, es importante destacar que las normativas y marcos regulatorios no son explícitos respecto a qué tecnologías son consideradas dentro del rubro de la Inteligencia Artificial. Por el contrario, la designación dentro de esa categoría se suele aplicar después de las evaluaciones a las que se somete a cada sistema. Además, el estudio de Gómez-González y su equipo evidencia un universo extenso de tecnologías que pueden ser sujetas a los marcos normativos en materia de IA. En lo subsecuente partiré de esta categorización para tratar los problemas específicos de las tecnologías utilizadas en los niveles de atención al paciente, dejando de lado aquellos usos en materia de investigación biomédica, experimentación humana,

edición genética, marketing y gestión empresarial, epidemiología y diseño de fármacos, entre algunos otros.

En este tenor, trataré exclusivamente las posibilidades de los marcos regulatorios en torno al uso de la IA/AP en los niveles primario, secundario y terciario de atención. Esto significa que centraré mis esfuerzos en abordar a las tecnologías usadas en la atención, evaluación, diagnóstico y tratamiento del paciente. A pesar de que este universo es, a su vez, muy amplio y transita desde el uso de aplicaciones móviles que recuerdan a qué hora tomar un medicamento, hasta la creación de neuroprótesis, todas estas tecnologías tienen en común una influencia directa en el seguimiento del paciente, lo que demanda acceso a la información en su uso.

Con la clasificación de Gómez-González en mente (e incluso cuando el argumento de esta investigación se pueda extender a otros rubros) podemos delimitar los objetivos de esta propuesta para las siguientes tecnologías:

- [1] Algoritmos para el diagnóstico asistido e inteligente, en donde se encuentran los sistemas de procesamiento de imágenes radiológicas e imagenología, diagnóstico de padecimientos basado en modelos probabilísticos y software para detección de anomalías cutáneas u oculares, entre otros.
- [2] Informes estructurados para la gestión de eSalud, que incluye a los sistemas de procesamiento de la información para crear reportes para el paciente y para el médico en torno a una decisión médica.
- [3] Herramientas para la visualización de la información en formato de realidad aumentada o virtual.
- [4] Tecnologías de teleoperación e intervención quirúrgica por imagen, así como robots cirujanos autónomos.
- [5] Sistemas de predicción de respuesta a tratamientos y fármacos.
- [6] Modelado del paciente para creación de prótesis o prueba de tratamientos.

- [7] Médicos digitales y sistemas de atención al paciente, como chatbots, modelos de lenguaje, interfaces de servicio o aplicaciones móviles para el seguimiento de tratamiento.
- [8] Robots sociales y de acompañamiento.
- [9] Internet de las cosas.
- [10] Wearables y aplicaciones de seguimiento y atención diseñadas para estos dispositivos.
- [11] Análisis de pruebas de laboratorio.

Es importante mencionar que algunas de estas tecnologías, ya sea por su novedad o por controversias en su uso, actualmente no están disponibles para los usuarios finales ni para los profesionales de la medicina, por lo que tienen un nivel de disponibilidad tecnológica bajo. Sin embargo, la rápida evolución y adopción de los sistemas inteligentes en el área de la salud demanda normativas preparadas para los retos tecnológicos del futuro y su posible implementación. Por ello, diversos países se han pronunciado en torno a la necesidad de regular el uso de estas tecnologías.

Con este marco conceptual en mente, en la siguiente sección abordaré algunos casos representativos de marcos normativos que se han implementado para el uso ético y seguro de la IA en el área de la salud, así como el papel que ha figurado la búsqueda de transparencia en estas regulaciones.

1.2 La transparencia como directriz en la regulación de tecnologías inteligentes

Todos los sistemas de Inteligencia Artificial son en mayor o en menor medida opacos. Esto significa que no existe una completa transparencia cuando hablamos de sistemas informáticos de este tipo, sino que siempre existen limitantes que impiden conocerlas a cabalidad, ya sean políticas de propiedad privada, falta de conocimientos técnicos o incapacidades cognitivas para seguir el ritmo de sus computaciones. En inicio, esto podría

poner en cuestión si la búsqueda por transparencia debe ser promovida. Sin embargo, como afirma la FDA, es justamente en el uso de la IA/AP que el tema de la transparencia adquiere mayor sentido, ya que hablamos de tecnologías que pueden aprender y cambiar con el tiempo, por lo que los usuarios deben saber cuál es el nivel de opacidad al que se pueden enfrentar en la interacción con estos algoritmos (2021, 5).

Esta opacidad ha llamado especialmente la atención de los organismos reguladores y de las agencias defensoras de los Derechos Humanos. La organización Artículo 19 ha afirmado, por ejemplo, que existe y debe promoverse una libertad de conocimiento que consiste en “el derecho a demandar y recibir información de los ostentadores del poder para la transparencia en la buena Gobernanza y para el desarrollo sostenible” (2019, p. 2). La misma organización ha hecho evidente que esta opacidad e inescrutabilidad ha dado pie a diversos esfuerzos para hacer a las tecnologías de IA transparentes. Sin embargo, como afirma Vidushi Marda (2017), en realidad no existe un consenso respecto a las formas de transparencia que debemos buscar. Es por ello que las diversas regulaciones de tecnologías inteligentes han propuesto diferentes vías para enfrentar este reto: desde la capacitación técnica de los profesionales hasta la apertura de las bases de datos que alimentan a los sistemas.

En materia de salud, al día de hoy los principales esfuerzos regulatorios de relevancia para los marcos normativos mexicano son el *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan* de la Administración de Alimentos y Medicamentos (FDA) en los Estados Unidos, la guía para la *Ética y Gobernanza de la Inteligencia Artificial para la Salud* de la Organización Mundial de la Salud (OMS), la *Sinopsis de políticas de la Organización Panamericana de la Salud (OPS)* y la *Recomendación sobre la inteligencia artificial de la UNESCO*. Los últimos cuatro documentos forman parte de las políticas internacionales para la promoción de los usos éticos de la IA en los que las delegaciones mexicanas han participado como firmantes, mientras que la política de la FDA forma parte de la NOM-241.

Sin embargo, existen algunos otros marcos regulatorios internacionales que tratan directa o indirectamente el problema de la transparencia en el sector salud, tales como *Uso estratégico y responsable de la inteligencia artificial en el sector público* de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), el *Proyecto de Ley 2338* del Senado Federal de Brasil, las *Directrices éticas para una IA fiable* de la Comisión Europea, la *Ley 2775* del Congreso de la República del Perú, la *Política Nacional de Inteligencia Artificial* de Chile y el *Diagnóstico sobre la inteligencia artificial en el Ecuador* de la Dirección de Fomento de Tecnologías Emergentes.

En lo subsecuente, abordaré las contribuciones de cada una de estas normativas, así como el rol que ha desempeñado la transparencia en ellas. Dejaré al final las regulaciones propuestas por la UNESCO, la OPS, la OMS y la FDA por su pertinencia, cercanía y aplicabilidad dentro del contexto mexicano.

1.2.1 Perú: *Ley 2775*

Uno de los primeros esfuerzos regulatorios en materia de IA en la región latinoamericana fue el impulsado por el Congreso de la República del Perú en el 2022. La propuesta de ley fue decretada por la Comisión de Ciencia, Innovación y Tecnología en agosto del mismo año y tiene como objetivo “fomentar [el] desarrollo económico y social del país, en un entorno seguro que garantice su uso ético, sostenible, transparente, replicable y responsable” (2023, 3) bajo seis principios:

- [1] Seguridad supervisada
- [2] Enfoque de pluralidad de participantes
- [3] Gobernanza de Internet
- [4] Sociedad digital
- [5] Desarrollo ético para una IA responsable
- [6] Privacidad de la IA

En relación con la transparencia, la ley solamente la promueve en su octava sección, al establecerla como un beneficio para las personas. En este apartado se relaciona a la transparencia con la sostenibilidad, la replicabilidad y la responsabilidad. Asimismo, la ley pone en juego a la explicabilidad como un medio para verificar el funcionamiento de los sistemas, aunque no establece cuáles son sus parámetros.

De esta ley, destaca su posicionamiento respecto al funcionamiento técnico de las tecnologías inteligentes, pues apuesta por la verificación de seguridad como medio para hacer un uso responsable de los sistemas. En contraste, no establece parámetros claros al hablar de transparencia ni explicabilidad y solo menciona a la salud como ejemplo de implementación de la IA, aun cuando afirma que es uno de los sectores prioritarios de la ley.

1.2.2 Ecuador: *Diagnóstico sobre la inteligencia artificial en el Ecuador*

Por su parte, el Ministerio de Telecomunicaciones y de la Sociedad de la Información de la República de Ecuador impulsó en el 2021 un diagnóstico sobre la IA como base para la creación de una *Estrategia de Inteligencia Artificial* en el país, que no ha sido publicada al 2023.

Si bien este estudio no es una normativa, sí delinea los intereses de la estrategia nacional para enfrentar los retos de la IA en sus diferentes aplicaciones. Además, promueve la transparencia y explicabilidad mediante el compromiso por explicar los resultados de la IA, aunque únicamente comprende este proceso como la facilidad de interpretación de los resultados de un sistema.

Esta propuesta parece proponer a la formación de expertos para la validación tecnológica o bien la dependencia de otras herramientas de verificación para asegurar a la transparencia técnica. Más adelante veremos las complicaciones que derivan de esta aspiración.

1.2.3 Chile: *Política Nacional de Inteligencia Artificial*

Chile, por el contrario, es uno de los países con mayor actividad en el manejo de la Inteligencia Artificial desde un punto de vista legislativo. En el 2021, el Ministerio de Ciencia, Tecnología, Comunicación e Innovación promulgó una *Política Nacional de Inteligencia Artificial* que, a pesar de no estar dirigida a un sector de aplicación en concreto, sí establece pautas para el manejo ético de las tecnologías.

En su Objetivo 3.1.2, la política busca impulsar la transparencia algorítmica, aunque establece que el Ministerio será el encargado de elaborar las recomendaciones que servirán a las empresas para identificar sesgos y para pilotear sus tecnologías en áreas de riesgo (2021, 53).

Por otro lado, la política establece a la educación como uno de los pilares para el uso de la IA en el país, apelando a la reformulación y monitoreo del currículo escolar para incorporar habilidades necesarias para el desarrollo de la IA (2021, 25). De este modo, la normativa únicamente considera la adquisición de habilidades técnicas para el desarrollo de tecnologías inteligentes y uso de datos.

Esta política forma parte de la legislación estatal de Chile, lo que la destaca de entre otros acercamientos en la región que únicamente han propuesto programas normativos. Contraria a las intenciones regulatorias o programas de ley, esta política establece lineamientos estratégicos claros para la derivación de leyes para diferentes sectores y aplicaciones tecnológicas.

1.2.4 Brasil: *Proyecto de Ley 2338*

En el cono sur destaca también el ejercicio realizado por el Senado Federal de Brasil, que en 2023 publicó una disposición sobre el uso de la Inteligencia Artificial como proyecto de ley. Este documento se diferencia del caso de Chile debido a que no forma parte de una política pública ratificada, sino que, hasta ahora, presenta solamente las pautas para la creación de un marco normativo de ley.

En este caso, la exigencia de transparencia forma parte de los fundamentos en el uso de la IA, y se ve acompañada por otros conceptos relacionados, como la explicabilidad y la inteligibilidad, pero destaca en el uso de la auditabilidad como parte de las estrategias para la transparencia. En este tenor, el proyecto de ley establece la importancia de ofrecerla cuando se habla de sistemas cuyo uso conlleva un alto riesgo o cuando pueden violar la seguridad de los usuarios.

Pero a diferencia de los casos de Perú, Ecuador y Chile, el *Proyecto de Ley 2338* es el primero en establecer la transparencia en dos sentidos. No solamente la trata desde un punto de vista técnico, sino que busca fomentar la transparencia en la gobernanza. En su artículo 19, el proyecto establece que se debe promover la transparencia “sobre las medidas de gobierno adoptadas en el desarrollo y uso de sistemas de inteligencia artificial” (2023, 13). Esta iniciativa deja ver un interés por pensar la transparencia no solo desde un punto de vista computacional o algorítmico, sino como una búsqueda de concientización tecnológica y política.

1.2.5 OCDE: Uso estratégico y responsable de la inteligencia artificial en el sector público

La OCDE cuenta con dos documentos que funcionan como recomendaciones para los países miembros en el manejo ético de la IA: *La inteligencia artificial en la sociedad*, de 2019, y *Uso estratégico y responsable de la inteligencia artificial en el sector público de América Latina y el Caribe*, de 2022.

Es en esta última en donde la OECD establece que la IA fiable debe poseer la capacidad de “explicar sus decisiones” como transparencia a los fines de una evaluación externa (2022, 79). Una limitante es que cada agente puede establecer qué es lo que exige al hablar de transparencia y que en estas recomendaciones la organización únicamente toma en consideración que la IA debe ser explicable para ser transparente. Esto

significa que si un agente entiende cómo funciona una máquina, no existe opacidad en su uso, restringiendo la transparencia al dominio técnico.

Por otro lado, la OECD también resalta la importancia de la especialización profesional como acompañamiento para mejorar el desarrollo de las tecnologías. De este modo, la deserción escolar es presentada como una limitante para el uso de la IA en la región.

1.2.6 UNESCO: *Recomendación sobre la inteligencia artificial*

Dentro del conjunto de marcos regulatorios para la Inteligencia Artificial que no se enfocan en el área de la salud, podemos destacar la Recomendación sobre la ética de la inteligencia artificial de la UNESCO, emitida en el 2021 y que ha destacado por un abordaje comprehensivo de sus usos en diferentes sectores.

Así como en los documentos antes descritos, la UNESCO pone a la transparencia y explicabilidad como ejes rectores de un uso ético de la Inteligencia Artificial. Esta es la única recomendación que es contundente en la forma en que se debe promover la transparencia, pues afirma:

Las personas deberían estar plenamente informadas cuando una decisión se basa en algoritmos de IA o se toma a partir de ellos, en particular cuando afecta su seguridad o a sus derechos humanos; en estas circunstancias, deberían tener la oportunidad de solicitar explicaciones e información al actor de la IA o a las instituciones del sector público correspondiente. (2021, 22)

Además, esta recomendación separa a la transparencia de la explicabilidad, poniendo a esta última del lado de la inteligibilidad del funcionamiento del sistema técnico, mas no sobre su uso e implementación en áreas específicas, como es la salud.

Por otro lado, la UNESCO no asume un papel educativo únicamente técnico, pues no solo promueve la alfabetización digital entre los usuarios, sino su sensibilización para conocer el impacto de las tecnologías. Sin embargo, en el apartado dedicado a las recomendaciones en el área de la

salud y el bienestar social, la Organización apela al manejo ético de los Estados Miembros en el uso de la IA para mejorar la salud humana, poniendo énfasis en la salud pública y la epidemiología, pero sin destacar la necesidad de empoderar a los pacientes respecto a su propia salud.

Incluso, la recomendación pone al conocimiento experto como tomador de decisión en materia de salud. Si bien esto promueve el uso tecnológico centrado en las personas, ignora la tendencia por crear prácticas médicas centradas en el paciente.

1.2.7 OPS: *Sinopsis de políticas*

En el 2023, la Organización Panamericana de la Salud promulgó una sinopsis de políticas titulada *Inteligencia Artificial*, en donde enlista algunas recomendaciones para el correcto manejo de la IA. En este estudio no se menciona la transparencia, salvo al afirmar que:

La falta de transparencia en los procesos de adquisición, transformación y explotación de datos y la necesidad de establecer normativas para el uso clínico y extrahospitalario de la información adquirida para pacientes, personal médico y profesionales de otras disciplinas, son retos complejos que deben abordarse para garantizar una IA en materia de salud que respete los derechos de las personas. (2023, 6)

Pero a pesar de esta iniciativa, únicamente se establece que la creación de evidencia sobre la aplicación de los sistemas puede fomentar su evaluación y que un objetivo rector debe ser la formación de personal calificado en materia de IA.

1.2.8 OMS: *Ética y Gobernanza de la Inteligencia Artificial para la Salud*

Por su parte, frente a los peligros del uso no regulado de tecnologías computacionales, el departamento de Salud Digital y el de Innovación e Investigación para la Salud de la Organización Mundial de la Salud (OMS)

trabajó durante 18 meses en la creación de una guía para la *Ética y Gobernanza de la Inteligencia Artificial para la Salud*. En ella, 21 expertos externos con formación clínica y biomédica concluyeron que la entonces futura implementación de las tecnologías de IA en el área médica dependería, en gran medida, de un esfuerzo colectivo para diseñar e implementar leyes y políticas bioéticamente defendibles en favor de los intereses de los pacientes y de las comunidades (2021, xi). En el mismo documento, la Organización apuntó que la carencia de estas normativas en la mayor parte de los países y contextos sociales en el mundo era un llamado para su creación.

De acuerdo con la OMS, estos marcos legales deberían considerar, al menos, seis factores de relevancia bioética para la implementación de las tecnologías de IA en el área médica:

- [1] La protección de la autonomía, que implica el aseguramiento de que las personas pueden tomar decisiones y actuar incluso cuando las máquinas hayan ofrecido un resultado.
- [2] La promoción del bienestar humano y de los intereses públicos, que conlleva el cumplimiento de estándares de seguridad y controles de calidad para proteger al usuario ante posibles riesgos.
- [3] El aseguramiento en la transparencia, explicabilidad e inteligibilidad, ofreciendo al usuario la información necesaria para abrir el debate en torno a cómo deberían ser utilizadas estas tecnologías.
- [4] El afincamiento de responsabilidades y rendición de cuentas, que esté acompañado por mecanismos para identificar obligaciones e incumplimientos cuando exista un mal funcionamiento en los algoritmos.
- [5] La garantía en la inclusividad e igualdad, que aspira a suprimir sesgos de género, raciales y socioeconómicos para asegurar el acceso a las tecnologías y su correcto funcionamiento para todos.

- [6] La búsqueda de IA sensibles y sostenibles, con el compromiso de reducir su impacto en el ambiente, en la oferta laboral disponible y en los sistemas de salud en general.

Cada uno de estos factores implica una serie de consideraciones sociales, filosóficas y legales que, de acuerdo con la Organización, deben ser exploradas no a partir de presupuestos éticos universales sino de acuerdo con los contextos sociales en los que las tecnologías serán implementadas. No obstante, tras dos años de su publicación, países como México, que han ratificado la autoridad directiva y coordinadora de la OMS en materia de salud (Secretaría de Relaciones Exteriores, 2022), siguen sin contar con un marco normativo para enfrentar los retos que conlleva el uso de estas tecnologías.

1.2.9 Estados Unidos: *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan*

Por último, es necesario mencionar el marco normativo de los Estados Unidos debido a que las políticas públicas en México relacionadas con la intersección entre Inteligencia Artificial y salud se basan en las regulaciones de la Administración de Alimentos y Medicamentos (FDA). Por su relevancia en el contexto mexicano y debido a que es la regulación más consistente y completa a nivel global, vale la pena conocer sus orígenes y alcances al día de hoy.

En febrero de 2020 la FDA autorizó la venta del primer software de ultrasonido cardíaco que mediante un conjunto de redes neuronales artificiales guiaba al usuario en la manipulación del dispositivo para que el sistema hiciera una evaluación de manera automática el material visual. El software, llamado *Caption Guidance*, era capaz de indicar el mejor método para llevar a cabo la toma de un video de ultrasonido cardíaco que, posteriormente, sería interpretado por un experto cardiólogo.

De acuerdo con Robert Ochs, directivo del Centro para Dispositivos y Salud Radiológica de la Administración Norteamericana, la autorización de

venta de este dispositivo permitía su utilización por parte de médicos no expertos en ultrasonografía, de enfermeros de una clínica de asistencia familiar o de otros usuarios. Esto se volvía especialmente importante ya que “demostraba el potencial de la Inteligencia Artificial y de las tecnologías de Aprendizaje Profundo para incrementar el acceso a seguros y efectivos diagnósticos cardíacos que podrían salvar la vida de pacientes” (2020). Sin embargo, la autorización de este dispositivo no estuvo regulada por una normativa adaptada a los riesgos que implica el uso de la IA/AP en el manejo de la salud.

La evaluación de la seguridad y efectividad del software utilizado como dispositivo médico (SaMD por sus siglas en inglés) ha sido, al menos desde el año 2013, una de las áreas de discusión más importantes para la Administración de los Estados Unidos. Para la FDA, el SaMD es definido como software que pretende ser utilizado por uno o más propósitos médicos y que no es esencial para que un hardware utilizado como dispositivo médico cumpla con sus propósitos. Esta definición general, así como los lineamientos de certificación de los SaMD, habrían sido los elementos utilizados para la evaluación en la seguridad y efectividad de *Caption Guidance* como dispositivo médico, pero no sería sino después de su autorización que se haría patente la necesidad por crear regulaciones diseñadas específicamente para la evaluación de tecnologías de Inteligencia Artificial. Esto hizo que la FDA impulsara la creación de manuales y regulaciones hechas a la medida de las tecnologías inteligentes para la salud.

Después de la publicación de esta autorización federal, entre el 25 y 26 de febrero de 2020, se llevaría a cabo uno de los primeros talleres sobre el rol de la Inteligencia Artificial en imagenología (The Academy for Radiology & Biomedical Imaging Research, 2020). Sólo entonces la FDA plantearía la discusión sobre las “buenas prácticas” para la validación del uso de software de Inteligencia Artificial en materia radiológica. En su reporte, la Administración afirmaría que esta primera autorización en la venta de *Caption Guidance* representaba la creación de una nueva

clasificación regulatoria. Casi un año después, la FDA presentaría el primer plan de acción en la regulación del software basado en Inteligencia Artificial/Aprendizaje Profundo como Dispositivo Médico.

En este plan de acción, la FDA ha afirmado que una de las grandes áreas de oportunidad que debe cubrir una agenda en materia tecnológica y de salud pública es promover objetivos centrados en el paciente que incorporen la búsqueda de transparencia a los usuarios. Algunas otras recomendaciones afirman que esto solo es posible mediante el aviso a los pacientes de temas relevantes a su salud de una forma simple y clara. Esto incluye su diagnóstico, la recomendación médica, los riesgos que implica la realización de un procedimiento, así como su omisión, los efectos físicos, emocionales, mentales o sociales, entre otros, que conlleva la decisión y otras consideraciones previas (Gobierno de Queensland, 2017, 9). Pero la guía de la FDA no menciona ninguno de estos elementos.

De acuerdo con el diálogo público sostenido por la FDA previo a la publicación de su plan, diversas partes interesadas han expresado la necesidad de que los desarrolladores de las tecnologías describan de manera clara la información que ha sido utilizada para el entrenamiento de los algoritmos, la relevancia de sus valores de entrada (inputs), las lógicas que utilizan (cuando sea posible), el papel que se espera que los valores de salida (outputs) representen y la evidencia del desempeño de los dispositivos (2021, 5).

A partir de estas exigencias, la FDA ha expresado su interés por promover la transparencia de las tecnologías por parte de la industria y de sus desarrolladores en aras de asegurar que los usuarios entiendan los beneficios, riesgos y limitaciones de los dispositivos, por ejemplo, mediante su etiquetado (labeling). Sin embargo, al día de hoy no existe un documento o estrategia clara sobre cómo debería llevarse a cabo este proceso.

1.3 Actualidad en la gobernanza digital y políticas públicas en México

Al 2023 en México no existen marcos regulatorios nacionales diseñados para enfrentar los retos que implican las tecnologías de Inteligencia Artificial. Además, los esfuerzos en materia legislativa no obedecen a una agenda tecnológica integral, por lo que muchos esfuerzos regulatorios únicamente sirven para sectores específicos, como en el caso de la salud. Por el contrario, el principal acercamiento al tema por parte de las políticas públicas ha girado en torno al papel de la educación como medio para asegurar el desarrollo de infraestructura tecnológica, la democratización de las Tecnologías de la Información y de la Comunicación (TIC) y la adquisición de habilidades computacionales.

1.3.1 Agenda digital educativa

En el país, la mayor parte de los planes de acción han promovido la idea de que la educación en materia computacional podría solventar la carencia de conocimientos para usar, gestionar, entender y evaluar a las tecnologías digitales. La Comisión Nacional de Derechos Humanos (CNDH), por ejemplo, ha señalado que la desigualdad e inequidad en el acceso a las tecnologías solamente puede ser contrarrestado mediante una alfabetización digital que funcione como herramienta para el desarrollo económico y social (2018). La elevación de los niveles de escolaridad y el reconocimiento de la necesidad por modificar los modelos educativos en relación con los nuevos estándares internacionales en materia tecnológica y de derechos humanos son las vías propuestas para erradicar las desigualdades educativas en el país. Sin embargo, la Comisión no ha propuesto mayores medidas prácticas para la implementación de este plan.

Complementariamente, la UNESCO ha promovido la idea de que los encargados de la formulación de políticas públicas deben atender el fenómeno del analfabetismo digital no sólo mediante la inclusión de las

tecnologías computacionales, sino también de IA. Esto se presenta como un medio para el aprendizaje que se ajusta a la rápida evolución de las tecnologías, así como un objeto mismo de estudio. En consonancia con esta invitación diversos países asumieron en el Consenso de Beijing del 2019, enfocado en crear recomendaciones para el correcto aprovechamiento de la Inteligencia Artificial para la Agenda de Educación 2030, la tarea de:

Promover las respuestas políticas adecuadas para lograr la integración sistemática de la inteligencia artificial y la educación, a fin de innovar la educación, la docencia y el aprendizaje, y para que la inteligencia artificial contribuya a acelerar la consecución de unos sistemas educativos abiertos y flexibles que permitan oportunidades de aprendizaje permanente, equitativo, pertinente y de calidad para todos (UNESCO, 2019, 29)

A pesar de la presencia de representantes nacionales durante la Conferencia en Inteligencia Artificial y Educación de Beijing, las políticas públicas en México han sólo recientemente tomado en cuenta las recomendaciones de la Organización.

De forma complementaria a las regulaciones en materia tecnológica, podemos resaltar el énfasis que las agendas digitales en el país han hecho en torno al despliegue de tecnologías informáticas. En el año 2013 entró en vigor la reforma a la Constitución Política de los Estados Unidos Mexicanos y, entre muchas otras, a la Ley General de Educación en el país. En esta última, la Secretaría de Educación indicaba la necesidad por desarrollar el potencial humano de los mexicanos con educación de calidad que incorporase y aprovecharse las TIC como medio para el proceso de enseñanza-aprendizaje y como objeto mismo de estudio. Sin embargo, el país no contaba en ese momento con modelos educativos desarrollados con ese fin.

En general, las políticas públicas impulsadas desde este año y sus correspondientes agendas digitales se han comprometido con la educación técnica como un medio para empoderar a los usuarios de las

tecnologías computacionales, promoviendo la alfabetización digital y fomentando la transparencia desde un punto de vista técnico. Pero estos esfuerzos han fallado en incluir a la IA/AP dentro de las iniciativas estatales. Este aspecto ha llamado la atención de investigadores y agencias (Artículo 19, ILDA, Centro Latam Digital, entre otros) debido a la falta de interés activo por parte de las estrategias nacionales en materia digital para incluir a las tecnologías de IA/AP dentro de los principales nodos de discusión en el futuro de la industria, la economía y el bienestar social.

En realidad, los retos y problemas que conlleva el uso de sistemas de Inteligencia Artificial no han figurado de manera clara en las políticas mexicanas. Esto se hace patente en que hasta ahora no ha habido un ejercicio consistente por parte de las dependencias públicas en la creación de una Estrategia Nacional para la implementación de estas tecnologías, como sí han hecho algunos otros países de la región. Por el contrario, al día de hoy los más importantes esfuerzos realizados para sondear el estado de la industria tecnológica y digital en materia de IA/AP ha corrido a cuenta de particulares (como Artículo 19) y externos (como la Embajada Británica en México). Sin embargo, ninguno de los esfuerzos ha tenido incidencia en las acciones tomadas por el Gobierno Federal.

En el 2018 la Embajada Británica en México, junto con la organización C Minds y Oxford Insights, en colaboración con la Coordinación de Estrategia Digital Nacional, publicó un reporte que pretendía proponer recomendaciones al gobierno federal para el impulso e implementación de una efectiva estrategia de IA. Destaca que este informe fue creado bajo la autoría de Emma Martinho-Truswell, directora de Oxford Insights, Hannah Miller, consultora tecnológica y experta en gestión de riesgos, Isak Nti Asare, especialista en ciberseguridad, André Petheram, líder de investigación de Oxford Insights en materia de gobernanza digital, Richard Stirling, CEO de la organización, y las mexicanas Constanza Gómez Mont, fundadora de C Minds, y Cristina Martínez, líder del PIT Policy Lab. El perfil del grupo de investigación sugiere una distancia respecto a las

condiciones económicas y sociales del país, pero también provee una mirada crítica sobre la actualidad de las políticas públicas desde la investigación académica, incluso cuando no cuenta con la colaboración de legisladores o tomadores de decisión normativa en el territorio.

Con especial interés en cinco áreas específicas, el reporte buscaba contribuir en la problematización de los espectros de [1] gobierno y servicios públicos, [2] investigación y desarrollo, [3] datos e infraestructura digital, [4] capacidades, habilidades y educación y [5] ética. La transparencia de las tecnologías es apenas mencionada en el texto, y, por el contrario, se destaca el papel de la adquisición de capacidades y habilidades técnicas por medio de la educación. La recomendación sugiere:

La educación en todos los niveles, en las escuelas, en la universidad y a lo largo de la vida de los ciudadanos, es importante tanto para aumentar las habilidades de la IA en México como para garantizar que los beneficios de la IA se compartan. Junto con la recomendación de aumentar el número de estudiantes en los cursos de inteligencia artificial y ciencia de datos, sugerimos enseñar pensamiento computacional en las escuelas mexicanas y los conceptos de la IA como parte de las ciencias sociales, así como las matemáticas y ciencias. (C Minds, 2018, 8)

De acuerdo con lo anterior, se esperaría que estas estrategias faciliten “el aprendizaje permanente” de los ciudadanos, a través de:

1. Desarrollo de herramientas para la educación continua en IA.
2. Ampliación del aprendizaje de la IA más allá de los estudiantes de informática y de matemáticas en universidades públicas y privadas.
3. Enseñanza de pensamiento computacional en escuelas.
4. Aumento en el número de estudiantes de maestría y doctorado en IA y en ciencias de datos.

A pesar de hacer un análisis comprehensivo de los aparatos legales mexicanos en relación con las estrategias implementadas a nivel mundial para la creación de competencias internacionales, el reporte no discute la falta de infraestructuras tecnológicas ni ahonda en el problema de las brechas digitales y el acceso tecnológico. Al ignorar la importancia de la inclusión digital y de los panoramas económicos en México, el documento plantea la necesidad por adoptar enfoques para los que las instituciones educativas mexicanas (principalmente públicas) no están acondicionadas.¹

Estos elementos son además limitantes contextuales que sólo adquieren sentido al atender la naturaleza de las infraestructuras tecnológicas del país. Por ello, Gutiérrez y Limón (2019) afirmarían necesario considerar, en primer lugar, el mejoramiento de las instalaciones educativas y las estructuras económicas en el país para dotar de las necesidades de infraestructuras y servicios básicos a las escuelas.

En esta línea, y en respuesta a los infructuosos intentos de inclusión digital previos, el Congreso General de México publicaría una nueva Ley General de Educación que abrogó en el 2019 a la Ley General de la Infraestructura Física Educativa. Por primera vez la legislación en materia de educación comprendería, en el artículo 18 de su Capítulo IV, no sólo el aprovechamiento de las plataformas digitales sino la promoción en “el conocimiento tecnológico, con el empleo de tecnologías de la información, comunicación, conocimiento y aprendizaje digital, manejo de diferentes lenguajes y herramientas de sistemas informáticos, y de comunicación”. La misma ley plantearía en su capítulo XI la creación de una Agenda Digital Educativa que, de manera progresiva, permitiría:

- i. El aprendizaje y el conocimiento que impulsen las competencias formativas y habilidades digitales de los educandos y docentes;

¹ En el reporte, por ejemplo, se plantea la necesidad por incorporar módulos de tecnología en cursos no tecnológicos al estilo del Tec de Monterrey (escuela privada en México de difícil acceso para las comunidades económicamente vulneradas del país) para preparar a las personas para “el mercado laboral del futuro, no del pasado”.

- ii. El uso responsable, la promoción del acceso y la utilización de las tecnologías de la información, comunicación, conocimiento y aprendizaje digital en los procesos de la vida cotidiana;
- iii. La adaptación a los cambios tecnológicos;
- iv. El trabajo remoto y en entornos digitales;
- v. Creatividad e innovación práctica para la resolución de problemas, y
- vi. Diseño y creación de contenidos.

Contraria a las iniciativas públicas que le antecedieron para promover la alfabetización digital, la Agenda Digital Educativa (ADE) promulgada por la Secretaría de Educación (2020) no tuvo como objetivo la distribución de dispositivos digitales, sino la inserción de las tecnologías en los procesos instruccionales para el aprendizaje técnico.

Asimismo, la ADE presenta por primera vez de manera clara en las políticas de estado un interés por incluir a las tecnologías de IA/AP tanto como medios de aprendizaje/enseñanza como objetos de estudio para los educandos del país, como ha sugerido la UNESCO en el Consenso de Beijing. Así, a partir de la incorporación y acceso a las TICCAD inmersivas (tecnologías de la información, comunicación, conocimiento y aprendizaje digitales) se pretende la utilización de recursos de realidad virtual y aumentada para el aprendizaje. Por otro lado, se apuesta por la enseñanza en robótica educativa que contempla su vinculación con la programación de computadoras, la inteligencia artificial y el diseño de ingeniería (ADE, 2020, 87). De este modo, la vigente Agenda busca contrarrestar los efectos de la brecha digital al apostar por una educación cada vez más involucrada con las tecnologías que reduzcan los índices de analfabetismo técnico, así como de escritura y lectura.

Para ello, la ADE tiene dentro de sus objetivos fomentar en las comunidades educativas la alfabetización digital mediante [1] la alfabetización técnica-tecnológica, basada en el conocimiento básico para el uso técnico y operativo de los equipos tecnológicos digitales, con sus

instrucciones y herramientas de manejo, y [2] la alfabetización informacional, que consiste en la educación en lenguajes y aplicaciones para conocer, acceder y utilizar los distintos servicios, recursos y contenidos digitales que ofrecen con miras en la protección de la seguridad de los usuarios (p. 78).

Complementaria a la ADE, la Coordinación de Estrategia Digital Nacional de la Presidencia de la República expediría en el segundo semestre de 2021 la Estrategia Digital Nacional 2021-2024, que tiene como fin “orientar los esfuerzos e iniciativas tecnológicos y de seguridad de la información en una misma dirección y sentido tecnológico” (DOF, 2021). El acuerdo refrendaría la importancia del cierre de las brechas de acceso a Internet para combatir la marginalización y la pobreza, así como para asegurar el uso democrático y universal de los recursos digitales.

El acuerdo llama la atención, además, en la necesidad por crear derivas educativas basadas en la igualdad que no privilegien a colectivos sociales específicos, pues apunta:

A su vez, los costos y la accesibilidad desigual de las TIC han devenido en una división más marcada de la sociedad y se habla de los que tienen acceso y de los que no tienen acceso a las tecnologías y al Internet. Asimismo, se mencionan las habilidades digitales para los trabajos especializados, lo que deriva en una jerarquía basada en cierto tipo de aprendizaje y de productividad; es decir, es posible que tenga mayor poder adquisitivo y oportunidades laborales quien cuenta con herramientas digitales que le permitan desempeñar un trabajo, a diferencia de quien no las tiene. Esto genera una exclusión con la que se debe acabar.

Ante este panorama se puede afirmar que los marcos normativos mexicanos en materia de educación han mostrado un interés activo por reducir los índices de analfabetismo técnico mediante el cierre de las brechas digitales en el país. Si bien algunas estrategias han favorecido el repartimiento de insumos y dispositivos computacionales, mientras que otras han apostado por el mejoramiento de las infraestructuras nacionales,

podemos constatar un ejercicio que busca solventar el problema de la falta de transparencia y de acceso a las tecnologías inteligentes. Afinar estas estrategias resultará necesario durante los siguientes años para evitar el fracaso de las iniciativas, así como para ponderar los alcances y límites reales en la búsqueda de la inclusión digital.

Si bien hace falta generar nuevas estrategias que respondan específicamente al problema de la transparencia de los sistemas de IA/AP, es importante resaltar que una política pública que parta de la transparencia como un derecho humano a la información y al conocimiento (como la que ha exigido Artículo 19, C Minds y la Embajada Británica) no puede ignorar la necesidad por masificar en primer lugar el uso y acceso a las tecnologías computacionales no inteligentes y plataformas como Internet.

De este acercamiento podemos afirmar que, contrario a los intereses de la industria para el desarrollo tecnológico, sólo en la medida en que haya una progresiva alfabetización de la población mexicana en su conjunto es que se puede esperar una efectiva agenda de educación en materia de IA/AP y, consecuentemente, una economía sujeta al desarrollo de las comunidades de la información.

1.3.2 Proyecto de decreto para expedir la Ley para la Regulación Ética de la Inteligencia Artificial para los Estados Unidos Mexicanos

En épocas más recientes, destaca de entre estos intentos el *Proyecto de decreto para expedir la Ley para la Regulación Ética de la Inteligencia Artificial para los Estados Unidos Mexicanos*, publicado en la Gaceta Parlamentaria el 30 de marzo de 2023. En él, el Grupo Parlamentario del Partido Acción Nacional (PAN), liderado por Ignacio Loyola Vera, ha buscado integrar a la Inteligencia Artificial como un elemento de relevancia en el Artículo 73 constitucional, que faculta al Congreso para:

...dictar leyes sobre vías generales de comunicación, tecnologías de la información y la comunicación, radiodifusión, telecomunicaciones, incluida

la banda ancha e Internet, postas y correos, y sobre el uso y aprovechamiento de las aguas de jurisdicción federal.

Este expediente deja en evidencia las capacidades del estado mexicano para crear marcos regulatorios más allá de Normas y que tengan el carácter de Ley.

En esta propuesta, llama la atención que se hable de la “opacidad en las “cajas negras”” para referirse únicamente a los sesgos de representación demográfica o ideológicos (Cámara de Diputados del H. Congreso de la Unión, 2023). Sin embargo, la iniciativa no contempla el problema de transparencia computacional, el analfabetismo técnico o tecnológico, las brechas digitales o la falta de infraestructura digital en el país (salvo en su mención al Plan de Desarrollo Nacional).

Por otro lado, destaca que el decreto establezca como uno de sus objetivos en el Artículo 1 de sus Disposiciones Generales:

...regular y normar el uso de la inteligencia artificial (IA) y la robótica en su uso con fines gubernamentales, económicos, comerciales, administrativos, comunicacionales y financieros para que su uso sea siempre basado en apego a la ética y en apego al derecho.

La falta de regularización del espectro médico y de atención a la salud entre los objetivos de la ley deja en evidencia el carácter preliminar de uno de los pioneros intentos de regulación de la IA en México. La especificidad de la salud pública, del desarrollo de tecnologías informáticas como dispositivos médicos y de la industria médica y farmacéutica en general exige la creación de lineamientos claros sobre el uso responsable de la IA. Por ello, una posible ley para la regulación ética de estas tecnologías no podrá distanciarse de las normas y legislaciones vigentes y debe sintetizarse en un arreglo de la Ley General de Salud, de la Ley General de Educación y de la Ley Federal de Transparencia y Acceso a la Información Pública.

En este contexto, la presente investigación busca posicionarse en los albores del interés legislativo por crear marcos normativos que incidan en el manejo de la salud pública en el país.

1.3.3 NOM-241

Por otro lado, en el escenario nacional es notable que, a pesar de ser el primer exportador de dispositivos médicos en Latinoamérica y uno de los más importantes a nivel mundial, México no cuenta con regulaciones claras para el desarrollo, venta y uso de sistemas de Inteligencia Artificial en el área médica, y las leyes existentes no son adecuadas ni lo suficientemente específicas para este propósito.

Al día de hoy, la única mención que existe en las normativas mexicanas respecto al uso de tecnologías computacionales en el área de la salud está contenida en la Norma Oficial Mexicana NOM-241-SSA1-2021 “Buenas prácticas de fabricación de dispositivos médicos”. Esta iniciativa sería propuesta por la Secretaría de Salud en colaboración con diversas cámaras, asociaciones y universidades² para modificar la Norma 241 “Buenas prácticas de fabricación para establecimientos dedicados a la fabricación de dispositivos médicos”. En el proyecto de modificación de norma, que al 2023 entró en vigor tras cinco años de su evaluación, solamente sería considerado el software como dispositivo médico en su inciso 3.102, definido como:

...al utilizado con uno o más propósitos médicos, que tiene como característica principal que no requiere de un hardware para cumplir con el propósito médico previsto; es capaz de funcionar en plataformas computacionales generales y puede utilizarse solo y/o en combinación con otros productos (ej., como módulo, otros Dispositivos Médicos, etcétera). Las

² Destacan el Consejo de Salubridad General, IMSS, ISSSTE, UNAM, IPN, CANACINTRA, CANIFARMA, CANAJAD, CANIPEC, la Academia Nacional de Ciencias Farmacéuticas, la Asociación Farmacéutica Mexicana, el Colegio Nacional de Químicos Farmacéuticos Biólogos México, Producción Químico Farmacéutica, AMELAF y AMID, así como sus respectivas dependencias y departamentos facultados para la evaluación de dispositivos médicos.

aplicaciones móviles que cumplen con esta definición son consideradas Software como Dispositivo Médico. El software que hace funcionar al Dispositivo Médico físico queda excluido de esta definición.

No obstante esta categorización, la NOM 241 no contempla ninguna especificación del debido tratamiento o especificidad de los sistemas de Inteligencia Artificial y solamente recurre a la noción de “dispositivo médico” de la regulación de Dispositivos Médicos promovida por la FDA. Esto significa que la transparencia no figura explícitamente en la normativa y no forma parte de las exigencias regulatorias en el contexto mexicano.

Inspirada en la regulación de la FDA, la normativa 241 especifica en su inciso 3.109 que la validación del software como Dispositivo Médico radica exclusivamente en “la evidencia documental generada a través de la recopilación y evaluación de la capacidad de un Software para generar con exactitud, integridad y precisión las funciones previstas a partir de los datos de entrada”. Asimismo, es importante resaltar que en el marco normativo mexicano a los sistemas computacionales se les define como “cualquier equipo, proceso u operación que tenga acoplada una o más computadoras y un software asociado o un grupo de componentes de hardware diseñado y ensamblado para realizar un grupo específico de funciones” (NOM 059, NOM 241). Sin embargo, sólo el software es considerado en la modificación de norma como dispositivo médico, mientras que el hardware y otros agentes involucrados en su desarrollo no.

Por otro lado, la actual normativa en materia de desarrollo de software como dispositivo médico en México no contempla ningún método de validación como un parámetro de evaluación del funcionamiento de los sistemas. La falta de especificidad en la normativa solamente contempla “evaluaciones científicas” (NOM 241: 3.107) que, en el caso de los sistemas computacionales, debe ser llevada a cabo por los establecimientos que pueden o no tomar en consideración los estudios de validación de los proveedores. Además, la NOM 241, en su inciso 3.109, sólo contempla que la evidencia documental del cumplimiento en la exactitud, integridad y

precisión de las funciones previstas a partir de los datos de entrada puede validar el uso de las tecnologías de software.

Adicionalmente, la NOM 241 contempla que la Gestión de Calidad tiene que considerar un análisis de riesgos que en el inciso 3.8 es designado como el “método para evaluar con anticipación los factores que pueden afectar la funcionalidad de: sistemas, equipos, procesos o Calidad de Insumos y producto”. Esta designación deja de lado al software como dispositivo médico tal como es definido por la misma normativa y no contempla el problema de la opacidad de los sistemas computacionales. Solamente en su inciso 3.55 se aborda a la gestión de riesgos de calidad como el “proceso sistemático para la valoración, control, comunicación y revisión de los riesgos a la Calidad de los Dispositivos Médicos a través de su Ciclo de Vida”. Sin embargo, una normativa no solamente debe considerar una validación técnica de los sistemas tras su utilización efectiva, sino una política de evaluación de riesgos previa a su aprobación y no en retrospectiva (Marda, 2017). Para ello, el apartado 11, correspondiente a la “Clasificación y validación” afirma en su inciso 11.11.2 que la Validación del proceso “debe contemplarse antes de la distribución y venta del producto”.

Un problema fundamental es que esto sólo contempla su realización con un enfoque de gestión de riesgos de calidad y, en última instancia, con el funcionamiento técnico de los sistemas y la eficiencia de sus resultados que no es justificable debido a la misma naturaleza de los sistemas. Por otro lado, la validación de sistemas computacionales (11.15) no contempla especificidades del software, mientras que la validación de software considerado como Dispositivo Médico no tiene parámetros de evaluación explícitos, pues relega autoridad al marco de categorización de riesgos formulado por la FDA de Software como Dispositivo Médico (2013). Es importante recordar que este último no establece instrumentos de evaluación definidos ni procesos de monitoreo más allá de la retroalimentación con los usuarios y la adecuación a manuales de buenas prácticas por parte de los desarrolladores.

Por lo anterior, podemos afirmar que, en su estado actual, la NOM 241 no es comprensiva ni aplicable a los sistemas de Inteligencia Artificial por los siguientes puntos:

1. Delega la responsabilidad a los establecimientos que deben diseñar, implementar, documentar y mantener el sistema de gestión de calidad a través de la creación de manuales de calidad propios. Por tanto, no existen instrumentos ni parámetros de evaluación aplicables a las tecnologías computacionales mismas.
2. La validación de los sistemas computacionales se basa en la recopilación de datos que verifican que el sistema se comporta de manera eficiente de acuerdo con los valores de entrada. Esto no contempla la opacidad epistémica que es producto de las facultades cognitivas de los evaluadores, para quienes no son transparentes los procesos internos del sistema. La NOM 241 solamente define la seguridad del equipo en términos inductivos y, por sí misma, la gestión de riesgos no es adecuada para liberarse de la incertidumbre operativa de los sistemas.
3. Para la gestión de calidad del software como dispositivo médico no se hace alusión a terminologías específicas del aprendizaje de máquinas. Por tanto, hace falta crear políticas que permitan la auditoría de algoritmos, códigos, modelos y bases de datos que permiten el funcionamiento operativo de las tecnologías.

Estas condiciones parecerían conducirnos a afirmar la necesidad por contar con mesas evaluadoras de expertos para la validación de las tecnologías y para su monitoreo que, además, estimen los riesgos reales en el uso de cada tecnología, diseñen parámetros de evaluación homogéneos y establezcan códigos de buenas prácticas replicables a gran escala. Pero, además, se necesitaría contar con el cumplimiento de estas exigencias y estándares por parte de productores y usuarios, así como de los gestores de salud miembros de las instituciones médicas.

En el siguiente capítulo cuestionaré si apelar a el conocimiento experto es una vía adecuada para promover la transparencia. Esta pregunta perseguirá como fin dejar en evidencia que el dominio técnico no nos compromete necesariamente con un aumento en la transparencia en el uso de las tecnologías.

CAPÍTULO 2. Perspectivas frente a la transparencia en el uso de sistemas de Aprendizaje Profundo

En su fundacional texto sobre la cibernética, Norbert Wiener definió a las cajas negras como “piezas del aparato [...] que llevan a cabo operaciones definidas en el presente y el pasado del potencial de un input, pero de las que no necesariamente tenemos la información de su estructura mediante la cual la operación es llevada a cabo” (2019, p. 285). Wiener tenía en mente el problema de los sistemas no lineales que durante la Segunda Guerra Mundial habían promovido la creación de complejos computadores digitales. Durante los años treinta se había asumido que los circuitos electrónicos podían ser descritos adecuadamente mediante un análisis de los mensajes enviados por un sistema en relación con las impedancias, entradas y voltajes del circuito. Pero con la emergencia de sistemas computacionales y digitales parecía ya no ser suficiente este marco de análisis. Por su complejidad no lineal, las nuevas tecnologías se presentaban como cajas negras cuya estructura y funcionamiento interno no era transparente a los usuarios.

Aproximadamente en la misma época, Frank Rosenblatt propondría el algoritmo que pretendía explicar matemáticamente el funcionamiento de las operaciones básicas del cerebro. En el año 1958, el equipo del Laboratorio Aeronáutico de Cornell presentó el perceptrón como un algoritmo inspirado en principios biológicos del funcionamiento de las células neuronales que permitía llevar a cabo operaciones lineales para la clasificación binaria. La propuesta del perceptrón serviría para desarrollar sistemas computacionales mucho más complejos que, de cierto modo, terminarían imitando la opacidad de los sistemas neuronales. A partir de la creación de sistemas más complejos para el procesamiento de la información inspirados en teorías neuronales, las máquinas computadoras adquirieron también este carácter de indescifrabilidad (Oliver, 2020, 43).

Al contar con tantas variables que vuelven opacos sus funcionamientos y operaciones, podríamos vernos conducidos a afirmar que los sistemas de Inteligencia Artificial se comportan, al igual que sus contrapartes biológicas, como cajas negras.

Al estar basados en lógica formal del estilo “si... entonces...”, los sistemas expertos clásicos tienen una estructura lógica completa que puede ser deducida a partir del conocimiento del algoritmo para la toma de decisiones. Esto naturalmente guarda una relación directa con la cantidad de información a procesar y con el relativamente bajo número de variables en él. La complejidad de estos sistemas de deducción es notablemente menor que el de arquitecturas que integran cálculos probabilísticos con operaciones recursivas para el aprendizaje automatizado y que consumen información no vectorial, tales como las Redes Neuronales Convolucionales, Recurrentes o Adversariales.

Si bien podemos afirmar que algunas tecnologías de Inteligencia Artificial como los sistemas expertos son sistemas cuyas partes podemos conocer en su totalidad para predecir cabalmente sus posibles estados finales, sería un error asumir que este carácter es propio de todas las arquitecturas de IA. Una de las problemáticas más ampliamente criticadas en torno a la implementación de sistemas de Aprendizaje Profundo es que al procesar cantidades de información tan amplias y al generar sus propias variables y pesos a través de complejas operaciones, mucha información, así como funciones operativas, resultan inaccesibles a sus operadores y usuarios.

Una de las críticas más ampliamente popularizadas respecto a este problema gira en torno a que los expertos en materia computacional podrían franquear estas limitantes epistémicas. Si suponemos que el nivel de opacidad depende del agente que interactúa con el sistema, podríamos afirmar que un desarrollador computacional o un ingeniero en sistemas podría conocer estos opacos y crípticos principios y mecanismos de la red. Pero una suposición de esta naturaleza sólo sería en teoría verdadera.

Por ejemplo, Julian Newman (2016) ha afirmado que la opacidad de estos sistemas debería ser un problema evitable para las ciencias de la computación en tanto se implementen prácticas ingenieriles y sociales correctas (p. 258). Para Newman las arquitecturas de software “bien construidas” no son opacas y, en realidad, esta atribución dependería solamente de la falta de adopción de los saberes básicos de las ciencias computacionales por parte de los agentes que interactúan con ellas (Duran, 2018, p. 646). En cierto modo, Newman supondría que el conocimiento sobre cómo funciona un sistema es suficiente para comprender la naturaleza de sus valores de salida; por tanto, el nivel de dominio del conocimiento de los agentes que interactúan con estos sistemas debería permitirnos reconocer que las tecnologías pueden ser epistémicamente transparentes.³

Esta es la misma directriz que han tomado la mayor parte de las normativas que exploré en el primer capítulo. Estas presuponen que el conocimiento técnico es suficiente para que no exista opacidad en el uso de las tecnologías. A su vez, apuestan por la validación técnica llevada a cabo por medio de los expertos en materia computacional para afianzar la seguridad en el uso de los sistemas.

En este segundo capítulo cuestionaré la propuesta de Newman, así como los acercamientos técnicos llevados a cabo por la industria y la academia para evaluar si las actuales estrategias para la transparencia técnica son suficientes para asegurar un uso transparente de las tecnologías. Asimismo, propondré que esta perspectiva parte de un privilegio basado en la autoridad epistémica que vulnera el derecho a la toma de decisión de las personas y que no resuelve el problema que surge de la intersección entre usos de la IA en el área de la salud y las prácticas médicas centradas en el paciente.

³ Llama la atención en este punto que hoy en día es relativamente sencillo construir modelos con base en otros pre-entrenados. Esto deja en evidencia que en algunos casos el conocimiento técnico no es un requisito para la construcción de tecnologías de IA complejas. Además, pone en crisis a la figura del desarrollador como autor.

2.1 Transparencia técnica y estrategias para la promoción de la explicabilidad

Contra el argumento de Newman, Juan Durán y Nico Formanek (2018) han afirmado que no es suficiente conocer el método a través del cual un sistema trabaja para afirmar que se entiende cómo llegó a su resultado y porqué. Por ejemplo, si hablamos de un sistema inteligente para la detección de anomalías cardíacas, no será suficiente entender los algoritmos que están detrás de la tecnología para saber por qué el sistema llegó a un diagnóstico concreto. Por el contrario, se requiere de más información sobre la industria médica, el perfil del paciente y la decisión médica para poder crear un panorama completo sobre el funcionamiento de un sistema. Para argumentar en favor de esta suposición, Durán y Formanek partirían de la definición de los procesos epistémicamente opacos de Paul Humphreys.

Podemos afirmar que un proceso es epistémicamente opaco en relación con un agente X en un momento t en caso de que X no conozca todos los elementos epistemológicamente relevantes del proceso (Humphreys, 2009, p. 618). Esta definición de opacidad epistémica sería planteada por Humphreys en su estudio sobre la necesidad por pensar filosóficamente los problemas que conlleva la relación entre humanos y máquinas de cómputo en la producción de conocimiento. Su conceptualización del fenómeno funcionará como el punto de partida que tomaremos para entender esta relación.

De acuerdo con Humphreys, la opacidad epistémica no es un problema exclusivo de las ciencias computacionales, sino que es una cuestión que compete a la filosofía de la ciencia en general, que se ha preguntado por cómo conocemos a través de instrumentos científicos. No obstante esta acotación, la opacidad sí es un problema particularmente comprometedor para la informática computacional ya que, contrarios a los instrumentos de medición y representación analógicos, “ningún humano puede examinar y justificar cada elemento de los procesos computacionales que producen

un valor de salida [...] o de otros artefactos de las ciencias de la computación” (p. 618). En realidad, no sería sino con el surgimiento de los sistemas computacionales complejos posteriores a los años cuarenta del siglo pasado que los detalles de las partes de un sistema se volverían progresivamente inaccesibles a las capacidades humanas. Así, el filósofo afirmaría que existe una infranqueable opacidad frente a las rápidas y complejas computaciones de los sistemas ya que “ningún humano ni grupo de humanos puede, en la práctica, reproducir o entender [sus] procesos” (p. 619).

De acuerdo con Duran y Formanek, asumir que un agente X no conoce todos los elementos epistémicamente relevantes de un proceso implica tomar en consideración, al menos, dos condiciones del proceso computacional: la velocidad y el volumen. El agente X, por sus capacidades cognitivas, puede estar incapacitado para seguir a la misma velocidad cada uno de los pasos en el procesamiento de la información por parte del sistema. Asimismo, estos agentes, por sus capacidades de memoria y percepción, pueden estar incapacitados para captar la totalidad de las operaciones y el flujo de la información dentro del mismo. Durán y Formanek afirmarían, así, que “un proceso es epistémicamente opaco en relación con un agente cognitivo X en el tiempo t sólo si X en t no tiene acceso a y no puede monitorear todos los pasos de la justificación [del sistema para arrojar un resultado]” (p. 650). Así, podemos asumir que la opacidad epistémica es un problema que se expresa en función de las capacidades cognitivas de un agente y de los conocimientos del mismo (encender una PC puede ser un proceso epistémicamente transparente para un ingeniero en sistemas, pero un suceso mágico para alguien sin conocimiento de las tecnologías digitales).

El argumento central del problema de la opacidad epistémica radica en que los algoritmos computacionales tienen tantos pasos y los sistemas procesan tan amplias cantidades de información que resultan inaccesibles e impenetrables para los agentes cognitivos humanos. De esto se sigue, además, que la creencia en sus resultados termina siendo imposible de

justificar. En contraposición a esta idea, Durán y Formanek han afirmado que, en realidad, sí contamos con diferentes recursos para generar confianza en los sistemas computacionales y que, en consecuencia, podemos justificar nuestras creencias en sus resultados. Este mismo principio ha sido rector en la búsqueda de transparencia en el uso de la IA.

De acuerdo con los autores podemos identificar al menos cuatro fuentes que nos permiten atribuir fiabilidad a los sistemas computacionales:

- [1] los métodos de verificación y validación,
- [2] el análisis en la robustez,
- [3] la historia de sus implementaciones exitosas/no exitosas y
- [4] el conocimiento experto.

En lo consecutivo, exploraré cómo estas cuatro fuentes pueden y han sido abordadas en relación con los sistemas de Inteligencia Artificial; esto tendrá como objetivo indicar la insuficiencia que tienen para, de hecho, promover la transparencia de las tecnologías.

Afirmaré que estas fuentes no satisfacen la creación de transparencia ni necesariamente de fiabilidad en los resultados de los sistemas cuando son pensados para asegurar la transparencia en el uso de tecnologías de IA.

2.1.1 Métodos de verificación y validación

Como he afirmado, la opacidad epistémica en sistemas complejos como los de Aprendizaje Profundo tiene un carácter necesario. Esto es, que no podemos asegurar la transparencia absoluta de los sistemas y que no importa el nivel de pericia de un agente cognitivo humano X, siempre habrá un grado de opacidad mínimo respecto a sus principios y mecanismos. Los agentes humanos están, por tanto, sometidos necesariamente a la indescifrabilidad de estos sistemas.

Como respuesta a este problema, en los últimos años los desarrolladores de sistemas de Aprendizaje Profundo hicieron notar la necesidad por crear vías para hacer descriptibles los procesos internos de estas tecnologías. Ante el problema de la incapacidad por monitorear en su totalidad a los sistemas, la industria e investigación en materia computacional recurrieron a la creación de una rama de la Inteligencia Artificial específicamente destinada a hacer entendibles a los sistemas: la Inteligencia Artificial Explicable (o XAI). De acuerdo con la Agencia de Proyectos de Investigación Avanzados de Defensa (DARPA por sus siglas en inglés) la Inteligencia Artificial Explicable, tiene por misión “entender, confiar de manera apropiada y administrar de manera efectiva una emergente generación de máquinas acompañantes artificialmente inteligentes” (2020). El XAI pretende crear un conjunto de técnicas de aprendizaje de máquinas que contribuya a crear modelos más explicables manteniendo un alto nivel de rendimiento. Así, se esperaría que se pudieran desarrollar técnicas que solucionen el conflicto entre explicabilidad-contradesempeño de los sistemas.

En diversas áreas, como el diagnóstico asistido computarizado, existe una importante necesidad de que los sistemas sean transparentes, entendibles y explicables para poder ganar la confianza de los médicos, reguladores e, incluso, de los pacientes. Idealmente, como afirma Singh (2020), un sistema de diagnóstico médico, por ejemplo, debería ser capaz de explicar completamente y a todas las partes involucradas la lógica a través de la cual toma una decisión. Para cumplir con esta misión, diversas estrategias para la validación de los resultados de un sistema han sido propuestas. Estas validaciones requieren, por un lado, la verificación en la correspondencia entre el resultado del sistema y el estado de cosas del mundo y, por el otro, la verificación del correcto funcionamiento operativo del mismo de acuerdo con los principios teóricos que lo determinan.

En el área médica se han implementado diversas estrategias para reducir el nivel de incertidumbre que provoca la operatividad de estas tecnologías. De acuerdo con Erico Tjoa y Cuntai Guan (2020) podemos

distinguir entre dos grandes categorías de estrategias para hacerlas explicables: por medio de una interpretabilidad perceptiva o por medio de una interpretabilidad matemática.

Los métodos que pretenden hacer interpretables a los sistemas son múltiples; por su naturaleza sensible, los perceptivos son principalmente aplicados para explicar el funcionamiento de sistemas de visión artificial o de procesamiento de sonidos. Por ejemplo, algunos métodos de XAI buscan visibilizar qué información visual está buscando el sistema en un momento determinado. Así, una tecnología construida para la detección de signos visuales de tuberculosis en radiografías de pulmones puede incluir un método de segmentación que indique al usuario qué partes de la imagen determinaron el índice de probabilidad arrojado por el sistema; estos métodos (también conocidos como saliency methods en inglés) hacen perceptibles por medio de imágenes los rasgos sobresalientes del material evaluado [Imagen 2].

Por el contrario, como afirmó David Ritscher en el taller público organizado por la FDA en 2020 en torno al rol de la IA en imagenología médica, la Inteligencia Artificial Explicativa tiene un problema nominal de gran relevancia que consiste en que el XAI en realidad no explica nada. Lo que hacen estos métodos, por el contrario, es simplemente dar indicios acerca de lo que está ocurriendo dentro del sistema; naturalmente, estos métodos pueden generar una mayor transparencia en la operatividad de las tecnologías, probar sus comportamientos o, incluso, encontrar fallas dentro de ellas. Sin embargo, los sistemas de XAI entran en un círculo de opacidad o de falta de transparencia que Ritscher diagnosticaría como “tener una Inteligencia Artificial envuelta en otra Inteligencia Artificial huésped”. El problema con los métodos de XAI es que no validan los modelos de Aprendizaje Profundo, sino que permiten, acaso, visibilizar o interpretar algunas partes del sistema, haciéndolo mayormente entendible a los agentes cognitivos.

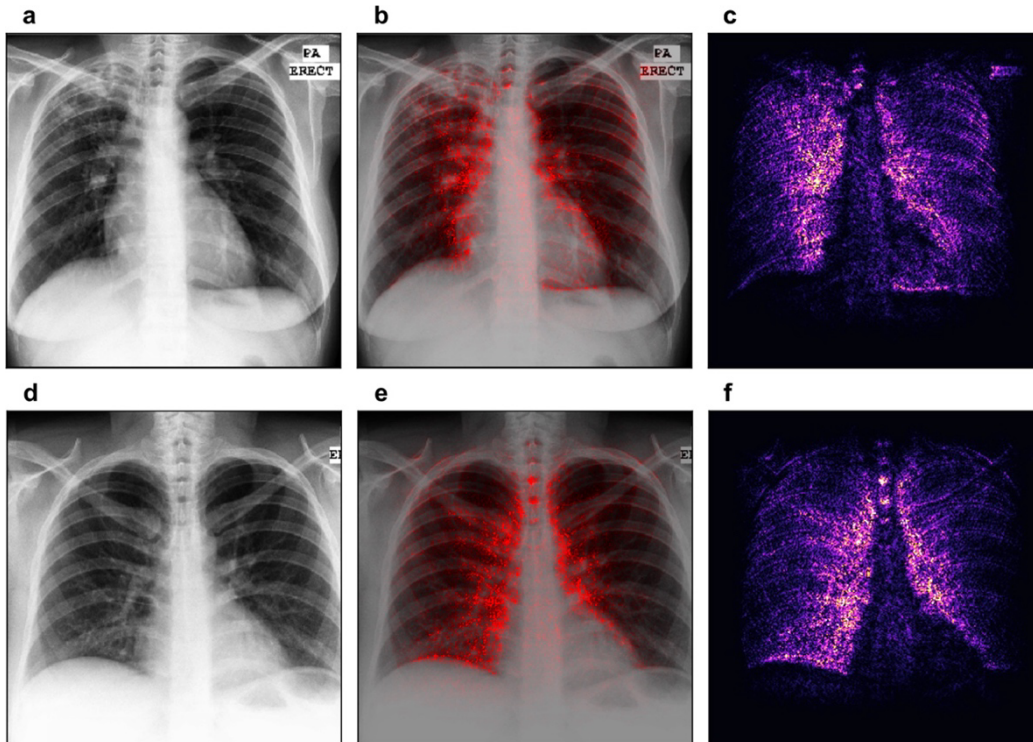


Imagen 2. Mapa de regiones salientes de dos casos correctamente identificados. Los paneles (a) y (d) muestran las imágenes radiológicas de los pacientes; los paneles (c) y (f) muestran los mapas de regiones salientes; los paneles (b) y (e) muestran los mapas de regiones salientes superpuestos a las imágenes radiológicas como medio de comparación. Las imágenes (a), (b) y (c) corresponden a un paciente con tuberculosis. Las imágenes (d), (e) y (f) a un paciente saludable. Extraído de Pasa, F., Golkov, V., et al. (2019) Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization.

Por lo anterior, podemos afirmar que los métodos de Inteligencia Artificial Explicable o bien aumentan la opacidad epistémica del sistema en su conjunto al añadir nuevas capas de algoritmos que no son descriptibles al usuario o bien permiten solamente disminuir el índice de incertidumbre sobre la operatividad de partes del sistema mismo o de sus resultados, pero no de forma integral.

2.1.2 Análisis de robustez

De la mano con los métodos propuestos por el XAI, los análisis de robustez, que incluyen la accesibilidad a las bases de datos que alimentan a los sistemas, han sido asumidos como medios para poder verificar el correcto funcionamiento de las tecnologías de Aprendizaje Profundo. Esta estrategia funge como respuesta al problema de la cantidad de información que un sistema debe procesar en relación con la compleja estructura operativa que le es propia.

Tomado de los germinales estudios en biología de Richard Levins, podemos afirmar que el análisis en la robustez es aquello que permite aprender sobre los resultados de un modelo dado, para saber si sus errores son artefactos (o errores) del sistema o si están relacionados con características centrales del mismo. Esto es, que un análisis de robustez permite distinguir los errores en la construcción de la arquitectura del sistema a partir de una comprobación de que en situaciones imprevistas se comportará de manera correcta.

Respecto a la Inteligencia Artificial, podemos encontrar un proceso de confirmación de la robustez del sistema en el paso de validación y prueba del aprendizaje. Cuando un sistema es construido para el reconocimiento de voz o para la detección de objetos en una imagen, es necesario que pase por un proceso de entrenamiento que contemple el procesamiento de una amplia base de datos en relación con la tarea que debe realizar. Si, por ejemplo, un sistema tiene una arquitectura diseñada para la detección de tumores en la mama, resultará necesario que el sistema sea alimentado con una cantidad amplia de imágenes tomográficas de mamas sanas y de mamas con tumores. Sólo a través de un proceso de entrenamiento el sistema será capaz de extraer las características relevantes de este conjunto de imágenes para producir un mapa de características que sirva como parámetro para la evaluación de los nuevos valores de entrada. Tras ello, y antes de enfrentarse con casos “reales”, el sistema debe pasar por un proceso de validación en el que imágenes no antes procesadas serán utilizadas para medir el nivel de éxito del sistema y para perfeccionar sus

parámetros. Sólo después de este proceso el sistema es puesto a prueba con un pequeño número de imágenes que confirman el nivel de efectividad del mismo para clasificarlas.

En este punto se esperaría que el sistema tuviera la capacidad de asignar en las correctas categorías las tomografías de mamas sanas y de mamas con tumores [Imagen 3] y de, por ejemplo, asignar una nula probabilidad de ser una imagen de una mama con un tumor a una imagen de un gato. Así, la robustez partiría de una generalización inductiva sobre el funcionamiento del sistema.

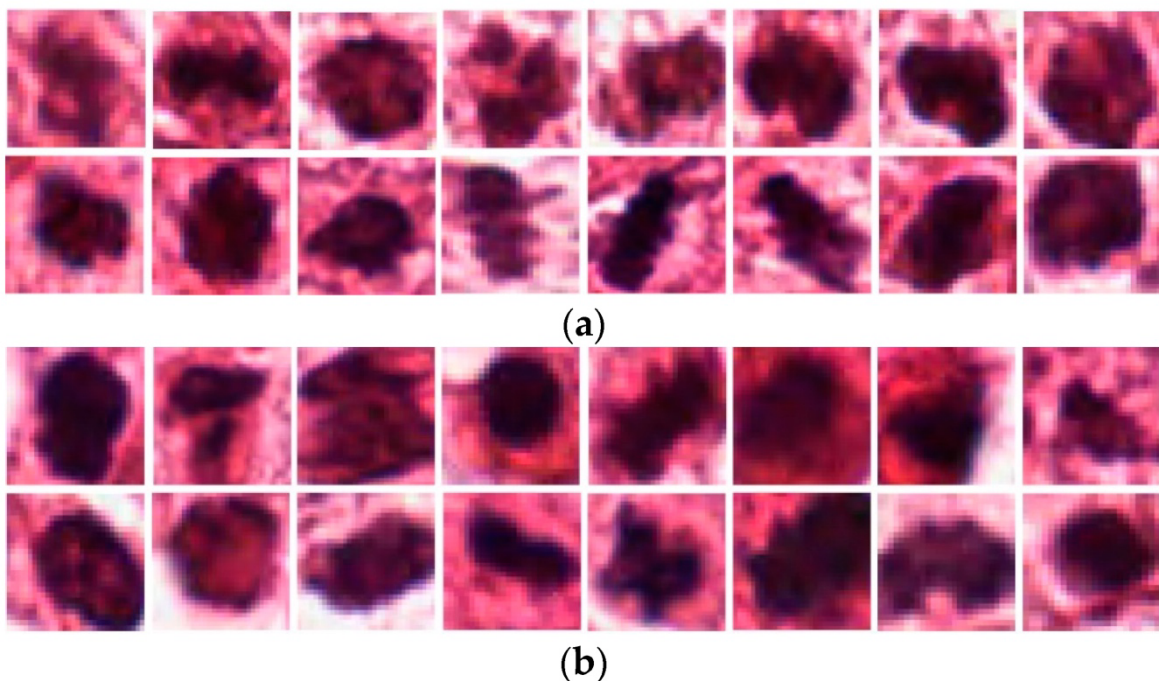


Imagen 3. Ejemplos de imágenes que sirven de alimento a un sistema para la detección de células mitóticas en pacientes con riesgo de cáncer de mama. En (a) células mitóticas y en (b) células no mitóticas. Extraído de Mahmood, T., Arsalan, M., et al. (2020) Artificial Intelligence-Based Mitosis Detection in Breast Cancer Histopathology Images Using Faster R-CNN and Deep CNNs

A mayor cantidad de casos en los que el sistema se comporta de manera correcta en situaciones imprevistas, tendría un carácter más robusto. Este modo de generar confianza en el sistema recuerda al viejo problema de la inducción. Para confirmar la robustez de un sistema tendríamos que poder ponerlo a prueba con todas las situaciones imprevistas posibles,

recurriendo, nuevamente, a un inabarcable conjunto de expedientes a comprobar.

Por otro lado, este método de validación se enfrenta con los ataques adversariales deliberadamente dirigidos por agentes humanos para hacer inoperante a un sistema. Esto significa que incluso después de las evaluaciones de desempeño, los algoritmos pueden ser “engañados” para arrojar resultados erróneos. Los ataques adversariales son además una limitante para el uso extensivo de la IA, ya que las redes neuronales aplicadas en el área médica son especialmente frágiles ante ellos (Kaviani, 2022). Además, en muchos casos el mal funcionamiento de las tecnologías y el entrenamiento sesgado o parcial de las tecnologías suelen escapar a las posibilidades de acción por parte de los usuarios, pues estos no necesariamente cuentan con las herramientas para solucionar los errores o para saber si la tecnología es, en efecto, robusta.

2.1.3 Historia de las implementaciones exitosas/no exitosas

Sería un error suponer, de acuerdo con Duran y Formanek, que un sistema computacional es un sistema acabado. En este sentido, es necesario reconocer que en el estado actual de cualquier sistema computacional hay una historia de sus exitosas y no exitosas ejecuciones, implementaciones y funcionamientos.

En el caso de los sistemas de Aprendizaje Profundo este carácter se vuelve patente con la amplitud del proceso de entrenamiento por el que deben pasar para ser capaces de cumplir con los objetivos para los que fueron diseñados. Un proceso de entrenamiento de una arquitectura promedio puede durar desde unas cuantas horas hasta varios días. Durante este proceso los sistemas hacen correr de manera iterativa el mismo banco de datos varios cientos de veces hasta extraer los pesos indicados para disminuir el error de clasificación. Al partir generalmente de una asignación de pesos iniciales de manera estocástica (o azarosa), el sistema, naturalmente, tendrá durante los primeros ciclos del

entrenamiento (llamados épocas) un índice de error amplio. Con la iteración continua de la misma información, el sistema es capaz de afinar de manera autónoma sus pesos para, al mismo tiempo, aumentar sus índices de éxito.

Paralelamente, podemos comprometernos con que el desarrollo tecnoevolutivo de estos sistemas ha estado determinado, en gran medida, por la aspiración de superación de otras tecnologías en competencia. En el caso de la detección de patologías por medio de sistemas de visión computacional, el perfeccionamiento de los sistemas parte de los resultados de otras arquitecturas y la utilización de estos conocimientos para el afinamiento de sus resultados.

Trasladando el argumento de Durán y Formanek, podemos presumir que este proceso sería una justificación para poder confiar en el sistema. La historia del progreso técnico de las tecnologías serviría, para los autores, de un argumento que deja en evidencia la confiabilidad de los sistemas actuales respecto a sus obsoletos antecesores. El problema fundamental con este recurso es que asume la comprensión de la historia de las tecnologías, información que, generalmente, no es del dominio público ni forma parte del saber colectivo. Esto implica una desventaja epistémica para ciertos actores.

Por todo lo anterior, podemos afirmar que ninguno de estos métodos de validación contrarresta el problema de la opacidad epistémica de las tecnologías de IA/AP. Por el contrario, las características técnicas de los sistemas de Aprendizaje Profundo eliminan la posibilidad de asegurar la transparencia de los sistemas no sólo para los usuarios últimos, sino para todo agente humano que interactúe con las tecnologías. En este sentido, solamente se puede promover la disminución en la opacidad epistémica que es relativa a los agentes que interactúan con ellos.

2.2 Limitantes y problemas en la búsqueda de transparencia técnica

De acuerdo con Durán y Formanek, el conocimiento experto es la cuarta fuente que provee justificaciones suficientes para poder confiar en un sistema computacional. Esto es, que la existencia de expertos en materia computacional, o en los principios que sirven de base a cualquier sistema opaco, es un argumento a favor de la posibilidad de reducir el índice de opacidad del sistema en cuestión y, por tanto, de asegurar una transparencia computacional.

Un problema de esta taxonomía de la fiabilidad computacional es que, en realidad, todas las fuentes propuestas por Durán y Formanek requieren de la existencia de expertos en el tema con un nivel de conocimientos en los principios del funcionamiento de las tecnologías. Así, para poder aplicar métodos de XAI resulta necesario contar con un alto nivel de conocimiento que permitan crear, implementar e interpretar los resultados del análisis. Asimismo, para hacer un estudio de robustez se requiere, al menos, estar familiarizado con la arquitectura en análisis y con los métodos de codificación y procesamiento que utiliza el sistema. Por otro lado, para comprender la tecnoevolución de un sistema y la historia de su afinamiento se requiere, si bien no ser partícipe directo en la construcción del mismo, al menos sí tener un amplio conocimiento del desarrollo de los sistemas y de la historia de la tecnología.

La problemática fundamental de la lectura de Durán y Formanek es que o bien asumen que la fiabilidad es una propiedad transitiva que va de los expertos en la materia hasta los usuarios últimos de los sistemas, o bien que sólo a través del conocimiento experto podemos generar transparencia en el uso de los sistemas mismos. Si bien en el caso de estudio de estos filósofos la simulación computacional, en general, sólo es utilizada con fines científico-industriales, en otros usos de los sistemas computacionales como la Inteligencia Artificial apelar a la confianza depositada por los expertos no es suficiente para, de hecho, asegurar la fiabilidad de los sistemas en su uso cotidiano.

Más aún, al asumir que solo los expertos pueden validar el uso de un sistema con fines médicos, se corre el riesgo de socavar la capacidad de toma de decisión del paciente, quien puede incluso desconocer si se ha utilizado un sistema inteligente para su diagnóstico, tratamiento o seguimiento clínico. En el siguiente apartado exploraré con mayor detalle porqué esto es una limitante para las prácticas médicas contemporáneas.

Preliminarmente, podríamos poner en duda que el usuario final de un sistema inteligente de producción de fármacos o el paciente cuyo diagnóstico es llevado a cabo por una máquina de aprendizaje profundo confíe en los resultados del sistema porque necesariamente tenga un nivel de pericia y conocimientos sobre la tecnología. En general, los usuarios finales (tanto médicos como pacientes) no tienen este conjunto de saberes. Asimismo, al ser la confianza un elemento no normativo de la vida humana sería dudoso afirmar que el usuario confíe en el sistema porque necesariamente confía en los expertos que la han validado.

El uso de estas tecnologías con fines médicos implica que, potencialmente, un amplio número de usuarios (personal radiológico, médico y de enfermería, así como pacientes e, incluso, personas sin preparación) se vean involucrados en el uso directo con estos sistemas. Podríamos esperar que, en el futuro, estas tecnologías se implementen de manera masiva y que, por tanto, nos veamos en la necesidad de confiar en estos sistemas para la procuración de nuestra salud. Si siguiéramos la propuesta de Durán y Formanek tendríamos que apelar a que todos deberían tener los conocimientos suficientes sobre los opacos y crípticos principios y mecanismos de los sistemas para poder generar confianza en ellos. Esta suposición, por demás utópica, resultaría, en lo inmediato, imposible de realizar debido a la actualidad en los índices de analfabetismo técnico.

Si bien es aún común hablar de analfabetismo al referirnos a la incapacidad para leer y escribir, es importante notar, como hace Itzel Casillas Avalos, que la alfabetización hoy no implica nada más el dominio de la lengua a partir de la lectura y la escritura, sino de otras formas, como

la digital, debido a las nuevas dinámicas de vida en el mundo que nos es contemporáneo (Dirección General de Comunicación Social, 2016).

En contraposición a una noción común, partiré de que el analfabetismo técnico no es un factor negativo y que debemos resolver como un objetivo universal, sino que, en realidad, resulta necesario para la existencia humana. Como afirma William Ross Ashby (2015), en la vida cotidiana estamos rodeados de objetos, tecnologías y sistemas que, en lo inmediato, resultan indescifrables para sus usuarios. Los picaportes, las bicicletas y las fuerzas interatómicas que mantienen unidas las partículas de los objetos del mundo son, en realidad, sistemas que no se presentan de manera transparente; en cambio, entenderlos requieren de un esfuerzo cognitivo por parte de los agentes que interactúan con ellos para conocer los mecanismos y principios que rigen su comportamiento. Sólo a través del desmontaje del cerrojo, del seguimiento del mecanismo de la bicicleta o de la instrucción en física podemos suprimir (o, al menos, disminuir) el grado de opacidad que implica toda relación entre algo y un agente que interactúa con ello. Así, podemos afirmar que todas las tecnologías conllevan por necesidad un analfabetismo técnico que a través de la interacción y educación es socavado. Es importante comprender, así, que el analfabetismo técnico se puede tipificar para cada tecnología específica. De este modo, puede hablarse de un analfabetismo de las técnicas escultóricas o, para el caso de interés, un analfabetismo de la IA/AP.

Pero como también afirma Ashby, en la vida cotidiana no es necesario preguntarnos por todos los principios y mecanismos que condicionan la existencia o la operatividad de la totalidad de los sistemas con los que interactuamos. A menos que nos encontrásemos en una situación de peligro en la que necesitaríamos forzar la cerradura de una puerta, que debiéramos reparar por nosotros mismos la cadena de una bicicleta o tuviésemos que salvar al mundo de una desintegración de la materia, no parece ser necesario que conozcamos íntegramente el funcionamiento y naturaleza de estos objetos. Podemos confiar en que el picaporte funcionará cuando nos levantemos por la mañana, que las llantas de mi

bicicleta seguirán rodando hacia adelante y que el mundo no se desintegrará pronto.

En realidad, el problema de la opacidad epistémica causado por el analfabetismo digital sólo adquiere sentido cuando tratamos con sistemas cuyas funciones o resultados tienen una notable importancia en la vida de los agentes humanos.

Por otro lado, resulta difícil pensar críticamente el problema del analfabetismo técnico/digital sin tomar en cuenta la naturaleza de las brechas que condicionan el repartimiento tecnológico en las comunidades (Sánchez, 2020). En realidad, la falta de conocimientos requeridos para la efectiva interacción entre humanos y sistemas de IA (o con sistemas computacionales en general) no solamente tiene una estrecha relación con el problema de la educación (específicamente computacional para nuestro caso de estudio), sino con la limitante que representa la inaccesibilidad a las tecnologías digitales. Esta frontera podría presumirse como el origen de la imposibilidad por tener las mínimas condiciones para la educación en materia digital, a saber, el mero acceso a los dispositivos computacionales y a las tecnologías mismas.

Micheli y Valle, por ejemplo, han afirmado en su estudio estadístico sobre el Índice de Desarrollo de las TIC (IDT) en México que el problema de la brecha digital está vinculado a la problemática del desarrollo local en su dimensión social y económica (2018, 51). Estados con mayores ingresos y distribución de productos tecnológicos tendrán mayor acceso a tecnologías complejas, mientras que comunidades con un bajo nivel de ingresos y sin acceso a las vías de distribución mercantil (tiendas, supermercados, comercios electrónicos, etcétera) tendrán menos posibilidades de acceder a tecnologías digitales, más allá de televisores, teléfonos celulares y otros dispositivos de orden cotidiano. En este sentido, podemos vernos conducidos a afirmar que la economía digital, así como sus consecuentes brechas, se constituye en un sistema que asocia consumo y producción.

En el apartado anterior afirmé que una limitante cognitiva restringe la posibilidad de hacer transparentes a las tecnologías de IA/AP y puse en evidencia que los niveles de opacidad epistémica dependen de la comprensión de los diversos agentes humanos respecto al funcionamiento de los sistemas. Hemos de reconocer que, para el desarrollo de las tecnologías, para la proyección del comportamiento de un dispositivo o para evaluar la robustez de un modelo es necesaria la existencia de expertos en materia computacional que puedan validar su funcionamiento.

Si bien he afirmado que para estos agentes expertos existen también diferentes grados de opacidad epistémica, concluí que las estrategias de XAI, los análisis de robustez y el conocimiento de la historia de implementaciones exitosas/no exitosas son medios para volver inteligibles las amplias dimensiones de los sistemas y verificar la seguridad en su uso. Pero, a su vez, argumenté que estos métodos no necesariamente elevan los índices de transparencia en la implementación de las tecnologías inteligentes.

El problema radica en que, en su mayoría, los agentes humanos no están facultados para validar el funcionamiento de los sistemas debido a que no cuentan con el nivel de pericia ni los conocimientos tecnológicos requeridos para comprender, en la medida que las capacidades cognitivas permiten, la operatividad de los sistemas. Habiendo afirmado que la transparencia técnica es una limitante a la que no podemos sobreponernos, resulta necesario poner a la falta de transparencia tecnológica como el problema real al que deben dirigirse las normativas que apuestan por un uso ético de la IA.

En otros términos, sensibilizar a las personas sobre los efectos de las tecnologías en aspectos esenciales de sus vidas es un paso necesario para abrir las cajas negras.

2.3 Abrir las cajas negras: conocimiento tecnológico contra autoridad epistémica

Hasta ahora, he afirmado que una de las grandes limitantes para la comprensión de las tecnologías inteligentes radica en que no conocemos el modo en el que operan los dispositivos informáticos. Sin embargo, he dejado en evidencia las limitantes materiales que imposibilitan que las personas contemos con los conocimientos necesarios para comprender a las tecnologías. En consonancia con esta idea, he defendido que esto es consecuencia de la falta de infraestructura en materia digital y de la incapacidad de las sociedades modernas para educar extensivamente a todas las personas en materia computacional. Esta afirmación pone en crisis la idea de que el conocimiento técnico es sinónimo de transparencia, y, al mismo tiempo, abre la puerta para pensar si existen otras derivas para disminuir la opacidad en el uso de la IA.

En primer lugar, hace falta hacer una diferencia entre el espectro de lo técnico y el de lo tecnológico. Etimológicamente, este término está compuesto de dos grandes conceptos griegos: *tecné* y *logos*. *Logos* es un término que hace alusión a la razón, mientras que *tecné* hace referencia a la técnica. Por lo tanto, podríamos decir que la tecnología puede comprenderse, al menos en origen, a la racionalidad de una técnica. Esta diferenciación nos permite establecer una diferencia entre las pautas o programas que sirven para llevar a cabo una acción (entendido como técnica) de un entramado mucho más complejo que implica la racionalidad que está detrás de ella.

Esta racionalidad podría entenderse, como hace Herbert Marcuse, como un proceso social en el que la técnica sólo tienen un papel parcial y en donde, en realidad, es necesario comprender la influencia de los aparatos técnicos de la industria que la acompañan (1941). Por ello, Marcuse afirma que:

La tecnología, entendida como un modo de producción, como la totalidad de los instrumentos, dispositivos y artefactos [contrivances] que caracterizan a

la era de las máquinas, es al mismo tiempo un modo de organizar y perpetuar (o cambiar) las relaciones sociales; una manifestación del pensamiento prevalente y de los patrones de conducta; un instrumento para el control y la dominación. (414)

Cuando hablamos de tecnologías de Inteligencia Artificial y de Aprendizaje Profundo podríamos suponer que hacemos referencia a los artefactos computacionales que sirven para ejecutar ciertas técnicas. En este caso, las técnicas se corresponderían con los programas, los algoritmos o las reglas para el procesamiento de la información. De lo anterior se sigue que las tecnologías requieren de técnicas para poder cumplir con un propósito, mientras que las técnicas necesitan de las tecnologías para efectuarse (el martillo no nos serviría para clavar si no poseemos la técnica necesaria para hacerlo y no podríamos clavar cosas con un martillo sin la herramienta).

Pero uno de los problemas en esta relación radica en que aplicar una técnica no solo implica la ejecución de una acción de acuerdo con un programa, sino que también involucra la intencionalidad, la proyección y la agencia humana. Es por ello que podemos conceptualizar a las tecnologías como sistemas mucho más complejos y abarcantes que no solamente se corresponden con los insumos materiales necesarios para ejecutar una tarea específica.

Como he afirmado, es común que hagamos referencia a la Inteligencia Artificial como “sistemas que pueden emular, aumentar o competir con el desempeño de humanos inteligentes en tareas específicas” (Vallor y Bekey, 2017, p. 339). Pero esta definición es controversial cuando introducimos el concepto de sistema y, más aún, cuando hablamos de sistemas computacionales o de Inteligencia Artificial. Esto se debe a que, al hablar de sistemas, a pesar de que sabemos que estos implican la interacción de diversos agentes o elementos, no tenemos claridad sobre qué cosas forman parte de ellos.

En el caso de la IA, sabemos que los sistemas computacionales involucran tanto a los algoritmos y reglas que hacen correr al sistema

como a los artefactos computacionales que sirven para procesar la información, o, en términos coloquiales, al software y al hardware. Pero, por sí mismos, estos dos elementos no son capaces de ejecutar una acción. Por el contrario, las tecnologías requieren de insumos externos para poder ser activadas y cumplir con sus funciones: necesitan abastecimiento eléctrico, un espacio físico, mantenimiento y, sobre todo, agentes que las usen. Esto significa que un sistema computacional es más que la unión de máquinas y de reglas. En cambio, un sistema computacional involucra aspectos mucho más abstractos y poco evidentes, tales como la intención de quien los usa, los objetivos sociales que cumple la tecnología y las condiciones económicas, culturales y políticas que están detrás de su desarrollo e implementación.

Esta distinción nos pone frente a dos limitantes que la búsqueda por transparencia en la IA debería poder superar: la opacidad técnica y la opacidad tecnológica. Dado que, como he afirmado en este capítulo, la limitante técnica resulta insuperable dadas las condiciones materiales, educativas y de infraestructura en la región, hace falta evaluar si frente a la limitante tecnológica tenemos vías para asegurar la transparencia.

2.3.1 Transparencia técnica y transparencia tecnológica

Un factor de desconocimiento que se suma a la falta de pericia técnica radica en que muchas personas utilizamos la tecnología con mínima o nula comprensión no sólo respecto a su funcionamiento, sino sobre las implicaciones de su uso y de dónde proviene. Este problema se aúna a que las estrategias de alfabetización técnica en materia computacional han promovido el desarrollo de programas formativos meramente instrumentales, que se centran en mayor medida en el aprendizaje para el uso de las herramientas que en las personas y sus necesidades (Travieso y Planella, 2008). Así, se ha privilegiado la alfabetización técnica de los individuos sin poner atención en el problema de la inclusión social. Ante esta omisión del factor humano en el uso de herramientas digitales, resulta

necesario comprender que las tecnologías se encuentran inscritas en contextos de producción y consumo que determinan su uso en nuestros circuitos sociales. En este sentido, las tecnologías de IA/AP forman parte activa dentro de los sistemas sociales en los que se implementan.

En este sentido, el problema epistémico de las tecnologías es más un desafío cultural y social que uno puramente tecnológico (Rojo, 2003) o, como lo he llamado, técnico. Así, en contraposición a la crítica de Ryszard Kapuscinski, hemos de afirmar que la tecnología no ha triunfado sobre la cultura (1995, 111), sino que ambos espectros se entretajan en un ensamblaje de agentes cuya vastedad apenas comenzamos a comprender. Entender que los usos de la IA están determinados no sólo por aspectos técnicos, sino culturales, es necesario para comprender cómo estas tecnologías se insertan en contextos específicos, así como para distinguir los problemas que conlleva su implementación para cada colectivo social.

Además de la revisión en la disponibilidad de las tecnologías con el que introduce el primer capítulo de esta investigación, el estudio de Gómez-González (2020) presenta una lista de los aspectos éticos y sociales de mayor relevancia presentes en las publicaciones examinadas. La necesidad por asegurar la privacidad de la información y la integridad de datos, así como las consideraciones en torno al efecto en el empleo a los profesionales de la salud son, por ejemplo, aspectos mencionados de manera constante en los reportes y estudios sobre el uso de IA en el campo de la medicina. Si bien la necesidad por regular las tecnologías figura en el rubro de consideraciones de relevancia particular para la aplicación de la IA en el área de la salud, llama la atención que el hacer accesible la información al público y los profesionales sea rara o nulamente mencionado como elemento necesario para su implementación [Tabla 3]. El estudio demuestra que la alfabetización tecnológica no figura actualmente para los desarrolladores de software de IA/AP como una de las áreas de oportunidad para la inclusión ética de las tecnologías en el área médica.

(G1) Actualmente bajo análisis, según se plantea en otras áreas de aplicación de la IA.	
Aspectos.	Cuestiones.
Privacidad, integridad de los datos.	Propiedad. Autorización para la recopilación, intercambio, minería, intercambio de datos.
Anonimato.	Ansiedad por la vigilancia.
Responsabilidad. Responsabilidad legal.	¿Quién es responsable en caso de mal funcionamiento?
Efectos en profesionales y empleo.	Pérdida de trabajos, nuevos trabajos. Cambios profundos en algunas especialidades médicas (algunas incluso pueden desaparecer). Necesidad de actualización profesional. Control de calidad, seguimiento.
Seguridad. Fiabilidad.	Vulnerabilidades Robo de datos. Manipulación de los datos utilizados para entrenar los sistemas.
Prestaciones.	Resultados de salud y circuitos clínicos mejorados. Reducción de errores médicos. 'Medicina personalizada'. Resultados psicosociales.
¿Humano "en-el-circuito" [de decisión]?	¿Debería un operador humano poder tomar el control de los sistemas de IA? ¿Incluso si el ser humano es más "propenso a errores"? ¿Qué pasa si no hay tiempo para actuar?
Aspectos.	Controversias.
Explicable ["explicabilidad"].	Actualmente requerido por la legislación. Algunos sistemas son (serán) demasiado complejos para ser entendidos por un humano. Pero pueden dar mejores resultados que un humano.
Confianza.	¿La "máquina" funciona mejor que un médico humano? ¿Qué hacer si ellos (sistema de IA, médico humano) dan opiniones contradictorias? "Estafadores de salud digital".
Calidad de datos. Sesgo / equidad.	¿Los sistemas de IA tienen sesgos / son justos con diferentes grupos (por ejemplo, étnicos, de género, de edad)? ¿Reciben datos adecuados y equilibrados para su aprendizaje? ¿Son válidos los resultados?
Empatía.	¿Decisiones compartidas? ¿[Un sistema de IA puede] ayudar (al humano) a tomar decisiones difíciles?
Opinión y participación ciudadana.	Bien común en la investigación financiada con fondos públicos, consentimiento informado, ciencia ciudadana. Reducción de la "asimetría" médico-paciente. Modelo "centrado en el paciente".
Prueba, comparativa.	¿Cómo evaluar los resultados? ¿Los procedimientos existentes para grupos promedio son válidos para tratamientos individualizados? ¿Comparación de sistemas de IA "contra humanos o máquinas"?
Regulación.	Retrasada respecto a la tecnología. Sin consenso internacional.
Asequibilidad. Impacto económico.	¿Tratamientos óptimos a precios "imposibles"? ¿Un factor de desigualdad? ¿Nuevos modelos de seguros y coberturas de salud?
Información para el público y profesionales.	Presión por nuevos productos. Avances reales frente a exageraciones e historias de éxito no confirmadas en áreas de gran interés (por ejemplo, curas para el cáncer). Riesgo de Medicina "basada en falsificaciones".
Decisiones de vida o muerte.	¿Deberíamos permitir que "una máquina" las tome (sobre nosotros, sobre un familiar)? El debate sobre los sistemas de armas autónomas letales.
Aspectos.	Temas importantes / conflictivos.
Humanización del cuidado.	Profesionales con IA: Más tiempo con el paciente, reducción de estrés. Sistemas de IA: [actualmente] carecen de examen físico / contacto con el paciente.
Ingeniería social, [selección de] perfiles basados en datos médicos, sanitarios y sociales fusionados.	Detección preventiva de eventos (por ejemplo, suicidio) frente a selección personalizada para marketing, seguros, atención médica y empleo. Cribado genético de la población.
Disponibilidad de múltiples datos (no supervisados, poco fiables), pruebas genéticas para cualquier persona.	Riesgo de Medicina "generada por el paciente".
¿Límites temporales al uso de datos? ¿herencia?	¿Uso post-mortem de información individual (por ejemplo, genética)?
Colaboraciones abiertas [crowd-sourcing] en algoritmos, potencia de procesamiento.	Intercambio libre de habilidades, conocimientos, experiencia. Solidaridad frente a riesgos de uso malicioso.
Lectura, decodificación de señales cerebrales.	Esperanza para personas con discapacidad severa frente a privacidad de aspectos básicos.
Interacción con procesos neuronales.	Ayuda para [personas con] enfermedades neurológicas y mentales frente al libre albedrío.
Edición genética como autoexperimentación.	Riesgo de resultados inesperados. Cambio de la herencia genética.
Edición genética de embriones humanos y embriones humano-animal.	Riesgo de resultados inesperados en recién nacidos. Creación de nuevos seres ("quimeras").
Los dos lados de la tecnología.	"Fácil" [conversión en] armamento. Alto riesgo de bioterrorismo.
Emulación / "trasplante" de cerebro.	Búsqueda de la inmortalidad. Definición de vida.
'Máquinas vivas' ('robots biológicos', 'biobots'). La búsqueda de formas de vida artificiales.	Definiciones de vida (natural, artificial) y muerte.
Beneficios frente a dificultades y riesgos.	¿Límites (o no) a la investigación y el desarrollo?

Tabla 3. Aspectos éticos y sociales de la inteligencia artificial y las tecnologías mediadas por la IA en Medicina y Salud ordenados en tres grupos (G1, G2 y G3). Se resaltan algunas cuestiones clave relevantes y conflictivas y controversias para cada aspecto.

Paralelamente, el equipo de investigación insistiría en que es necesario no pensar a las tecnologías solamente desde el éxito de sus resultados o desde un punto de vista técnico, sino a través de la revisión de aspectos éticos y sociales. Por ello, el estudio afirmararía que:

Si bien parece evidente que los criterios objetivos y profesionales deben definirse para un juicio adecuado, el impacto extraordinario de la Medicina y la Salud en los seres humanos y la sociedad trasciende una "evaluación técnica" y requiere tener en cuenta también los criterios culturales, éticos y sociales. Para aumentar la complejidad del análisis, la aceptación y la expansión potencial de las aplicaciones de IA se relacionan con equilibrios multifacéticos de efectos de bienestar y aspectos económicos, geográficos, incluso políticos, aunque están esencialmente vinculados con las cualidades personales, intangibles –basadas en la confianza– que configuran la relación (actual) de médico y paciente (2020, 13).

Así, este problema de la ignorancia en la disponibilidad tecnológica se aúna al del desconocimiento general por parte de los involucrados en la relación humano-humano y humano-IA de los papeles epistémicos que desempeñan y la importancia de las tecnologías como mediadoras entre ellos. En este sentido, Gómez-González afirmararía que con el uso de las tecnologías se configuran nuevas relaciones entre médicos y pacientes; entender la naturaleza de estas relaciones es, por tanto, un elemento fundamental para la comprensión de la tecnología y sus problemas epistémicos en su espectro ético y social.

En realidad, el desconocimiento del papel que desempeñan las tecnologías dentro de la sociedad es también un problema epistémico que concierne a una discusión sobre el nivel de transparencia de las herramientas y que forma parte de la falta de transparencia que las acompaña. En este sentido, es importante mencionar que esta forma de ignorancia no es un problema menos tecnológico que el del desconocimiento operativo y técnico de los dispositivos digitales; en

cambio, representa una expresión en el orden social del opaco papel que desempeñan las tecnologías en nuestras vidas.

Por lo anterior, afirmo que existe un problema de conocimiento y de acceso a la información que, además del analfabetismo técnico-digital compromete nuestro entendimiento de las tecnologías. Este no está dado por instancias puramente técnicas, sino que implica un desconocimiento de las estructuras políticas, económicas y sociales que permiten el afianzamiento de los sistemas tecnológicos en el mundo, a saber, una opacidad o falta de transparencia tecnológica.

A pesar de que hoy en día vivimos rodeados de tecnologías digitales es importante reconocer, como hemos afirmado hasta ahora, que en general no estamos epistémicamente equipados para tomar decisiones bien fundamentadas respecto al funcionamiento de las herramientas computacionales y que incluso en menor medida estamos habilitados para pensarlas de manera crítica. En este sentido, podemos afirmar que existe un entendimiento pobre de las características esenciales de la tecnología (Young, Cole y Denton, 2002), pero también de la influencia que tienen en nuestras sociedades y de cuáles son las agencias que dependen de nosotros mismos para afectar su desarrollo.

En su estudio sobre los índices de analfabetismo tecnológico en los Estados Unidos de Norteamérica, Young, Cole y Denton han afirmado que el americano promedio consume productos sin conocer su composición y sin tener consciencia sobre cómo han sido desarrollados, producidos, empacados y distribuidos. Por ello, a pesar de los altos niveles de producción y venta de tecnologías digitales en la región, de los altos índices de formación técnica y de la consecuente inclusión tecnológica que de estas economías se deriva, sería difícil afirmar que los niveles de analfabetismo en el país son equivalentes a los índices de consumo.

En cambio, resulta necesario pensar que el desconocimiento del modo en que opera la industria tecnológica es una seria limitante para la comprensión de la digitalidad. En este sentido, los tecnólogos afirman que la posesión de habilidades y conocimientos técnicos específicos no

garantizan la alfabetización tecnológica. Esto se debe a que incluso los agentes con niveles de pericia en la materia pueden no tener el entrenamiento o la experiencia necesaria para pensar las implicaciones sociales, políticas y éticas de su trabajo (2002).

Por el contrario, una mirada más amplia de la tecnología nos compromete con que el conocimiento de estos espectros debería ser tan valioso como el conocimiento técnico cuando hablamos de empoderamiento digital. Pero, para lograrlo, hace falta reinventar nuestra noción de tecnología y entenderla como un sistema.

2.3.2 La IA como sistema tecnológico

Una definición estándar de lo que es un sistema computacional dirá, en resumidas cuentas, que es el conjunto de dos ontologías principales llamadas software y hardware. La primera comprende a todos los elementos abstractos del sistema, como el algoritmo, los lenguajes de programación y los programas mismos. Por otro lado, el hardware comprende a los soportes materiales que tienen una existencia física, como los procesadores, los discos y las máquinas de cómputo.

En realidad, esta distinción clásica ha tenido, desde los orígenes de las ciencias computacionales, un amplio número de críticas y detractores. Por ejemplo, James Moor (1978) habría afirmado que la falsa dicotomía entre ontologías materiales y abstractas era uno de los grandes mitos de la computación. Por ejemplo, para Moor no existían programas como entes abstractos si no tenían una realización a través de soportes materiales. Por ello, para él ambos espectros ontológicos se complementaban y la relación existente entre ellos era, en resumidas cuentas, la única ontología que debería importar a la computación. Si bien esta histórica distinción tenía su validez en un espectro práctico, en realidad la frontera entre ambos lados de la ecuación no es tan clara.

Este problema de corte ontológico ha dado a lo largo de los años motivo suficiente para que diversos filósofos (Colburn y Shute, 1999; Irmak, 2012;

Duncan, 2017) hayan planteado múltiples categorías para diferenciar, o bien homologar, el software y el hardware como partes de un sistema complejo que, para funcionar, requiere de ambos espectros.

Para resolver esta aparente dicotomía, otra metodología ha sido utilizada para explicar la ontología de los sistemas computacionales. Apelar a la existencia de niveles de abstracción (LoA, por sus siglas en inglés) ha sido una de las más fecundas propuestas para abordar este problema. Los niveles de abstracción permiten comprender que un sistema tendrá elementos que son fácilmente perceptibles por el usuario de una computadora, por ejemplo, pero que también hay una serie de elementos menos tangibles, o más abstractos, que forman parte, del mismo modo, de la unidad del sistema computacional.

Así, el sistema podría comprender desde sus insumos puramente físicos y materiales hasta elementos tan abstractos como la intencionalidad en la creación del algoritmo o el lenguaje utilizado para su programación. Giuseppe Primiero (2016), por ejemplo, habría afirmado que un sistema computacional tendrá, al menos, cinco diferentes niveles de abstracción: la definición cuantitativa de bits, la construcción sintáctica de las operaciones, el significado de las instrucciones, el formato abstracto del algoritmo y la intención del desarrollador. Posteriormente, Primiero mismo, en colaboración con Nicola Angius y Raymond Turner (2021), afirmarían que estos niveles de abstracción se pueden resumir en seis elementos que comprenden la ontología de los sistemas computacionales digitales: [1] la intención como acto cognitivo que define los problemas a resolver por el sistema y que dependen de los agentes humanos que se los proporcionan; [2] especificaciones entendidas como la formulación de los requisitos necesarios para el cumplimiento de la asignación; [3] el algoritmo que expresa los procedimientos provistos al sistema para la ejecución de su empresa; [4] las instrucciones en lenguajes de programación de orden superior que fungen como la implementación lingüística del algoritmo propuesto; [5] las operaciones de la máquina que son el código compilado para la ejecución; y [6] la ejecución misma del

código que se da en el nivel físico dada la arquitectura material del sistema computacional para hacer correr el programa.

Una lectura de la ontología de los sistemas computacionales de esta naturaleza permite comprender que los sistemas no se cierran exclusivamente a los elementos técnicos de la construcción del mismo. No solamente el sistema comprenderá, entonces, a los programas de software y a los componentes físicos del hardware, sino que también será necesario incluir elementos tan abstractos como la intencionalidad de los desarrolladores, así como las elecciones de lenguaje o, incluso, los materiales de composición del soporte físico del computador. Podemos afirmar, entonces, que un sistema computacional no es solamente el dispositivo técnico que procesa la información, sino que, como sugiere Piotr Sztompka (1999), es un sistema abstracto que involucra otras esferas de la vida humana y que los pone en relación directa, por tanto, con sistemas complejos.

Como afirma James Steinhoff (2021), el desarrollo de la Inteligencia Artificial es producto de intereses en los que se entrelazan el capital, el trabajo y las plataformas que permiten el surgimiento de los entramados tecnológicos. Por su parte, Kate Crawford afirma que la IA/AP debe ser entendida como un atlas en el que se ven involucrados no solo los dispositivos técnicos mismos, sino múltiples sistemas de poder interconectados. Así, la IA puede ser utilizada para hablar de las formaciones industriales masivas que incluyen política, trabajo, cultura y capital (2021, 9). Es importante tomar en cuenta que estas condicionantes son las que, en gran medida, han determinado el inequitativo acceso a las tecnologías digitales y a la educación técnica. En suma, si las tecnologías computacionales únicamente consistieran en la suma de software y hardware, de técnicas y dispositivos, no existiría una explicación directa sobre las brechas digitales, el analfabetismo técnico ni la sectorización de productores tecnológicos. Por eso, el espectro político no está añadido a la tecnología, sino que la tecnología es producto de la política.

Al final, como afirma Ivan Illich (2005, 164), en un sistema, el usuario, el administrador, por la misma lógica del sistema, se convierte en parte de él. Sin embargo, el operador no exclusivamente el usuario final, sino también el agente productor, de distribución y de venta de las mercancías. Esto hace del sistema un entramado de agencias que raramente dependen del paciente y que, cada vez más, parten de los intereses de los poseedores de la industria tecnológica.

Por ello, cuestionar el papel de la política, de la industria y del desarrollo tecnológico en los marcos económicos actuales resulta necesario para comprender las dimensiones que se ven afectadas por el uso de sistemas inteligentes. Como hace Kapuscinski, es necesario comenzar a entender la estructura de red que tiene el uso de las tecnologías respecto a las condiciones económicas y de inequidad de las sociedades globalizadas actuales (2006, 34-35). Así, relaciones mutables y dinámicas sin puntos de referencia permanentes modifican el modo en que los sistemas tecnológicos contemporáneos ponen en relación a los agentes cognitivos humanos y a las máquinas inteligentes. Esto incluso puede derivar en propuestas de reconsideración de lo humano y lo técnico como partes de un sistema que integra a la política y al derecho (Albert, 2021, 224).

En el caso de México y de América Latina, es necesario recalcar la existencia de un colonialismo tecnológico que ha aprovechado las redes de producción algorítmica e informática globales para insertarnos en un circuito de consumo de tecnologías extranjeras (principalmente americanas y chinas) (Scasserra, 2021). Esto limita las capacidades de los agentes nacionales para acceder a información dirigida a los públicos hispanohablantes (así como a los nacionales que hablan otras lenguas distintas al español). A su vez, esto puede representar un riesgo en el uso de tecnologías no construidas para los perfiles médicos de la región, que producen sesgos y posibles errores en los sistemas.

Como afirma Renata Ávila Pinto (2018), las TIC, la innovación en inteligencia artificial y la capacidad de desplegar sistemas e infraestructuras rápidamente están en manos de unos pocos países que

compiten por mantener el control de la industria (p. 16). Nuevamente, la falta de infraestructura nacional, así como de “recursos, tanto [...] de capital (propiedad y control de cables y servidores, además de datos) como recursos intelectuales (los más avanzados técnicos e instituciones de investigación)” (Ávila, 2018), son también condicionantes de la falta de transparencia a la que nos vemos sometidos en su uso. En tanto las políticas de propiedad intelectual y algorítmica se rigen por códigos locales, en muchas ocasiones los usuarios de otras latitudes, como en el caso de México, quedan exentos de poder realizar evaluaciones y auditorías algorítmicas, imponiendo nuevas barreras para un uso ético, transparente y seguro de las tecnologías.

Dadas estas premisas, es importante señalar que este tipo de opacidad tecnológica se enmarca, además, dentro de los límites de los campos en los que los dispositivos digitales y de IA/AP son implementados. En este sentido, cada área de aplicación conlleva particularidades y distintas partes interesadas de relevancia, así como grados de opacidad que trascienden el aspecto técnico de las herramientas inteligentes. Así, el uso informado y transparente de estas tecnologías para la videovigilancia militar, para la educación inteligente o para la creación de obras de arte de manera autónoma requiere la comprensión del papel de diferentes actores que se ven involucrados en cada contexto de aplicación y que en muchas ocasiones son excluyentes entre sí: la educación primaria que utiliza plataformas de IA para la enseñanza no involucra la participación de un alto comando militar, mientras que un artista que utiliza un sistema de *DeepDream* para crear obras de arte probablemente no sea requerido para liderar una emboscada mediante las cámaras de vigilancia equipadas con arquitecturas de reconocimiento facial.

Dada la naturaleza de esta investigación centraré mi atención en cómo las tecnologías de IA/AP se insertan en la industria médica, así como en la comprensión social del funcionamiento tecnológico de los sistemas de salud en el país.

Una intuición común que reflexione sobre quiénes son las partes interesadas de relevancia al uso de tecnologías en los niveles de atención afirmaría que la relación médica primordial es la que existe entre el paciente y el médico. En la práctica, podemos presumir que los principales agentes de interés para la industria médica son, por un lado, los profesionales de la salud y, por el otro, los individuos que se ven atravesados por las prácticas de estos (incluidos en muchos casos los médicos mismos). Pero es importante comprender que las tecnologías no solamente están condicionadas por su uso de acuerdo con ciertos marcos epistémicos, sino también por los intereses que anteceden su producción, sus procesos productivos mismos e, incluso, por las prácticas que se dan fuera de estos marcos. En este sentido, sería desafortunado pensar que la compleja red entre agentes de relevancia se enmarca exclusivamente en la práctica común de gabinete. En realidad, el uso de un dispositivo médico no requiere la relación directa entre pacientes y doctores.

En este sentido, los desarrolladores informáticos, las mesas evaluadoras, los tomadores de decisión en la industria privada médica, los agentes legisladores en el terreno de la salud pública, así como los agentes de ventas, los médicos mismos y los pacientes, entre muchos otros, forman parte del entramado tecnológico que interesa al problema de la transparencia. Cómo estas relaciones afectan el desempeño de un sistema, qué efectos puede tener en la atención, el diagnóstico o el tratamiento del paciente y qué implicaciones socioculturales y económicas implica su despliegue son cuestiones que deberían poder responder los usuarios finales de las tecnologías para presumir que hay una relación transparente entre el sistema y las personas.

2.3.3 La autoridad epistémica como límite para la transparencia tecnológica

Como he remarcado en esta investigación, generalmente los usuarios finales de las herramientas de IA/AP desconocen la operatividad de los sistemas utilizados, lo cual limita la comprensión de la relación que se

establece entre tratante y paciente. Pero, en muchos casos, ninguno de los dos es consciente del funcionamiento de los dispositivos. En este sentido, la crítica a la posición de vulnerabilidad epistémica del paciente y del médico se ha vuelto un tema de discusión de importante actualidad.

De acuerdo con Alex Dragomir (2021), un problema fundamental del uso de los sistemas de IA en el área de la salud radica en que si los profesionales no comprenden cómo el dispositivo ha llegado a ciertos resultados, estos están imposibilitados para explicarle la naturaleza del diagnóstico a sus contrapartes. Así, el profesional de la salud que desconoce la operatividad del sistema se encuentra también imposibilitado para cumplir con sus obligaciones epistémicas respecto al paciente. Por ello, nos veríamos conducidos a afirmar que o bien hemos de apostar por el uso de tecnologías que prometen un alto índice de éxito y precisión en sus resultados a pesar de que no conozcamos su funcionamiento o bien a asegurar el compromiso epistémico y renunciar al uso de estas tecnologías en el área de la salud.

En paralelo al problema de la autoridad epistémica del tratante, la ininteligibilidad de las tecnologías nos compromete con que el paciente pierde autonomía pues no posee la información requerida para comprender el resultado de la evaluación y, por otro lado, no existen herramientas para hacérselo explicable (Dragomir, 2021). De acuerdo con el pensamiento de Tom Beauchamp (2009), podríamos sugerir que este fenómeno es un tipo de paternalismo, caracterizado por la relación dispar entre el paciente y el profesional, que se traduce en que el médico tiene un entrenamiento y conocimiento superiores, lo cual lo sitúa en una posición de autoridad para determinar los intereses de aquellos que caen bajo su cuidado y administración (2009, 79). Así, el usuario final/paciente que desconoce el funcionamiento del sistema se ve comprometido a asumir una posición de confianza o sometimiento respecto a la autoridad epistémica del profesional de la salud, sea el caso de que este último tenga la información para llevar a cabo una toma de decisión médica informada o no.

Carel y Kidd (2014) han afirmado que en estos escenarios el paciente sufre, además, una vulneración debido a la injusticia epistémica que es producto del desconocimiento y la falta de elementos necesarios para entender las condiciones de interacción tecnológica y para comunicar sus propios intereses. Así, aunque el médico no poseyera las herramientas para entender a la tecnología y asegurar sus resultados, sí sería poseedor de una autoridad epistémica que lo sitúa en una posición de privilegio sobre el paciente.

Podría sugerirse que la toma de decisión informada en materia de IA funciona como deriva de solución al problema de la “medicina de caja negra”, incluso cuando exista un límite en la transparencia técnica respecto a los sistemas. Si bien Bjerring y Busch asumen que la IA conlleva un índice de opacidad que, como hemos argumentado en este capítulo, no puede ser sorteado por ningún agente humano, en realidad todas las tecnologías empleadas en el área médica implican una no transparencia. En este sentido, el uso de un estetoscopio, de un nuevo reactivo para la detección de enzimas en la sangre o de un sistema de IA/AP para la interpretación automatizada de ultrasonidos no se diferencia desde un punto de vista epistémico. En realidad, en mayor o en menor medida los dispositivos se presentan con un índice de opacidad que, en todos los casos, depende de los conocimientos y capacidades cognitivas del sujeto interactuante.

Pero el factor de relevancia que diferencia a las tecnologías de IA respecto a otras herramientas analógicas radica en el conocimiento de los principios básicos de operación del sistema y de los riesgos que implica su uso. Como hemos afirmado, las tecnologías de IA/AP están acompañadas de una opacidad epistémica que imposibilita su transparencia para los agentes humanos, contraria a la de muchas otras que forman parte de los conocimientos básicos de las prácticas médicas; por ello, sólo el reconocimiento de estas limitantes puede ser propuesto como el medio para la creación de tomas de decisión médica informadas.

Esto, evidentemente, sostiene la posición de que los sistemas no son oscuros en sí mismos, sino que las capacidades de comprensión de los agentes humanos relativiza los índices de opacidad epistémica y transparencia de los sistemas. A su vez, implica que los agentes que interactúan con las tecnologías pueden extender su comprensión de los sistemas para hacer un uso responsable de su autoridad epistémica en consonancia con los límites de sus capacidades cognitivas. La formación de recursos humanos capacitados en materia tecnológica en el área de salud resulta, por tanto, una necesidad de primera importancia.

Pero la vía de solución al conflicto del desconocimiento general de la operatividad de los sistemas radica no sólo en el entrenamiento de los practicantes de la salud en materia digital y tecnológica, sino, principalmente, de todo agente que se encuentre atravesado por las prácticas médicas. De forma paralela a la limitante del analfabetismo técnico, el problema radica en cómo hacer de dominio público la información necesaria para el desciframiento de las tecnologías a conjuntos no homogéneos de individuos (de diferentes contextos culturales y económicos, latitudes geográficas, comunidades tecnológicas y lingüísticas, etcétera).

Esta problemática se ha presentado como un reto que aspira a que todo usuario pueda comprender las dimensiones sociales del uso de las tecnologías específicamente en los contextos médicos y de la salud. Sin embargo, un problema práctico que acompaña a estas propuestas radica en que una agenda en materia digital no puede contrarrestar los efectos del repartimiento tecnológico y económico a corto ni mediano plazo. En realidad, podríamos esperar que una estrategia de alfabetización técnica y tecnológica requiera varias décadas tan sólo para el desarrollo de la infraestructura requerida para la inclusión digital de los habitantes del territorio. Como afirma Kapuscinski, cada revolución tecnológica requiere que los individuos revisen sus marcos conceptuales, y esto generalmente toma tiempo (2003, 13). Por ello, podríamos vernos conducidos a buscar normativas que trabajen mientras se generan agendas de promoción

técnica y tecnológica, que protejan la integridad de los pacientes y que, al mismo tiempo, provean las herramientas para la defensa de nuestro derecho a la información y a la toma de decisión informada.

En el panorama mexicano, he abordado en el primer capítulo las iniciativas de inclusión digital y la problemática de la inclusión social requerida para una efectiva alfabetización técnica. Este abordaje ha evidenciado, a su vez, la falta de instrumentos legales y administrativos por parte de las dependencias públicas para enfrentar estos retos. Es evidente que, como afirman Travieso y Planella, a lo largo de los años se ha visto que las instituciones públicas han evitado su responsabilidad como garantes de la cohesión social y del desarrollo equitativo de la sociedad (2008). Por ello, es importante notar que, ante este panorama, han sido las organizaciones civiles y de protección de los derechos digitales las encargadas de luchar por sociedades tecnológicas justas y democráticas, mientras que las empresas han impulsado el desarrollo de manuales de buenas prácticas para responder a la falta de iniciativa estatal para la regulación de los sistemas y la creación de agendas de interés público.

Este interés por parte de la industria privada ha estado en tensión con los intereses productivos de las empresas, quienes han creado sus manuales y promovido la creación de normativas que regulen el uso de las tecnologías con un interés en resguardar su propiedad intelectual. Desde el punto de vista del usuario, esto se traduce en una inaccesibilidad al funcionamiento de las herramientas, al código que sirve para su programación e, incluso, a las bases de datos que alimentan al sistema. Esto ha condicionado, en gran medida, un tercer tipo de falta de transparencia caracterizado por haber sido deliberadamente creado por las corporaciones (o por los mismos desarrolladores) debido a políticas de seguridad, propiedades intelectuales o códigos comerciales.

En consonancia con lo anterior, es importante pensar cómo las regulaciones en nuestro contexto legislativo pueden enfrentar el reto del acceso a la información no sólo frente a los riesgos que conlleva el uso de

las tecnologías, sino para la toma de decisión informada por parte de los usuarios finales de los sistemas de salud. Por ello, en el siguiente capítulo exploraré los medios y alcances actuales para una gobernanza tecnológica que promueva, de acuerdo con los límites marcados hasta ahora, la transparencia en el uso de tecnologías de IA/AP como dispositivos médicos en el país.

Para ello, en el siguiente apartado analizaré el papel que las instituciones públicas han de asumir en el país para la efectiva y transparente gobernanza en materia digital, así como las estrategias implementadas por las industrias y ONG como medio para asegurar un uso responsable y seguro de las tecnologías de IA/AP.

CAPÍTULO 3. Materiales para una regulación de sistemas de IA/AP en México

La Inteligencia Artificial está transformando a todos los sectores de la economía a lo largo del globo. Sin embargo, es importante notar que en México existe una falta de coordinación institucional en materia tecnológica, tanto a nivel estatal como por parte de la industria que fomente el uso seguro, ético y confiable de la IA. Este fenómeno nos invita a establecer estrategias de largo plazo para el desarrollo de tecnologías de Inteligencia Artificial que logren aprovechar las oportunidades y minimizar los riesgos que estas conllevan. De entre estos riesgos podríamos afirmar que resolver su carácter de “cajas negras” es una de las más grandes necesidades para garantizar la seguridad en el uso de estos sistemas. Pero ¿cómo podemos apostar por un desarrollo sustentable y seguro de estas tecnologías en México?

Una propuesta radicaría en promover el incremento en la investigación y el desarrollo de Inteligencia Artificial para disminuir la dependencia de tecnologías que provengan del exterior. Esto naturalmente requeriría un tránsito de una economía basada en la mano de obra a una basada en el conocimiento y el manejo de la información. Adicionalmente, dos aspectos de inminente relevancia serían propuestos por Luis Enrique Sucar Succar durante el Foro “Aplicaciones de Inteligencia Artificial” para proyectos de impacto social (2021) como necesidades para el uso y desarrollo de estas tecnologías en el país: [1] la educación y capacitación de la población para prepararse a las nuevas oportunidades tecnológicas y [2] la adecuación de nuestros marcos legales y políticos para su efectivo aprovechamiento.

A lo largo de esta investigación he demostrado la importancia de ofrecer a las personas y usuarios finales de tecnologías de aprendizaje profundo y automático en el área médica las herramientas cognitivas y la información necesaria para tomar decisiones respecto a su propia salud

como una respuesta al problema de la opacidad epistémica. Esta propuesta ha surgido como resultado de las imposibilidades que conlleva la aspiración por dotar de conocimiento técnico y tecnológico a la población en el país. Pero, hasta ahora, no hemos dado respuesta a una pregunta fundamental que debería regir el núcleo de una regulación en materia de IA/AP: ¿qué implica dar a los pacientes información para llevar a cabo la toma de decisión respecto a su propia salud?

Si, como he afirmado en el capítulo anterior, asumimos que un paciente está sometido a una relación de paternalismo epistémico y que no cuenta con los conocimientos técnicos ni tecnológicos para evaluar el uso de una tecnología de IA/AP dentro de la atención que se le brinda, la respuesta al problema de la opacidad epistémica no debe residir en aquellas falencias que, presumiblemente, han condicionado sus capacidades cognitivas y que determinan sus niveles de analfabetismo en la materia. Por el contrario, una propuesta de regulación que considere dar respuesta al problema de la falta de transparencia en el uso de la IA debe tomar en cuenta qué es lo que un paciente considera necesario para llevar a cabo una toma de decisión.

La definición de esta opacidad formulada al inicio de mi investigación, y que fue tomada de Humphreys, afirma que un proceso es epistémicamente opaco en relación con un agente X en un momento t en caso de que X no conozca todos los elementos epistemológicamente relevantes del proceso (Humphreys, 2009, p. 618). De este modo, el objeto de interés es cuáles son los elementos epistemológicamente relevantes que un agente X (entendido como paciente) debe conocer para que [1] el sistema no sea epistémicamente opaco y [2] pueda tomar una decisión informada respecto a su propia salud.

Estrictamente no existe una única respuesta sobre qué información debería proveer un prestador de servicio de salud para que los pacientes tomen decisiones informadas. E, incluso, en muchos casos las prácticas médicas centradas en el paciente se ven imposibilitadas ante la incapacidad por integrar las necesidades, valores y deseos de las personas

en la toma de decisión, ya sea debido a la falta de infraestructura, el contexto cultural o la incapacidad de empoderamiento por parte de los pacientes. Sin embargo, sí existen algunos tópicos comunes que podemos hallar en las discusiones sobre la naturaleza de las prácticas médicas que privilegian la autonomía en la gestión de salud del paciente.

Proponer un abordaje normativo para la regulación de estas tecnologías no puede por ello simplemente concentrarse en asegurar que los usuarios finales poseen toda la información relevante para tomar decisiones respecto a su propia salud. Por el contrario, el esfuerzo regulatorio debe tomar en cuenta los marcos normativos existentes y las necesidades en materia legislativa para asegurar la idoneidad de la regulación al contexto en el que se aplica.

En la primera parte de este último capítulo plantearé dos alternativas complementarias para afrontar el problema de la opacidad que conlleva el uso de estas tecnologías y que tienen como objetivo afianzar un uso seguro y responsable de los sistemas: el principio precautorio y la gestión de riesgos.

Para ello, partiré de que el conjunto de los elementos epistémicamente relevantes de un proceso o sistema son contingentes y relativos a las necesidades de diferentes agentes humanos; por tanto, argumentaré que la transparencia buscada en defensa del derecho al conocimiento también debe considerar hacia quiénes se dirige y de qué modo se puede promover. Afirmaré que, de acuerdo con esta hipótesis, las características de la opacidad epistémica de cada contexto tecnológico conllevan limitantes para la búsqueda de transparencia, explicabilidad e inteligibilidad que deben ser consideradas en estas regulaciones.

Tras llevar a cabo esta ponderación, en el último apartado elaboraré algunas nociones, conceptos y temas que deberían formar parte de una estrategia de regulación en el uso de sistemas de IA/AP en el área de la salud como respuesta a algunas de las controversias propias de la propuesta de la presente investigación. En última instancia, defenderé que las políticas públicas en materia de regulación de sistemas de IA en el área

de la salud no pueden ni deben comprometerse con una absoluta transparencia, explicabilidad e inteligibilidad en su uso si quieren promover al mismo tiempo una práctica médica centrada en el paciente en la que este sea capaz de tomar decisiones informadas respecto a su salud.

3.1 El principio precautorio como herramienta de previsión estatal

Como he revisado, al relacionarnos con los sistemas de Inteligencia Artificial y de Aprendizaje Profundo existe una limitante cognitiva para poder rastrear los valores de entrada en su totalidad y el procesamiento interno de la información. Por tanto, una regulación en el uso de sistemas de IA/AP como dispositivos médicos no puede comprometerse solamente con asegurar una transparencia técnica basada en que “suficiente información sea publicada o documentada antes del diseño y uso de una tecnología de IA”, como sugiere la OMS, sino con una ponderación de riesgos debido a la inabarcable operatividad de los sistemas. Esto significa que la información presentada a los usuarios no debe ser exhaustiva, continuando la idea de que el desarrollo de factores técnicos es suficiente para que entiendan el funcionamiento de un sistema, sino solamente suficiente. Y esta debe poder expresar en términos claros la utilidad e incertidumbre de los sistemas, los riesgos de caer en sesgos, la atribución de responsabilidades y el papel de la IA en la toma de decisión clínica (Farah, et. al., 2023).

Frente al problema de la opacidad epistémica podríamos optar por dos principales vías para otorgar información relevante que les ayude a tomar decisiones informadas respecto a su salud: [1] afirmar que la única información de relevancia para un paciente debería ser la seguridad y eficacia que ofrece un sistema para llevar a cabo su tarea de manera correcta o [2] presumir que el interés de los pacientes debería girar en torno a los riesgos que implica someter su diagnóstico, tratamiento o evaluación al uso de estas tecnologías.

A pesar de que esto puede parecer una simple diferencia de grado entre el binomio seguridad-riesgo, lo cierto es que es un matiz que las regulaciones deben tratar como piedra angular de una posible normatividad. Y esto ha sido precisamente uno de los tópicos más discutidos en materia legislativa a la hora de tratar con tecnologías que pueden representar un riesgo a la salud e integridad de las personas.

Una primera alternativa para ofrecer protección legal a los miembros de la comunidad nacional por parte de las instituciones gubernamentales y legislativas consistiría en apelar a la creación de mesas de expertos evaluadores que estimen el riesgo que un paciente podría enfrentar al someterse a un tratamiento o proceso médico que incluya la intervención de tecnologías de IA/AP. Esto permitiría que los pacientes y usuarios finales únicamente puedan considerar dentro de sus opciones aquellas que hayan demostrado evidencia de funcionar con un buen margen de seguridad.

Esta medida de prevención podría ser considerada como una modalidad del «principio precautorio». Este principio es una herramienta jurídica que desde el siglo pasado ha servido en diversos contextos legislativos para advertir sobre los posibles efectos de un peligro ambiental o de salud que pueda afectar el bienestar de un ecosistema o de un conjunto de individuos.

La Declaración de Wingspread sobre el Principio de Precaución presentada por la Red de Salud Ambiental y Científica (SEHN) (1998) estipula lo siguiente:

...cuando una actividad se plantea como una amenaza para la salud humana o el medio ambiente, deben tomarse medidas precautorias aun cuando algunas relaciones de causa y efecto no se hayan establecido de manera científica en su totalidad. En este contexto los proponentes de una actividad, y no el público deben ser quienes asuman la carga de la prueba.

Esta declaración deja ver que la responsabilidad en la gestión de asuntos ambientales, de salud humana y de biotecnología debe corresponder a los

desarrolladores, distribuidores y a la industria en general que propone la creación o uso de nuevas tecnologías (Medina, 2022, 28). En el caso del uso de sistemas de IA/AP en el área de la salud, la responsabilidad precautoria correspondería entonces a las instituciones públicas que las implementan, así como a los agentes detrás de su desarrollo.

Si bien en sus orígenes el principio precautorio surgió como respuesta a los peligros ambientales de siglo pasado, como afirma Lourdes Medina, “el medio ambiente no es el único bien jurídico tutelado por el principio precautorio” (Medina, 2022, 31). Por el contrario:

...este principio tiene como misión y función informar al ordenamiento jurídico en su dinámica, sirviendo de guía, igualmente a los operadores y funcionarios que interactúan [con base en] la legislación y otras normas, sobre temas del medio ambiente, salud, seguridad alimentaria o biodiversidad que son los bienes jurídicos protegidos donde preponderantemente encuentra su radio de acción el PP.

Si bien esta figura legal no es unívoca a lo largo de diferentes territorios y legislaciones, suele ser considerado un principio general de la ley. Algunos tratados, como los de la Unión Europea, protegen esta herramienta como un recurso para enfrentar situaciones que ponen en peligro la integridad humana y de los ecosistemas. La flexibilidad de este principio hace que figure como una alternativa funcional para la regulación en el uso de sistemas de IA/AP en el área de la salud en el país a falta de marcos normativos aptos para su tratamiento.

Es precisamente el Parlamento de la Unión Europea quien ha ratificado el principio precautorio como una herramienta encausada a “evitar impactos adversos en situaciones de incertidumbre científica” (2015, 6), pero que no necesariamente compromete a los representantes de un contexto legal a emitir prohibiciones o medidas regulatorias, sino alertas sobre los efectos que puede tener su uso, consumo o comercialización. Así, el principio precautorio se basa en la promoción de tomas de decisión informadas, transparentes y responsables que ayuden a las personas en

situaciones en las que la evidencia científica no es suficiente para asegurar que un producto, sustancia o dispositivo producirá un daño en el futuro.

En el Séptimo Programa de Acción Ambiental del Parlamento de la Unión Europea, se definen a las acciones en torno a este principio de la siguiente manera:

...los legisladores y las partes interesadas deben tener bases más informadas que les permitan desarrollar e implementar políticas ambientales y climáticas que incluyan un mejor entendimiento de los impactos ambientales de las actividades humanas y la medición de los costos y beneficios de sus acciones y de sus inacciones (2013, 19).

Este principio se ha vuelto popular en los últimos años debido a la importancia que ha tenido en el control y alerta pública sobre los efectos de los organismos genéticamente modificados, la rápida adaptación de bacterias y su multiresistencia a fármacos, las afectaciones provocadas por el cambio climático, los daños producidos por el asbesto en humanos y la extinción de especies de abejas y otros organismos. Si bien muchos de estos fenómenos no cuentan con evidencia científica que asegure un daño directo al ambiente o a la salud humana, el principio precautorio no exige la certeza científica absoluta de que existe una relación entre causas y efectos (Medina, 2022, 34). En estos casos, vale la carga de una perspectiva de daño grave o irreversible para invertir la carga de la prueba, o bien, a probar la ausencia de riesgos potenciales de una acción.

Es por ello que el principio precautorio podría ser una respuesta para superar la opacidad epistémica de las tecnologías de IA/AP utilizadas en el área médica: si un paciente recibe información sobre los potenciales riesgos en el uso de un sistema de esta naturaleza, en teoría debería ser capaz de tomar una decisión informada respecto a su uso para la administración de su bienestar.

En el caso de la IA/AP la evidencia de su funcionamiento es denotada con los reportes de éxito del sistema. No hace falta más que corroborar que una tecnología cumple con su objetivo de manera reiterativa y

constante para presumir que se seguirá comportando del mismo modo a lo largo del tiempo. Pero esto no evita el problema epistémico que he planteado a lo largo de la investigación: la incertidumbre que produce su uso debido a la complejidad de sus cálculos y la indescifrabilidad de su operatividad. La idealidad del principio precautorio como medida de control en estos casos radica en que la desconfianza en su operatividad no depende de la evidencia que aportan las ciencias computacionales, sino en la opacidad que produce su funcionamiento respecto al usuario final.

Este mismo principio sería extendido a la tecnología en el Séptimo Programa de Acción Ambiental del Parlamento, donde afirman:

...hay una necesidad por entender mejor los potenciales riesgos ambientales y a la salud humana asociados a las nuevas tecnologías con el fin de evaluarlas y manejarlas de mejor forma. Esta es una condición previa para la aceptación pública de las nuevas tecnologías, así como para la capacidad de la Unión [Europea] para identificar y responder a potenciales riesgos asociados con el desarrollo tecnológico de forma pronta y efectiva. Las innovaciones tecnológicas deben ir acompañadas de diálogos públicos y de procesos participativos (2013, 19).

Estrictamente, en el contexto mexicano no existe una figura de derecho constitucional que proteja el acceso al principio precautorio y mucho menos en materia de salud. De acuerdo con el estudio realizado por Lourdes Medina sobre la aplicación del principio precautorio en México, el estado únicamente está obligado a cumplir con tres tratados en materia ambiental y que son vinculantes para el derecho nacional:

1. El Convenio sobre la Diversidad Biológica de 1992 (que incluye a los protocolos de Cartagena y de Nagoya),
2. La Convención Marco de las Naciones Unidas sobre el Cambio Climático de 1992, y
3. El Convenio de Estocolmo sobre Contaminantes Orgánicos Persistentes, de 2001.

Estas obligaciones adquiridas en materia de derecho internacional dejan en evidencia que, hasta ahora, el principio precautorio no es un recurso ampliamente utilizado y no tenemos testimonios de su aplicación en el área computacional. Sin embargo, en el país existen diversos expedientes de aplicación del principio para la regulación tecnológica que, aunque no ha sido catalogada de este modo, apela a la creación de medidas preventivas en la producción, uso y venta de tecnologías, como son los organismos genéticamente modificados, pesticidas, macrogranjas, desforestación y hasta publicidad en salud (Medina, 2022).

El sostén legal para su uso radica en la protección a las garantías individuales y derechos humanos de todos los mexicanos, por lo que el principio precautorio ha sido generalmente utilizado mediante juicios de amparo tanto por particulares como por agencias estatales y colectivos, principalmente indígenas.

Para empezar, en el tercer párrafo del capítulo primero de la Constitución Política de los Estados Unidos Mexicanos se asegura que, al menos desde su modificación en el 2011, “el Estado deberá prevenir, investigar, sancionar y reparar las violaciones a los derechos humanos” en los términos de la ley.

Esta obligación adquirida por el Estado compromete a las instituciones a promover estrategias de prevención ante una serie de escenarios que pongan en peligro la salud, vida e integridad de todas las personas, incluido el uso de las tecnologías que interesan a esta investigación. Es en este punto en donde entra en juego el principio precautorio como una herramienta útil para prevenir, investigar, sancionar y reparar las potenciales violaciones a, al menos, el derecho a la vida, la integridad y seguridad personales, los derechos sexuales y reproductivos, la salud y la verdad, entre otros, con especial énfasis en el derecho a la información.

Mi objetivo a lo largo de la investigación ha sido promover la tesis de que la transparencia técnica y tecnológica en el uso de sistemas de IA/AP en el área de la salud no solo no es posible, sino que, en muchos casos, es

indeseable. Por ello, una legislación que se comprometa con la reducción (o, incluso, la eliminación de la opacidad epistémica) debe asumir un compromiso por hacer de dominio público la información necesaria para que un paciente pueda tomar decisiones informadas respecto a su propia salud. Presumiblemente, esta información debe ser relevante para conocer los riesgos que implica su uso, que, por motivos que trataré más adelante, es el factor de relevancia epistémica más importante a la hora de tomar una decisión en materia de salud.

En el caso de las tecnologías de IA/AP, el riesgo no radica en que el sistema pueda generar resultados imprevistos a lo largo del tiempo. Por el contrario, una de las ventajas de estos sistemas es que tienden a tener un comportamiento estable y cada vez más preciso. Si bien los casos de falsos negativos y positivos, así como los sesgos raciales y de género, son detonantes de peligros a la integridad de los individuos durante un proceso de tratamiento o diagnóstico, la relación que estos factores guardan con el problema de la opacidad epistémica es principalmente de acceso a la información. En este sentido, el principio precautorio únicamente funcionaría como una medida provisional para advertir sobre la falta de acceso a la información, así como los riesgos que esto puede significar a la hora de tomar decisiones respecto a la salud de las personas. Desde un enfoque centrado en el paciente, este principio debe empoderarlo para actuar y decidir autónomamente.

La urgencia de aplicación de un principio precautorio en México respecto al uso de estas tecnologías radica en su novedad y en la falta de conocimiento público sobre su implementación y potenciales riesgos. Este factor permite asegurar una falta de certidumbre científica sobre sus efectos, que, de acuerdo con la literatura en derecho comparado e internacional, es necesario para la aplicación del principio.

Los efectos de la aplicación de este principio dependerán de las necesidades y recursos públicos en materia de salud, que pueden ir desde una prohibición en el uso de las tecnologías en tanto no se asegure el acceso a la información, la inclusión de indicaciones visuales para los

pacientes cuando una tecnología de este tipo sea utilizada durante su atención médica o la obligación de los prestadores de servicios de salud para hacer de conocimiento del paciente el uso de las tecnologías.

Hasta este punto, el principio precautorio parecería ser una respuesta ideal para afrontar las limitantes epistémicas que conlleva el uso de sistemas de IA/AP. A partir de la inclusión colectiva, los legisladores y partes interesadas en el uso de las tecnologías podrían democratizar la toma de decisión y contribuir al uso responsable de las tecnologías, pero esto solo funcionaría como una solución temporal y no determinante, ya que, a final de cuentas, no reduce la opacidad epistémica ni asegura algún tipo de transparencia en el uso de las tecnologías para el usuario final. Asimismo, es incapaz de por sí mismo crear protocolos de acción definidos para asegurar el acceso a la información.

El principal riesgo que se corre al aplicar este principio en el manejo de las tecnologías de IA/AP radica en que la evaluación de seguridad y riesgos en su uso no depende de la opinión pública, sino del asesoramiento de expertos computacionales, desarrolladores informáticos, médicos y otros especialistas (como, acaso, el escritor de esta tesis). Esto resulta problemático para sortear el fenómeno de la opacidad epistémica ya que son nuevamente los agentes expertos quienes estiman los parámetros de fiabilidad en un sistema tecnológico y no los usuarios finales de la tecnología, quienes pueden ver sesgado su juicio para tomar decisiones respecto a su salud.

Por ello, si queremos asegurar que los pacientes obtengan la información necesaria para tomar una decisión (entendida como el elemento epistémicamente relevante del proceso) hace falta superar las acciones tomadas para la protección del usuario y posibilitar que este tome decisiones respecto a su propia salud. Para lograrlo, la respuesta estatal puede resultar insuficiente, por lo que el esfuerzo regulatorio debe recaer también en la esfera de la responsabilidad social de las empresas y particulares operantes en el territorio nacional.

3.2 La gestión de riesgos como agencia de responsabilidad social

De acuerdo con un estudio publicado por Carmen L. Lewis y Michael Pignone (2009) en el North Carolina Medical Journal, la toma de decisión informada por parte de un paciente es especialmente compleja cuando este no tiene acceso a información médica o cuando no la entiende por efecto de su carencia de alfabetización en materia de salud (y, podríamos añadir, tecnológica). De acuerdo con Lewis, el paciente (o su representante en caso de existir alguna discapacidad física o cognitiva que le impida considerar o tomar una decisión autónoma) debe conocer las opciones disponibles para su tratamiento y los posibles resultados de cada elección; además, debe ser otorgada la oportunidad de que considere y evalúe sus propios valores en relación con el tratamiento elegido.

Un factor relevante para tener en mente es que para el Parlamento Europeo el principio precautorio y la buena gobernanza deben ir de la mano (2015, 19). Es por ello por lo que este principio debe estar acompañado de una gobernanza de riesgos en la que tanto la sociedad como la administración estatal tomen decisiones en relación con los costos y beneficios en el uso de las tecnologías. Esta gobernanza debe contemplar, al menos, los tres siguientes aspectos:

1. Evaluación de riesgos, que consiste en la identificación, determinación y caracterización de los posibles efectos negativos ambientales o a la salud de un fenómeno, producto o agente.
2. Gestión de riesgos, que permite a los responsables diseñar planes e implementar acciones que reduzcan, transfieran o contengan los riesgos.
3. Comunicación de riesgos, que contempla las acciones necesarias para exponer a la población objetivo los riesgos identificados y las afectaciones que pueden tener en ella.

Si bien es común que el principio precautorio funcione como una estrategia proteccionista en favor de las comunidades vulneradas tecnológica y socialmente, lo cierto es que este recurso es una forma de paternalismo que impide la toma de decisión autónoma por parte de los pacientes. Por ello, resulta contraproducente pensar esta licencia proteccionista como la solución al problema de la opacidad epistémica.

Por su parte, en el año 2019 la Organización Internacional para la Estandarización hizo pública una versión revisada de la ISO 14971 “Dispositivos médicos/productos sanitarios (MD) Aplicación de la gestión del riesgo a los MD” en compañía del posterior reporte técnico ISO/TR 24971 “Productos sanitarios. Orientación sobre la aplicación de ISO 14971”, en los que se incluye la categoría de software como dispositivo médico. En consonancia con el plan de acción de la FDA, los documentos de estandarización afirman que, para enfrentar la no transparencia de las tecnologías, el etiquetado de los dispositivos podría ser una vía de concientización pública sobre los riesgos que implica el uso de los sistemas. En aras de promover la seguridad, la industria de dispositivos y fármacos, por ejemplo, ha impulsado la categorización de los recursos de acuerdo con estándares de riesgo comprensibles para los usuarios. Esta misma idea, que acompaña a la propuesta de la FDA, se ha planteado como un recurso útil para la presentación sintética de los riesgos y posibles efectos que conllevaría el uso de las tecnologías. En su reporte técnico, la ISO afirma que para promover el acceso a esta información para la seguridad en el uso de los dispositivos (entre los que podemos incluir a las tecnologías de IA/AP) resulta necesario considerar:

- Si la información será provista directamente al usuario por parte del desarrollador o si tendrá que involucrar la participación de terceros, como instaladores, asistentes y cuidadores, profesionales de la salud, directores de laboratorio o farmaceutas, y eso tendrá implicaciones para el entrenamiento;

- Comisionar y transferir al usuario final en atención a si la instalación o manejo de los dispositivos puede ser llevada a cabo por personas sin las habilidades requeridas;
- Basado en el tipo y esperanza de vida de los dispositivos médicos si el re-entrenamiento o la re-certificación de los usuarios o del personal es necesario (2020, 33)

En este sentido, la ISO propondría una evaluación de los alcances epistémicos de los usuarios finales para ponderar en qué medida el acceso a la información y operatividad de las tecnologías debería ser asegurado. En torno a la pregunta sobre cómo debe ser exhibida la información, la norma establece que deben tomarse en cuenta diversos factores, como “la visibilidad de los ambientes, la orientación, la capacidad visual del usuario, las poblaciones y perspectivas, la claridad e la información presentada, unidades, códigos de colores y la accesibilidad a información crítica” (2020, 34). Así, la apertura a la información es relegada a los desarrolladores, que tienen que considerar paralelamente los riesgos de mal manejo que implica el acceso a la información o su acceso de manera no autorizada a datos y funciones de los dispositivos (2020, 35). Ahondando en estas consideraciones, la Organización afirma que para la gestión de riesgo de las tecnologías deben considerarse, a su vez, los siguientes factores:

- Concientización del usuario en caso de que un dispositivo médico con cierto grado de autonomía genera un error, alarma o fallo;
- Concientización del usuario cuando se requiere la intervención en un acto realizado autónomamente;
- La habilidad del usuario para intervenir o abortar una acción que se está llevando a cabo de manera autónoma; y
- La habilidad del usuario para seleccionar y llevar a cabo acciones de corrección.

Estas recomendaciones tienen por objetivo, principalmente, advertir sobre la necesidad por pensar los futuros riesgos en el comportamiento de los dispositivos médicos de una manera clara, sintética y entendible para los agentes cognitivos promedio. Por su parte, la Canadian Medical Protective Association (CMPA) considera que la responsabilidad de otorgar información para auxiliar en la toma de decisión por parte del paciente es también responsabilidad de los médicos y tratantes de salud.

En este nivel de atención, los códigos de buenas prácticas y leyes pueden quedar en un segundo nivel, por lo que las instituciones médicas deben ser capaces de concientizar al personal sobre la importancia de asegurar el acceso a la información propia de la gestión de riesgos. La CMPA (2022) considera las siguientes directrices para ayudar en este proceso:

- Establecer cómo prefiere el paciente recibir información, tanto en formato como en lenguaje e idioma.
- Considerar los valores culturales y creencias del paciente, así como nivel de alfabetización médica (y tecnológica).
- Preguntar si el paciente requiere el acompañamiento de un tercero para participar en la discusión.
- Evaluar con el paciente los niveles de riesgo que considerarán aceptables.
- Dar respuesta oportuna a las inquietudes y dudas del paciente sobre su salud y tratamiento.
- Presentar al paciente fuentes de información fiables.
- Utilizar otros recursos que puedan ayudar a los pacientes a entender y evaluar las opciones.
- Considerar un abordaje por etapas que aseguren que el paciente está absorbiendo la información antes de tomar una decisión.

Si bien la Constitución Política de los Estados Unidos Mexicanos da cabida a la aplicación de herramientas como el principio precautorio para

proteger los derechos humanos de las personas, en relación con la gestión y gobernanza de riesgos existe menos libertad de acción.

La Ley General de Salud, actualizada al 2022, no reconoce en ninguno de sus artículos la gestión de riesgos en el uso de sistemas tecnológicos computacionales. A lo sumo, en su Artículo 266 hace referencia a los reactivos biológicos administrados en seres humanos, que deben estar etiquetados con indicaciones, precauciones y forma de aplicación. Complementariamente, la Ley establece en su artículo 265 que esta misma exigencia debe ser cubierta cuando se trata de agentes de diagnóstico empleados en dispositivos o equipos médicos, dejando a la deriva los recursos y tecnologías utilizadas para el seguimiento, tratamiento y evaluación de pacientes y sin hacer mención de los recursos tecnológicos computacionales. Asimismo, la NOM-241 tampoco establece derivas claras de acción ante sistemas de IA/AP entendidos como dispositivos médicos y la ley no exige la adecuación empresarial a las normas y estándares internacionales anteriormente propuestos.

Es por ello que la agencia estatal encuentra una frontera de acción frente a la regulación de sistemas de IA/AP en el área de salud en su estado actual. Esto deja ver que a menos que haya una reforma de los marcos normativos oficiales, los esfuerzos regulatorios no pueden depender enteramente de las instituciones públicas. Otra posible vía de acción es la coordinación empresarial; Esta estrategia buscaría promover la responsabilidad social de la industria en el país mediante un esfuerzo por regularizar sus usos tecnológicos, estandarizar las exigencias de acceso a la información por parte de los desarrolladores y elevar los niveles de demanda de protección de los derechos humanos de los pacientes involucrados. Sin embargo, llevar a cabo este ejercicio de forma aislada puede derivar en malas prácticas y omisiones. De ello se sigue la necesidad por crear ecosistemas jurídicos y empresariales que compensen la falta de regulación tecnológica en el país.

Una de las propuestas de acción de responsabilidad social impulsada por parte de las corporaciones (y que no necesariamente son exclusivas

del sector privado) es la creación de certificaciones de calidad que tomen los estándares de las ISO para denotar que las instituciones cumplen con las exigencias de acceso a la información y uso ético de las tecnologías. Esta estrategia permitiría descentralizar la evaluación tecnológica y permitir que agentes externos a las empresas validen con parámetros de objetividad más elevados los usos tecnológicos de las empresas.

Por su parte, debido a su carácter externo a las exigencias de ley, las certificaciones terciarizadas, que van más allá de las buenas prácticas adoptadas por las instituciones, pueden ser diseñadas a la medida de la actualidad sociotécnica del país con el fin de ofrecer lineamientos y parámetros de presentación de la información. De entre ellos, destacan:

- Inclusión de otras lenguas como medio de comunicación de los riesgos en dispositivos médicos que utilicen IA/AP. Esto incluye no solo el uso del español como lenguaje de uso corriente en el país, sino al inglés y lenguas indígenas representativas del conjunto demográfico en cuestión.
- Diseño de sellos de seguridad de fácil interpretación para los pacientes que compensen el analfabetismo técnico y tecnológico, así como las propias capacidades cognitivas de los agentes diana.
- Estimación de riesgo, fracaso y éxito en el uso de las tecnologías y presentación de la información en códigos cromáticos (p. ej. rojo, amarillo y verde o alguna otra combinatoria) para indicar niveles de seguridad, así como sus equivalencias numéricas.
- Indicaciones e instructivos de uso cuando la operación del dispositivo requiera la intervención del usuario final.
- Datos de contacto de los representantes legales en el desarrollo y utilización de las tecnologías para la correcta delegación de responsabilidades.

Esta iniciativa, además, deberá ir acompañada de un trabajo de capacitación en materia de comunicación verbal y escrita para los desarrolladores tecnológicos y para los prestadores de servicios de salud, así como en asuntos éticos, epistémicos y de protección a los derechos humanos.

De lo anterior, podemos concluir que sólo en tanto los esfuerzos colectivos se sumen para mejorar las condiciones técnicas y tecnológicas en el país es que podemos aspirar a crear un uso responsable y ético de los sistemas de IA/AP en el área de la salud.

Como se ha visto a lo largo de este tercer y último capítulo, uno de los problemas a los que se enfrenta la creación de una regulación tecnológica de este tipo en el país radica en que en la actualidad no se cuentan con suficientes herramientas para adaptar los marcos normativos vigentes ni para cubrir todos los usos de sistemas de IA/AP en el área de la salud. Asimismo, el derecho a la información, protegido por instituciones como la CNDH o el INAI, no cuenta con un panorama de acción lo suficientemente amplio para asegurar la transparencia tecnológica.

Pero si bien todas estas estrategias buscan promover la seguridad del paciente y del usuario a través de proporcionarles información de seguridad, esto no debe comprometernos a que el paciente conozca todas las funcionalidades técnicas de los dispositivos ni que tenga acceso total a la información que forma parte de las bases de datos que los alimentan. Esto parecería ser un límite para asegurar la toma de decisión informada por parte del paciente y, en última instancia, pone en duda la viabilidad de asegurar transparencia a los usuarios de las tecnologías inteligentes.

De acuerdo con el abordaje que he realizado en esta investigación resultaría necesario promover el planteamiento de preguntas éticas y sociales dentro de los circuitos de desarrollo técnico, a la vez que se debe llamar la atención sobre los efectos negativos, controversias y aspectos faltos de reflexión para un futuro uso de las tecnologías en el área de la salud. Sólo en la medida en que estas problemáticas se hagan patentes a los públicos usuarios y a sus agentes de relevancia es que se podrían

comprender las dimensiones éticas y sociales de las tecnologías. Siguiendo esta lógica, las instituciones regulatorias deben asegurar la promoción de los espectros benéficos que conlleva el uso de la IA/AP, lo cual permitiría reducir el peligro a los pacientes mediante la disponibilidad de información justa, confiable y contrastada que sea de acceso público.

Es por ello que resulta importante cuestionar los límites, márgenes y fronteras de transparencia a los que los agentes humanos deben someterse para asegurar un uso ético de la IA.

3.3 Limitaciones en el tratamiento del acceso a la información para una democratización digital ética

De acuerdo con Jenna Burrell (2016), es necesario distinguir, al menos, tres tipos de opacidad que se presentan en los sistemas de Inteligencia Artificial y, particularmente, de Aprendizaje Profundo: [1] la opacidad de los algoritmos deliberadamente creada por las corporaciones (o por los mismos desarrolladores) debido a políticas de seguridad, propiedades intelectuales o códigos comerciales; [2] la opacidad de los sistemas debido al analfabetismo tecnológico; y [3] la opacidad que surge de las características de los algoritmos computacionales y de la escala de información requerida para su efectiva utilización (p. 1-2).

A pesar de que a lo largo de esta investigación he defendido que la gobernanza de riesgos es el punto hacia donde una regulación debe dirigir su atención para responder al problema de la opacidad epistémica, lo cierto es que, como afirma Burrell, la opacidad no solo se ve comprometida por cuestiones cognitivas y de alfabetización digital, sino por aspectos humanos, tanto políticos como éticos y económicos, que deliberadamente son creados para evitar el acceso a la información.

Esta creación deliberada de opacidad ha permitido, a su vez, que la creación de riesgos se codifique dentro de la necesidad tecnológica de producción, haciendo no solo que el riesgo sea un efecto secundario latente, como los llamaría Ulrich Beck, sino un inevitable aspecto a asumir

como consecuencia del uso de las tecnologías inteligencias. Esto hace que la IA, por sus mismos principios técnicos, esté comprometida con el riesgo. Y, además, su producción se somete a la condición de creación deliberada de fronteras a la accesibilidad con el fin de impedir la violación de la propiedad privada e intelectual, detonando un nuevo peligro que consiste en que la gente no puede acceder a la información de relevancia para la toma de decisiones sobre su propia vida.

En este sentido, Mike Ananny y Kate Crawford (2016) han sido particularmente críticos en el abordaje de la transparencia como uno de los objetivos primordiales de las políticas públicas en materia de IA/AP. En paralelo a esta investigación, el dúo ha afirmado que la completa transparencia epistémica no solo es imposible de conseguir, sino que en muchos casos puede ser indeseable debido a las afectaciones políticas y sociales que la apertura a la información conlleva, así como por la infertilidad ética en la accesibilidad a la información.

La transparencia ideal a la que Annany y Crawford critican es aquella en la que se pretende no solo hacer completamente claras y entendible, sino controlable, a un sistema. De acuerdo con ellos, esta transparencia se encuentra con algunas limitaciones no solo inevitables, sino que deben ser promovidas por los mismos agentes involucrados, entre los que destacan los legisladores y tomadores de decisión. En su estudio, Annany y Crawford enlistan diez factores de relevancia que los llevan a afirmar que la absoluta transparencia no debe ser parte de las agendas regulatorias en materia de IA/AP. Estos motivos son:

- La transparencia puede estar desconectada del poder, lo que conllevaría que las prácticas de corrupción continúen independientemente de que se asegure la transparencia.
- La transparencia puede ser dañina si no existe un motivo por el cual se revela la información, vulnerando a los individuos.
- La transparencia puede ocluir los procesos de revisión, ya que demasiada información puede incapacitar a los auditores a evaluarla.

- La transparencia puede crear falsos binarismos en torno a la confianza y desconfianza.
- La transparencia puede invocar a un modelo neoliberal de agencia, en el que los individuos presuman su completa libertad y anulen la capacidad del estado para regular a los sistemas.
- La transparencia no necesariamente contribuye a la creación de confianza.
- La transparencia conlleva limitantes profesionales y laborales.
- La transparencia puede privilegiar el ver sobre el entender.
- La transparencia tiene limitaciones técnicas.
- La transparencia tiene limitaciones temporales.

Aunados a estos factores, a continuación, enlistaré algunas otras consideraciones relevantes que una propuesta regulatoria debe evaluar ante la tendencia por asegurar la transparencia en el uso de los sistemas. Es en estos niveles en donde se vuelve necesario establecer pautas legales para la protección del usuario final ante malas prácticas llevadas a cabo por la industria médica, los responsables del área de la salud e incluso el sector productor de dispositivos médicos.

3.3.1 Daños a la privacidad

Como hemos visto, asegurar una transparencia total en el uso de tecnologías de IA/AP no solo implica que los agentes cognitivos humanos sean capaces de rastrear cada operación realizada por el sistema y que conozcan a todas las partes involucradas en el funcionamiento de la tecnología, sino que tengan acceso íntegro a las bases de datos que sirven para entrenar a las arquitecturas.

Por sí mismo, el acceso a datos sensibles compromete a los usuarios en su privacidad ya que su información demográfica, imágenes y diagnósticos podrían adquirir un carácter público (especialmente cuando no han cedido los derechos de esta información con fines científicos). Esto

puede derivar en un mal uso de la información por parte de aseguradoras, empresas financieras o por instituciones de salud, quienes pueden aprovechar los datos para estimar tasas de riesgo poblacionales o estadísticas sobre morbilidad.

Es por ello que no se puede asegurar ni promover una total transparencia en el uso de los sistemas, ya que viola los derechos a la protección de la información y vulnera la confidencialidad médica.

3.3.2 Daños a la propiedad intelectual

Los algoritmos y las arquitecturas de sistemas de IA/AP suelen ser propiedad intelectual de sus desarrolladores y, en algunos casos, de las empresas a las que pertenecen los ingenieros que las diseñan. Al promover políticas de completa transparencia se vulneran los derechos a la propiedad intelectual tanto de particulares como de organizaciones y se permite la replicación de soluciones de software de forma indiscriminada.

Los efectos de este tipo de reproducciones han sido ya ampliamente tratados por otros investigadores, como Deshpande, Pan y Foulds (2020). Por ejemplo, los sesgos demográficos pueden ser una seria limitante para la utilización de algoritmos en poblaciones para las que no han sido diseñadas. En el espectro médico, su uso puede significar falsos negativos, fallas en la interpretación de imágenes o estimación incorrecta del éxito de un tratamiento.

Por otro lado, cada vez está menos diferenciada la separación entre agentes que protegen a los usuarios de los agentes productores de las tecnologías. Contraria a la idea de Ulrich Beck que afirma que los aseguradores no están en el mismo barco que los productores y que los primeros son quienes toman posición junto con las «potenciales víctimas» (2000, 17), hoy en día las fronteras tecnológicas son menos claras. Esto implica que los tomadores de decisión a nivel legislativo y regulatorio se encuentran en tensión con los intereses de los particulares.

Del lado contrario, al proponer una transparencia absoluta en el uso de los sistemas, se apertura al uso clandestino y no monitorizado de las tecnologías, poniendo el riesgo la integridad de los individuos y diluyendo la autoridad epistémica del médico y tratante.

3.3.3 Creación de mayor opacidad epistémica

A pesar de parecer contraintuitivo, la promoción de transparencia en el uso de sistemas de IA/AP puede ser causa de mayores niveles de opacidad epistémica. Esto tiene que ver con la formulación realizada por Humphreys respecto al fenómeno. Para él, así como para los fines de esta investigación, un proceso será epistémicamente opaco si un sujeto X no conoce todos los elementos epistemológicamente relevantes de un proceso.

A lo largo de esta discusión he intentado demostrar que los elementos de relevancia epistémica son contingentes a los sujetos (o pacientes) que interactúan con el sistema. Y, además, he afirmado que el factor de riesgo es el elemento mínimo que las normativas deben exigir para ser de conocimiento del paciente. Sin embargo, al aplicar un principio de transparencia absoluta se apertura la visibilización de la red de agentes involucrados en el desarrollo tecnológico del sistema, así como la naturaleza técnica del mismo.

Esto hace que el panorama de elementos epistémicamente relevantes sea inabarcable para el usuario final y que, por tanto, sea imposible evaluar la pertinencia en el uso de las tecnologías.

3.3.4 Otorgamiento de un exceso de agencia al paciente

Uno de los aspectos más cuestionables de la propuesta que acompaña a esta investigación radica en que el argumento principal que presento parece defender la completa autonomía del paciente para tomar decisiones respecto a su salud. Este argumento parecería apelar al buen juicio de las personas a la hora de evaluar los niveles de riesgo y seguridad

en el uso de las tecnologías, los pronósticos de éxito en su empleo y la pertinencia de su utilización para la atención de su salud.

Este aspecto es controversial ya que contar con información mínima de seguridad en el uso de un dispositivo médico puede llevar a los pacientes a poner en duda la efectividad de su uso, incluso si los índices demuestran un alto nivel de exactitud. Este fenómeno ha sido explorado anteriormente en otras áreas de relevancia médica, tales como la automedicación o el uso de tratamientos alternativos que no han demostrado eficacia comprobable. En materia de IA, podemos presumir que si un usuario está sujeto a narrativas falaces puede verse afectada su capacidad para razonar adecuadamente, alejándolos de un entendimiento real sobre los modelos (Babic, Gerke, Evgeniou y Cohen, 2021).

Ante esta crítica resulta necesario reconocer una posición tecnocrítica respecto a la tecnología médica y a sus prácticas de gestión de la salud. Solo en la medida en que una regulación que contemple la gobernanza de riesgos sea complementada con una estrategia de alfabetización técnica y tecnológica idónea es que se podrá esperar que los usuarios finales tomen decisiones correctas sobre su salud dentro del marco epistémico y moral de una sociedad médicamente tecnologizada. Además, debemos promover el reconocimiento y adopción de los estándares debidos para la toma y codificación de la información que alimenta al sistema o bien su adecuación a los marcos legales internacionales.

A su vez, se debe propiciar la evaluación ciega de las tecnologías de acuerdo con parámetros establecidos en relación con su exactitud, integridad y precisión. Paralelamente, se debe contar con un análisis de la medición de índices de seguridad a los pacientes, de protección de su información y de su comportamiento de acuerdo con los códigos de Derechos Humanos vigentes y de las implicaciones en los ambientes sociotécnicos que implica el uso de las tecnologías de IA/AP.

En este tenor, los firmantes de la Recomendación sobre la ética de la inteligencia artificial de la UNESCO, entre los que se incluye México, afirma que “las personas deberían estar plenamente informadas cuando una

decisión se basa en algoritmos de IA o se toma a partir de ellos” (UNESCO, 2021). Este punto ha sido también ampliamente tratado en la Ley de la Inteligencia Artificial de la Unión Europea (2023), para la que asegurar una mínima transparencia puede ser una vía para asegurar que los usuarios tomen decisiones informadas. Esto implica hacerlos conscientes de que están interactuando con sistemas que funcionan mediante Inteligencia Artificial con el fin de que decidan si quieren seguir usándola o no.

En suma, sólo en tanto no se evalúe exclusivamente el desempeño técnico de los sistemas es que se pueden crear normativas bioéticamente responsables que respondan a las necesidades de las comunidades tecnológicas en las que se implementan. Mientras tanto, el derecho a la información y la autodeterminación de la gestión de la salud individual debe ser privilegiada. O, como afirmó Ivan Illich a Gordian Troeller y Marie-Claude Deffarge (1976):

Nosotros simplemente no podemos dejar en manos de los expertos hacer una evaluación de su propio impacto en la sociedad. Es una tarea de las personas legas. Y es mucho más importante hacer a todos responsables de lo que la industria médica y tecnológica hace con nosotros que dejárselo a sólo algunos especialistas.

CONCLUSIONES

En marzo de 2023, más de mil expertos en materia de desarrollo tecnológico y de Inteligencia Artificial firmaron una carta en la que exhortaban a la detención de todos los grandes experimentos de IA por al menos seis meses. Esta iniciativa tenía como detonante la creación y apertura al público de herramientas de procesamiento de lenguaje, como Chat-GPT, que prometían revolucionar el modo en el que las personas interactuaban con la IA y, naturalmente, nuestra capacidad para entenderla y controlarla.

La carta pública afirmaba que la IA avanzada representaba un profundo cambio en la historia de la vida en la tierra, y que debía ser planeada y administrada con cuidados y recursos medidos (Future of Life, 2023). El llamado buscaba evidenciar que a pesar de la urgencia en crear códigos de buenas prácticas y políticas que nos permitiesen hacer un uso responsable de las tecnologías, al momento no se estaba tomando con la debida seriedad su administración ni gobernanza.

Durante esta pausa, los responsables deberían tomar la oportunidad para establecer los riesgos en el uso de las tecnologías y las medidas necesarias para contrarrestarlos con el fin de que “los sistemas de IA sean desarrollados una vez que se tenga la confianza suficiente en que sus efectos serán positivos y que sus riesgos serán manejables” (2023). Sin embargo, tras los seis meses propuestos, el desarrollo de tecnologías de IA no ha sido suspendido y, a pesar del trabajo de investigación y planeación de políticas públicas, en muchos lugares no ha habido un ejercicio real de creación de marcos regulatorios.

A pesar de esto, llama la atención que hayan sido los mismos desarrolladores quienes hayan alzado la voz respecto a los peligros que

un desmedido o irresponsable uso de las tecnologías puede acarrear en la vida de las personas. Esto arroja la pregunta sobre cómo la gobernanza digital debería enfrentar el problema de la falta de transparencia en su uso y, naturalmente, sobre si los esfuerzos regulatorios de hecho satisfacen las necesidades de protección a los usuarios.

En materia de salud, a través de su plan de acción para la regulación de software de IA/AP como Dispositivo Médico (2021), la FDA ha promovido la idea de que los marcos normativos actuales no satisfacen las necesidades que la implementación de sistemas de aprendizaje automatizado demanda. Por tanto, se ha propuesto la creación de marcos normativos “hechos a la medida” para la implementación de estas tecnologías. En este sentido y contrario a la propuesta de la OMS del 2021 y a los ejercicios legislativos que se han llevado a cabo en el país, podríamos sugerir que la legislación mexicana no debe adaptar sus actuales marcos regulatorios a las nuevas tecnologías, sino diseñar marcos específicos para los nuevos sistemas y medios computacionales. Sólo de este modo se puede promover un análisis y una normatividad que responda integralmente a los espectros éticos y sociales en la implementación de estas tecnologías.

Al 2023, México no cuenta con estos marcos normativos. Adicionalmente, los esfuerzos en materia regulatoria no han producido resultados tangibles para la creación de códigos, normas o leyes que protejan a los usuarios de las tecnologías inteligentes. Y aun menos ha habido un posicionamiento claro en materia de salud.

Por su parte, la regulación en el uso de dispositivos médicos impulsados por motores inteligentes ha mostrado un retraso significativo en el país. Una única normativa hace mención de este concepto: la NOM 241 *Buenas prácticas de fabricación de dispositivos médicos*. Pero, como he discutido, esta normativa solo refiere al marco regulatorio impulsado por la FDA actualizado hasta el año 2017, por lo que no atiende los desarrollos tecnológicos de los últimos años. Esta normativa habría pasado, además, 4 años como propuesta de norma. No sería sino hasta el 2023 que este

proyecto de norma entraría en vigor, a pesar de que desde el 2021 se encontraba en espera de ser publicada en el Diario Oficial de Federación.

Llama la atención que a pesar de que esta Norma se basa en la regulación de la FDA, ignore la invitación por crear marcos normativos hechos a la medida del contexto social de su aplicación. Por el contrario, el código de buenas prácticas recurre a las normas estadounidenses, pero sin el rigor suficiente para tomarlas como guías de acción, sino como referencias conceptuales.

A pesar de este panorama, los últimos años han significado una aceleración en el desarrollo de tecnologías informáticas inteligentes. Esto ha motivado que diversas organizaciones y grupos de investigación hayan puesto el ojo sobre el aspecto ético y epistemológico de la IA, como el seminario de investigación LEEIA de la Universidad Panamericana, así como en la creación de materiales regulatorios, como el Seminario Permanente de Inteligencia Artificial y Derecho del Instituto de Investigaciones Jurídicas de la UNAM.

A lo largo de esta discusión he afirmado que la opacidad epistémica es un límite para el uso seguro y responsable de la IA/AP en el área de salud. Asimismo, he argumentado que existen métodos para reducir los niveles de opacidad en el desarrollo, implementación y uso de estas tecnologías para todos los agentes relevantes que se ven relacionados con ellas, promoviendo una visión de las tecnologías como sistemas que involucran tanto los dispositivos técnicos como la agencia humana misma. Sin embargo, esta perspectiva conlleva un gran peligro: evidencia que alcanzar la absoluta transparencia de los sistemas se vuelve un objetivo imposible.

En realidad, podríamos afirmar que, aunque las políticas y regulaciones públicas impulsaran la reducción en la opacidad de los sistemas, los agentes que interactuamos con ellas siempre nos encontraremos con un límite para saber cómo funcionan y cómo utilizarlas; en este sentido, no habría transparencia porque seguiríamos estando imposibilitados para ver

dentro de los sistemas (Ananny y Crawford, 2018). Pero, entonces, ¿qué podemos hacer con la opacidad epistémica que permanece?

Para superar esta limitante, y de acuerdo con Kate Crawford, nuestra tarea consiste en dejar de ver a la IA como una caja negra que es necesario abrir. En este sentido, no hay un secreto que exponer sobre su operatividad ni su naturaleza (2021, 10). En realidad, lo que debemos hacer es comenzar a entender a la IA como una multiplicidad de sistemas de poder interconectados. Sólo de este modo se posibilitaría abrir un panorama de comprensión que nos habilite para entender las dimensiones sociales, políticas, económicas y culturales de dichas tecnologías, promoviendo un efectivo reconocimiento de la importancia de las herramientas inteligentes y digitales en nuestras vidas.

Esta invitación permitiría, además, superar las limitantes que representa la opacidad producida por el analfabetismo tecnológico. Cuando los usuarios finales de las tecnologías conocen el papel que estas herramientas desempeñan en su vida, cuando son conscientes de las estructuras que están detrás de ellas y cuando saben a quién responsabilizar por su mal uso, podemos hablar de una disminución efectiva de los índices de opacidad tecnológica. Esto es algo que la comisión de Buena Gobernanza e IA de Oxford llama una “transparencia política” (2020).

En consonancia con el trabajo de Ulrich Beck, es importante pensar que a la falta de conocimiento como lo que está detrás del miedo frente al uso de las tecnologías inteligentes, pero, al mismo tiempo, el exceso de conocimiento sectorizado es lo que ha permitido la producción de nuevos riesgos para las sociedades modernas. Si “no hay mejor abono para los riesgos que negarlos” (2000), la invitación de esta investigación es evidenciarlos con el fin de crear nuevas y *responsables* vías de acción frente al uso de la IA para la salud.

Desde un punto de vista técnico, probablemente el anhelo de transparencia deba ser abandonado para considerar derivas más realistas para el tratamiento de la IA en el área de la salud. Al final del día, las

limitantes cognitivas y de distribución y educación tecnológicas son brechas que difícilmente podríamos superar. Sólo en tanto reconozcamos integralmente su complejidad, sus estructuras físicas y sus contextos de uso en relación con el uso de la política para su regulación, podemos entender su influencia en la vida humana y obtener una mejor comprensión del papel que las tecnologías desempeñan en el mundo.

Pero más allá del diagnóstico de un problema, la alternativa que he defendido a lo largo de esta investigación para enfrentar esta necesaria e inevitable opacidad radica en asegurar que proveemos a las personas de la mínima información epistémicamente relevante para tomar decisiones informadas respecto a su salud.

En concreto, la propuesta que he buscado defender a lo largo de la investigación apuesta por la aplicación de principios precautorios que protejan a los individuos ante el uso de tecnologías que puedan poner en riesgo su integridad. Asimismo, he defendido la necesidad de acompañar este principio con una buena gobernanza digital basada en la gestión de riesgos que dé a los pacientes y médicos tratantes información sobre los posibles perjuicios, niveles de éxito, sesgos y datos de desarrollo y fabricación de cada tecnología que incluya a la IA dentro de sus elementos constituyentes para evaluar la incidencia que tendrán en su diagnóstico, tratamiento y seguimiento clínico. Esto implica que la industria tecnológica debe incorporar guías éticas con validez oficial en el desarrollo de sistemas impulsados por IA como parte de sus obligaciones regulatorias.

En tanto las regulaciones estatales se alejen de un paternalismo epistémico y reconozcan que los pacientes deben ser dueños de la toma de decisión respecto a su propia salud, este acceso a la información se vuelve necesario. Esta estrategia sería consistente con prácticas médicas centradas en el paciente y podría incidir en la concientización sobre el papel que la IA está desempeñando en nuestras vidas.

Para terminar, sirvan las palabras enunciadas por Iván Illich:

O trabajamos para encontrar instrumentos temibles y nuevos que hablen de un modo cada vez más opaco e impenetrable para el hombre, o fijamos las condiciones para una nueva era en que la tecnología se utilice para hacer a la sociedad más sencilla y transparente, de manera que todos los hombres puedan volver a tener los conocimientos y utilizar las herramientas que moldean sus vidas. (Troeller y Deffarge, 1976)

POSFACIO

Cuando era pequeño viajaba mucho en autobús. Durante muchos años mi familia y yo solíamos trasladarnos a la central de autobuses de Observatorio, en la Ciudad de México, para tomar un camión que nos llevaría al pueblo de mi padre. Si bien la estadía en el pueblo era generalmente grata, los viajes en autobús eran, por demás, incómodos, ruidosos y molestos. Sin embargo, recuerdo que una chispa de emoción se disparaba en mí cada vez que me subía a uno de esos autobuses con olor a alfombra vieja: al subir, uno no sabía con qué gran obra de la cinematografía (siempre americana) se iba a encontrar siendo proyectada en las diminutas pantallas que se distribuían dentro de la nave. A veces una película romántica interrumpía el sonido de los cláxones que traspasaba las ventanas del autobús; otras veces, cuando se corría con suerte, el conductor proyectaba de un DVD pirata alguna película de la cartelera actual.

Mi padre era un gran entusiasta de esas películas a las que yo solía referirme con el término “películas de camión”. Una práctica de comunión que llevamos durante varios años era contarnos las historias de las películas que veíamos en nuestros trayectos (aunque él a veces me las contaba dos o tres veces ya que olvidaba que la había visto anteriormente). Los títulos que conocí en esos viajes en carretera son muchos e inolvidables, pero hay algunos que destacan por las malas actuaciones, las producciones de bajo presupuesto e, incluso, por el aburrimiento que me causaban.

De entre estas últimas recuerdo una película que comenzaba con un escenario desértico y precario (escenario que sólo después supe era el de las calles de Iztapalapa). Probablemente pensé que sería una historia biográfica de algún personaje famoso que salió de la pobreza, así que, seguramente, decidí prestar atención a los paisajes por la ventana mientras mi padre, al lado de mí, intermitentemente dormía y veía la

película. A los pocos años mi padre fallecería por una leucemia diagnosticada tardíamente y las películas de camión dejaron de disparar esas chispas de emoción en mí.

Unos años más tarde me reencontré por azar con esa película. Recordé ese escenario desértico y algunas escenas que llegué a atisbar en el camino y, sin mucho ánimo y con suficiente ocio, me dispuse a verla. Descubrí que la película llevaba por nombre *Elysium*, que había sido filmada en el año 2013 y que fue dirigida por Neill Blomkamp (mismo director que dirigió *Chappie*, una de las películas de camión favoritas de mi padre). La película se sitúa en el año 2154 y narra la historia de una Tierra sumida en la pobreza que contrasta con la vida de lujos de los habitantes de un satélite artificial construido en forma de aro que gira alrededor de la Tierra.

Antes que la lucha por los medios de producción, por la educación o por los derechos de los trabajadores, los “buenos” de la película tenían por misión sabotear el sistema de Elysium (un satélite artificial en forma de aro que flotaba en la órbita de nuestro planeta) para acceder a unos curiosos dispositivos llamados MedBays. Estos no eran sino lujosas máquinas que permitían a los habitantes de Elysium detectar, diagnosticar y tratar cualquier enfermedad de manera rápida y automatizada. En la película, sólo bastaba recostarse en la camilla del estilizado aparato para que el brazo de un escáner hiciera una rotación completa sobre el cuerpo y notificara el tipo de padecimiento que aquejaba al paciente. Con sólo un toque en la pantalla táctil el escáner daba una vuelta más y el cuerpo del paciente era reconstruido, liberado de la enfermedad o, incluso, rejuvenecido. ¡Qué maravilloso hubiera sido para mi padre contar con una tecnología similar unos años antes!

En la película, los habitantes de la Tierra, debido a la contaminación y a la radicación a la que se ven sometidos en las fábricas que alimentan el satélite artificial, padecen diversas enfermedades que no tienen cobertura por los sistemas de salud en el planeta. Mientras los habitantes de Elysium cuentan con los recursos tecnológicos para ser virtualmente inmortales,

los terrícolas tienen que arriesgarse a viajar clandestinamente al satélite para acceder a los MedBays. Presumiblemente, los MedBays de la película funcionarían con sistemas de Inteligencia Artificial que, combinados con complejos escáneres radiográficos y con sensores ópticos de biopartículas, podrían llevar a cabo las tareas de diagnóstico y tratamiento en segundos.

Aunque estos sistemas parecerían encontrarse a la distancia en el tiempo, actualmente diversas tecnologías son desarrolladas para cumplir con las tareas de esos mágicos dispositivos. Sistemas de tratamiento no invasivo, escáneres personales y sistemas de visión computacional son, al día de hoy, tecnologías en pleno desarrollo.

Si bien puede resultar sorprendente saber que tecnologías tan futuristas forman parte de las agendas actuales de investigadores y desarrolladores, resultaría menos sorprendente a la lectora reconocer que el problema político, económico y social imaginado por Blomkamp es algo que, desde hace muchas décadas, forma parte de las injusticias tecnológicas en el mundo. La precarización de las condiciones de vida de millones de personas, la falta de políticas públicas que protejan universalmente a los habitantes del mundo y la injusticia tecnológica producida por un repartimiento no equitativo de los recursos, instrumentos y dispositivos para el cuidado de la salud son asuntos que no son exclusivos de las ficciones cinematográficas.

Por ello resulta necesario hacernos una serie de urgentes preguntas: ¿qué políticas públicas deberían implementarse para el seguro uso de estos sistemas en el área médica? ¿cómo podemos asegurar una repartición tecnológica universal de estos sistemas? ¿qué riesgos corren los pacientes ante el aparentemente inminente uso de estas tecnologías?

Esta investigación pretende ayudar a la resolución de estas preguntas en aras de promover una implementación eficiente, segura y universal de estos recursos. Pero, sobre todo, pretende contribuir a que estas tecnologías ayuden en el futuro a que padres e hijos sigan viendo películas de camión.

REFERENCIAS BIBLIOGRÁFICAS

- Albert, M. (2021). Posthumanismo, inteligencia artificial y Derecho. *Persona y Derecho* 84, 207-230.
- Annany, M. y Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New media & society*, 20 (3).
- Ammanath, B. (2021, Julio 29). The Future is Intelligent... are you ready? [Conferencia]. *AI Congress LATAM*.
- Ashby, W. R. (2015). *An introduction to cybernetics*.
- Ávila, R. (2018). ¿Soberanía digital o colonialismo digital?. *Sur* 27 15(27), 15-28.
- Babic, B., Gerke, S., Evgeniou, T. y Cohen, G. (2021). Beware Explanations from AI in Healthcare. *Science*, 373, 6552, 284-286.
- Bau, I., Logan, R., Dezii, C., Rosof, B., Fernandez, A., Paasche-Orlow, M. y Wong, W. (2019). Patient-Centered, Integrated Health Care Quality Measures Could Improve Health Literacy, Language Access, and Cultural Competence. *National Academy of Medicine*.
- Beauchamp, T. (2009). The Concept of Paternalism in Biomedical Ethics. *Jahrbuch Für Wissenschaft Und Ethik*, 14 (1), 77-92.
- Beck, U. (1998). *La sociedad del riesgo*. Paidós.
- Beck, U. (2000). Retorno a la teoría de la «sociedad del riesgo». *Boletín de la AGE*, 30.
- Bjerring, J. y Busch, J. (2020). Artificial Intelligence and Patient-Centered Decision-Making. *Philosophy and Technology* 34, 349-371.
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 205395171562251. <https://doi.org/10.1177/2053951715622512>
- Carel, H. y Kidd, I. (2014). Epistemic injustice in healthcare: a philosophical analysis. *Medicine, Health Care and Philosophy*, 17, 529-540.

- Colburn, T., & Shute, G. (2007). Abstraction in Computer Science. *Minds and Machines*, 17(2), 169-184. <https://doi.org/10.1007/s11023-007-90612-7>
- Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Deshpande, K., Pan, S. y Foulds, J. Mitigating Demographic Bias in AI-based Resume Filtering. *UMAP '20*.
- Dragomir, A. (2021, Julio 5). Luke, I'm NOT Your Father: Beyond Technological Paternalism, towards Mutual Cooperation between Patients, Medical Staff and AI. Ponencia en *CEPE/IACAP Joint Conference 2021: The Philosophy and Ethics of Artificial Intelligence*.
- Duncan, W. D. (2017). Ontological distinctions between hardware and software. *Applied Ontology*, 12(1), 5-32. <https://doi.org/10.3233/AO-170175>
- Durán, J. M., & Formanek, N. (2018). Grounds for Trust: Essential Epistemic Opacity and Computational Reliabilism. *Minds and Machines*, 28(4), 645-666. <https://doi.org/10.1007/s11023-018-9481-6>
- Ekmekci, P. y Arda, B. (2020). *Artificial Intelligence and Bioethics*. Springer.
- Farah, L., Murriss, J., Borget, I., Guilloux, A., Martelli, N. y Katsahian, S. (2023). Assessment of Performance, Interpretability, and Explainability in Artificial Intelligence-Based Health Technologies: What Healthcare Stakeholders Need to Know. *Mayo Clinic Proceedings: Digital Health*, 1 (2), pp. 120-138.
- Fierens, M., Rossello, S. y Wauters, E. (2021). Setting the Scene: On AI Ethics and Regulation. En *Artificial Intelligence and the Law*, pp. 51-81.
- Gómez-González, E., Gomez, E., Márquez-Rivas, J., Guerrero-Claro, M., Fernández-Lizaranzu, I., Relimpio-López, Ma. I., Dorado, M., Mayorga-Buiza, Ma. J., Izquierdo-Ayuso, G. y Capitán-Morales, L. (2020). *Artificial intelligence in medicine and healthcare: a review and classification of current and near-future applications and their ethical and social Impact*. arXiv:2001.09778
- Gutiérrez, H. y Limón, C. (2019). Eficacia de los programas de inclusión digital en educación primaria en México. Ponencia en XV Congreso Internacional de Investigación Educativa.

- Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, 169 (3), 615-626. <https://doi.org/10.1007/s11229-008-9435-2>
- Illich, I. (2005). *The Rivers North in the Future*. Anansi.
- Irmak, N. (2012). Software is an abstract artifact. *Grazer Philosophische Studien*, 86(1), 55-72. https://doi.org/10.1163/9789401209182_005
- Kapuscinski, R. (1995). *Lapidarium II*. Varsovia: Czytelnik.
- Kapuscinski, R. (2003). *A Reporter's Self-portrait*. Cracovia: Znak.
- Kapuscinski, R. (2006). *The other*. Cracovia: Znak.
- Kaviani, S., Han, K. y Sohn, I. (2022). Adversarial attacks and defenses on AI in medical imaging informatics: A survey. *Expert Systems with Applications*, 198 (116815).
- Lewis, C. y Pignone, M. (2009). Promoting Informed Decision-Making in a Primary Care Practice by Implementing Decision Aids. *North Carolina Medical Journal*, 70(2): 136-139.
- Marcuse, H. (1941). *Some social implications of modern technology en Studies in Philosophy and Social Science*. Vol. IX. Institute of Social Research.
- Marda, V. (2017). Machine Learning and Transparency: A Scoping Exercise. *SRRN*.
- Mariscal, J. (2005). Digital divide in a developing country. *Telecommunications policy*, 29 (5-6).
- Mariscal, J. (2007). Regulación y competencia en las telecomunicaciones mexicanas. CEPAL.
- Medina, L. (2022). *La aplicación del Principio Precautorio en México*. Tirant Lo Branch.
- Micheli, J. y Valle, J. (2018). La brecha digital y la importancia de las tecnologías de la información y la comunicación en las economías regionales de México. *Revista Internacional de Estadística y Geografía*, 9 (2), 38-53.
- Miralles, F. (2021, septiembre 14). Sector salud y bienestar: Pruebas de concepto de referencia [Conversatorio]. *AI & Big Data Congress*, Barcelona.

- Moor, J. H. (1978). Three Myths of Computer Science. *The British Journal for the Philosophy of Science*, 29(3), 213-222. <https://doi.org/10.1093/bjps/29.3.213>
- Newman, J. (2016). Epistemic Opacity, Confirmation Holism and Technical Debt: Computer Simulation in the Light of Empirical Software Engineering. En F. Gadducci & M. Tavosanis (Eds.), *History and Philosophy of Computing* (Vol. 487, pp. 256-272). Springer International Publishing. https://doi.org/10.1007/978-3-319-47286-7_18
- Oliver, N. (2020). *Inteligencia artificial, naturalmente. Un manual de convivencia entre humanos y máquinas para que la tecnología nos beneficie a todos*. ONTSI.
- Ossa, L., Starke, G., Lorenzini, G., Vogt, J., Shaw, D. y Elger, B. (2022). Re-focusing explainability in medicine. *Digit Health*, 8.
- Primiero, G. (2016). Information in the philosophy of computer science. En L. Floridi & Routledge (Firm) (Eds.), *The Routledge handbook of philosophy of information* (pp. 90-106). Routledge, Taylor & Francis Group.
- Primiero, G., Angius, N. y Turner, R. (2021). The Philosophy of Computer Science. En *The Stanford Encyclopedia of Philosophy*.
- Rojo, P. (2003). Analfabetismo tecnológico en la sociedad de la información. *Revista Latinoamericana de Comunicación CHASQUI*, 081, 48-55.
- Sánchez, M. (2020). La Brecha Digital como causa del Analfabetismo Digital en las Personas más Desfavorecidas. *MATI*. Recuperado de: http://www.webmati.es/index.php?option=com_content&view=article&id=45:la-brecha-digital-como-causa-del-analfabetismo-digital-en-las-personas-mas-desfavorecidas&catid=15&Itemid=164
- Scasserra, S. (2021). La desigualdad automatizada. Industrialización, exclusión y colonialismo digital. *Nueva Sociedad*, 294 (pp. 49-60)
- Singh, A., Sengupta, S., & Lakshminarayanan, V. (2020). Explainable Deep Learning Models in Medical Image Analysis. *Journal of Imaging*, 6(6), 52. <https://doi.org/10.3390/jimaging6060052>
- Steinhoff, J. (2021). *Automation and Autonomy: Labour, Capital and Machines in the Artificial Intelligence Industry*. Palgrave Macmillan Cham.

- Sztompka, P. (1999). *Trust: A sociological theory*. Cambridge University Press.
- Tjoa, E. y Guan, C. (2020). A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE transactions on neural networks and learning systems*, 32 (11).
- Toudert, D. (2014). Bercha digital y perfiles de uso de las TIC en México: Un estudio exploratoriop con microdatos. *Culturales*, 3 (1).
- Travieso, J. y Planella, J. (2008). La alfabetización digital como factor de inclusión social: una mirada crítica. UOC Papers. *Revista sobre la sociedad del conocimiento*, 6.
- Vallor, S., & Bekey, G. (2017). Artificial Intelligence and the Ethics of Self-Learning Robots. En P. Lin & K. Abney (Eds.), *Robot ethics 2.0: From autonomous cars to artificial intelligence* (pp. 338-353). Oxford University Press.
- Young, T, Cole, J. y Denton, D. (2002). Improving Technological Literacy. *Issues in Science and Technology*, 18 (4).
- Yudkin, J. (2022). Advancing patient-centered care: moving from outcome-based to risk factor-based models using the big four risk factors. *Revista Panamericana de Salud Pública*, 46.
- Wiener, N. (2019). *Cybernetics: Or, Control and communication in the animal and the machine* (Second edition, 2019 reissue). The MIT Press.
- Wortman, J., Wallach, H. (2021). A Human-Centered Agenda for Intelligible Machine Learning. En M. Pelillo y T. Scantamburlo (Eds.), *Machines we trust: perspectives on dependable AI* (pp. 123-138). The MIT Press.

DOCUMENTOS

- Administración de Alimentos y Medicamentos de los Estados Unidos (2013). *Software as a Medical Device (SaMD): Key Definitions*. FDA.
- Administración de Alimentos y Medicamentos de los Estados Unidos (2017). *Software as a Medical Device (SAMd): Clinical Evaluation. Guidance for Industry and Food and Drug Administration Staff*. FDA.

Administración de Alimentos y Medicamentos de los Estados Unidos (2020). *FDA Authorizes Marketing of First Cardiac Ultrasound Software That Uses Artificial Intelligence to Guide User*. Actualizado al 07 de febrero de 2020 de: <https://www.fda.gov/news-events/press-announcements/fda-authorizes-marketing-first-cardiac-ultrasound-software-uses-artificial-intelligence-guide-user>

Administración de Alimentos y Medicamentos de los Estados Unidos (2020). *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)* (2021). FDA.

Administración de Alimentos y Medicamentos de los Estados Unidos (2021). *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan* (2021). FDA.

Agencia de Proyectos de Investigación Avanzados de Defensa (2018). *Explainable Artificial Intelligence (XAI)*.

Article 19 (2019). *Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence*. Article 19.

Ashford, N., Barret, K., Bernstein, A., Constanza, R., Costner, P., Cranor, C., deFur, P., Geiser, K., Jordan, A., King, A., Kirschenmann, F., Lester, S., Maret, S., M'Gonigle, M., Montague, P., Peterson, J., O'Brien, M., Ozonoff, D., Raffensperger, C., Regal, P., Resor, P., Robinson, F., Schettler, T., Smith, T., Sperling, K., Steingraber, S., Takvorian, D., Tickner, J., von Moltke, K., Wahlstrom, B. y Warledo, J. (1998). *Wingspread Statement on the Precautary Principle*. Centro de Conferencias de Wingspread, Winsconsin.

Cámara de Diputados del H. Congreso de la Unión. (2019). *Ley General de Educación*. DOF.

Cámara de Diputados del H. Congreso de la Unión. (2022). *Constitución Política de los Estados Unidos Mexicanos*. DOF.

Cámara de Diputados del H. Congreso de la Unión. (2022). *Ley General de Salud*. DOF.

Cámara de Diputados del H. Congreso de la Unión. (2023). *Anexo III. Iniciativas*. Gaceta Parlamentaria. Año XXVI, No. 6246 del jueves 30 de marzo de 2023.

- Canadian Medical Protective Association. (2022). *Helping patients make informed decisions*. CMPA.
- Comisión de Ciencia, Innovación y Tecnología (2023). *Dictamen del Proyecto de Ley 2775/2022-CR, Ley que promueve el uso de la Inteligencia Artificial en favor del desarrollo económico y social del país*. Congreso de la República del Perú.
- Comisión de IA y Buena Gobernanza de Oxford (2020). *Global Attitudes Towards AI, Machine Learning & Automates Decision Making*.
- Comisión Nacional de los Derechos Humanos. (2018). *Comunicado de prensa DGC/258/18: Demanda CNDH alfabetización digital inclusiva para más de 4.7 millones de personas que no saben leer y escribir, con acceso a las TIC's, para que puedan hacer efectivos sus derechos fundamentales*. Dirección General de Comunicación de la CNDH.
- Coordinación de Estrategia Digital Nacional. (2021). *Acuerdo por el que se expide la Estrategia Digital Nacional 2021-2024*. DOF.
- Dirección General de Comunicación Social (2016). *Boletín UNAM-DGCS-607. Se suma al analfabetismo la brecha digital*. UNAM.
- Future of Life (2023). *Pause Giant AI Experiments: An Open Letter*.
- Gobierno de Queensland (2017). *Guide to Informed Decision-making in Health Care*. Queensland Health.
- ISO (2019). *ISO 14971:2019 Dispositivos médicos/productos sanitarios (MD) – Aplicación de la gestión del riesgo a los MD*.
- ISO (2020). *ISO/TR 24971 Medical devices – Guidance on the application of ISO 14971*.
- Martinho-Truswell, E., Miller, H., Asare, I., Petheram, A., Stirling, R., Gómez, C. y Martínez, C. (2018). *Hacia una Estrategia de IA en México: Aprovechando la Revolución de la IA*. Embajada Británica en México, Oxford Insights y C Minds.
- Ministerio de Ciencia, Tecnología, Conocimiento e Innovación (2021). *Política Nacional de Inteligencia Artificial*. Gobierno de Chile.
- Ministerio de Telecomunicaciones y de la Sociedad de la Información (2021). *Proyecto: Diagnóstico sobre la inteligencia artificial en el Ecuador*.

OECD/CAF (2022). *Uso estratégico y responsable de la inteligencia artificial en el sector público de América Latina y el Caribe*. Estudios de la OCDE sobre Gobernanza Pública.

Organización Mundial de la Salud (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*.

Organización Mundial de la Salud (2015). *WHO global strategy on integrated people-center health services 2016-2026*.

Organización Panamericana de la Salud (2023). *Inteligencia Artificial. Caja de herramientas de transformación digital*.

Pacheco, R. (2023). *Dispõe sobre o uso da Inteligência Artificial*. Senado Federal.

Parlamento Europeo (2013). *DECISION No 1386/2013/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 20 November 2013 on a General Union Environment Action Programme to 2020 ‘Living well, within the limits of our planet’*.

Parlamento Europeo (2015). *The precautionary pinciple: Definitions, applications and governance*. EPRS.

Parlamento Europeo (2023). *EU AI Act*.

Secretaría de Educación Pública (2020). *Agenda Digital Educativa*.

Secretaría de Relaciones Exteriores (2022). *Salud*. Actualizado al 07 de febrero de 2022 de: <https://mision.sre.gob.mx/oi/index.php/areas-tematicas/salud>

Secretaria de Salud (2021). *Norma Oficial Mexicana NOM-241-SSA1-2021, Buenas prácticas de fabricación de dispositivos médicos*. DOF.

The Academy for Radiology & Biomedical Imaging Research (2020). *FDA Workshop: “Evolving Role of Artificial Intelligence in Radiological Imaging”*.

UNESCO (2019). *Consenso de Beijing sobre la inteligencia artificial y la educación*.

UNESCO (2021). *Recomendaciones sobre la ética de la inteligencia artificial*.

OTRAS FUENTES

Ritscher, D. (2020). *Explainable AI and Regulation in Medical Devices*. Conferencia durante taller público.

Sucar Succar, L. (2021). *Aplicaciones médicas de la Inteligencia Artificial*. Presentación durante el Foro Aplicaciones de Inteligencia Artificial para proyectos de impacto social

Troeller, G. y Deffarge, C. (1976). *Im Namen des Fortschritts - Kein Respekt vor heiligen Kühen*. 45:29 min.