ISSN 1870-4069

Problemas filosóficos y sociales en la implementación de sistemas opacos de IA en el área de la salud

Diego Vázquez Díaz

Universidad Nacional Autónoma de México
Facultad de Filosofía y Letras,
México

dikvaz@gmail.com

Resumen. Muchas de las guías éticas y regularizaciones en el uso de sistemas de Inteligencia Artificial (IA) han asumido el compromiso de asegurar la transparencia, explicabilidad e inteligibilidad de las tecnologías para sus usuarios. Sin embargo, esta búsqueda tiene como límite la opacidad que conlleva el alto nivel de complejidad de las tecnologías y los índices de analfabetismo tecnológico por parte de los agentes humanos que interactúan con los sistemas. Este problema se vuelve especialmente importante cuando hablamos de las aplicaciones clínicas de la IA si partimos del supuesto de que es necesario promover prácticas médicas centradas en el paciente que aseguren que estos pueden tomar decisiones sobre su salud de manera informada. En el presente trabajo se exploran algunas problemáticas filosóficas y sociales que conlleva la implementación de estas tecnologías en el área biomédica, así como algunas consideraciones éticas y epistemológicas para su resolución.

Palabras clave: Opacidad epistémica, práctica médica centrada en el paciente, transparencia, analfabetismo.

Philosophical and Social Problems of Medical Applications of Opaque AI Systems

Abstract. Many of the ethical guidelines and regulations on the use of AI systems have made a commitment to ensure the transparency, explainability and intelligibility of the technologies for their users. However, this search is limited by the opacity that comes with the high level of complexity of the technologies and the rates of technological illiteracy of those who interact with the systems. This problem becomes especially important when we talk about the clinical applications of AI if we state that it is necessary to promote patient-centered medical practices that ensure that patients can make informed decisions about their health. In the present work, some philosophical and social problems involved in the implementation of these technologies in the biomedical area are explored, as well as some ethical and epistemological considerations for their resolution.

Keywords: Epistemic opacity, patient-centered medical practices, transparency, analphabetism.

1. Introducción

En su guía para la Ética y Gobernanza de la Inteligencia Artificial para la Salud [1], publicada en el mes de junio de 2021, la Organización Mundial de la Salud (OMS) afirmó que la futura implementación de las tecnologías de Inteligencia Artificial (IA) en el área médica deberá considerar, al menos, seis factores de relevancia bioética a saber:

1. la protección de la autonomía,
2. la promoción del bienestar humano y de los intereses públicos,
3. el aseguramiento en la transparencia, explicabilidad e inteligibilidad de estas tecnologías,
4. el afincamiento de responsabilidades y rendición de cuentas,
5. la garantía en la inclusividad e igualdad,
6. la búsqueda de Inteligencias Artificiales sensibles y sostenibles.

En este artículo exploraré de manera específica las consideraciones sociales y filosóficas relevantes al tercer factor de interés bioético enlistado por la OMS para el uso de sistemas de Inteligencia Artificial en el área médica; a saber, el problema de la transparencia, explicabilidad e inteligibilidad de las tecnologías. En la guía, la OMS afirma que el aseguramiento de este factor, así como de los otros enumerados en el listado, dependerá en gran medida de un esfuerzo colectivo para diseñar e implementar políticas bioéticamente defendibles en favor de los intereses de los pacientes y de las comunidades.

Sin embargo, como afirman Bjerring y Busch, una práctica médica centrada en el paciente es incompatible con los usos de tecnologías de Inteligencia Artificial y, específicamente, de Aprendizaje Profundo (que serán referidas con la abreviatura IA/AP en este artículo) en el área de salud [2]. Esto se debe a que, generalmente, los usuarios finales de estas herramientas desconocen la operatividad de los sistemas utilizados. Esto limita la comprensión de la relación que se establece entre tratante y paciente, debido a que, en muchos casos, ninguno de los dos es consciente del funcionamiento de los dispositivos y, por tanto, existe una opacidad en el uso de los sistemas para la procuración de la salud.

En el primer apartado de este trabajo propondré una definición de esta falta de transparencia, a la que me referiré como «opacidad epistémica». 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" [3].

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 [4], en realidad no existe un consenso respecto a las formas de transparencia que debemos buscar. En el primer apartado de este trabajo caracterizaré algunos tipos de opacidad que limitan nuestra comprensión del

funcionamiento de estos sistemas. Posteriormente, argumentaré que estas limitantes imposibilitan el aseguramiento de la transparencia, explicabilidad e inteligibilidad en el uso de estas tecnologías, lo cual contraviene las aspiraciones en la regulación de la IA en el área de salud.

Para finalizar y como respuesta a este problema, presentaré algunas estrategias basadas en la buena gobernanza digital que pueden ayudar a promover la toma de decisión informada por parte de los pacientes a pesar de la existencia de esa opacidad, y presentaré algunas consideraciones finales.

2. Marco filosófico-conceptual

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. Por su parte, podemos hablar de Aprendizaje Profundo (AP) cuando las tecnologías de aprendizaje automático son capaces de procesar información compleja (como imágenes o sonidos) mediante transformaciones múltiples para hacer más efectivo el entrenamiento y aprendizaje del sistema. La particularidad de estos sistemas es que no solamente sirven para el procesamiento de la información de acuerdo con valores ingresados por los agentes humanos, sino que son capaces de aprender y modificar sus propios valores para la obtención de resultados cada vez más precisos y exitosos.

A través del procesamiento iterativo de información, estos sistemas tienen la habilidad para extraer sus propias variables y asignarles pesos específicos para el análisis e interpretación de los datos. Al procesar amplios volúmenes de información en espacios de tiempo cada vez más reducidos, estas tecnologías representan importantes bastiones del aumento en la complejidad de los sistemas computacionales que, en múltiples sentidos, han superado las capacidades de los agentes humanos y que, por tanto, obstaculizan un derecho al conocimiento de las tecnologías que sirven a la gestión de la salud pública e individual.

Debido en parte al analfabetismo digital, a las políticas de privacidad de las empresas tecnológicas y a las capacidades cognitivas de los agentes humanos, actualmente nos encontramos rodeados de tecnologías que al estar condicionadas por *opacos y crípticos principios y mecanismos* [5] impiden que, en muchos casos, sus usuarios las comprendamos a cabalidad. Si bien es común hacer referencia a estas limitantes mediante la denominación de las tecnologías de IA/AP como “cajas negras”, para los fines de este trabajo llamaremos a este fenómeno el *problema de la opacidad epistémica*. 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 [6].

Esta definición de opacidad epistémica sería planteada por Paul Humphreys en su estudio sobre la necesidad por pensar filosóficamente los problemas que conlleva la relación entre humanos y máquinas de computo en la producción de conocimiento. De acuerdo con Jenna Burrell [7], 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 que surge de las características de los algoritmos computacionales y de la escala de información requerida para su efectiva utilización;

(2) la opacidad de los sistemas debido al analfabetismo tecnológico; y (3) 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.

Tomando como punto de partida la clasificación de Burrell, partiré de la hipótesis de que la creación de códigos bioéticos y regulaciones que promuevan la transparencia, explicabilidad e inteligibilidad en el uso de sistemas de Inteligencia Artificial en el área de la salud requiere una exploración de las características específicas de cada uno de estos tipos de opacidad epistémica en relación con los miembros de las comunidades tecnológicas en las que serán utilizadas.

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 para la formulación de principios bioéticos. En última instancia, defenderé que estas consideraciones para el uso de sistemas de IA/AP 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. La opacidad epistémica como producto de las limitantes cognitivas

Los agentes humanos de las comunidades digitales contemporáneas nos encontramos en medio de una contradicción que parece ser infranqueable. Por un lado, en nuestras culturas hemos aceptado la idea de que el derecho al conocimiento es absoluto e ilimitado, mientras que, por el otro, nos enfrentamos con la creciente incapacidad de conocer a cabalidad las tecnologías con las que día a día convivimos.

Si bien esta incapacidad no es un problema exclusivo del siglo XXI, podemos afirmar con suficiente seguridad que la emergencia de las tecnologías digitales y, específicamente, de Inteligencia Artificial y de Aprendizaje Profundo, ha ensanchado la brecha existente entre las capacidades humanas de comprensión del mundo y el estado real de cosas en él.

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 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” [6].

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 y, por tanto, que la creencia en sus resultados termina siendo imposible de justificar.

Durán y Formanek [8], por el contrario, han afirmado que, en realidad, sí contamos con diferentes recursos para generar confianza en los sistemas computacionales que, en consecuencia, podrían justificar nuestras creencias. 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.

Estas cuatro fuentes han sido exploradas como medios para asegurar la transparencia en la utilización de tecnologías de IA/AP. En contraposición a esta idea, en lo consecutivo afirmaré que estas fuentes no contribuyen realmente a superar la limitante que representa la opacidad epistémica en el uso de los sistemas.

Esto dejará en evidencia, además, que, partiendo de la búsqueda por defender el derecho al conocimiento y la información, las estrategias que buscan asegurar la transparencia, explicabilidad e inteligibilidad de las tecnologías no son compatibles con una práctica médica centrada en el paciente.

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 los sistemas.

3.1. Métodos de verificación y validación

El problema de la opacidad epistémica en sistemas complejos, como los de Aprendizaje Profundo, tiene un carácter necesario. Esto es que no importa el nivel de pericia de un agente cognitivo humano X, siempre habrá un grado de opacidad respecto a sus principios y mecanismos. Los agentes humanos están, por tanto, sometidos necesariamente a la indescifrabilidad de estos sistemas. Por ello, 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 recurrió a la creación de una rama de la Inteligencia Artificial específicamente destinada a hacer explicables a los sistemas. De acuerdo con la Agencia de Proyectos de Investigación Avanzados de Defensa (DARPA por sus siglas en inglés) la Inteligencia Artificial Explicable (XAI), tiene por misión "entender, confiar de manera apropiada y administrar de manera efectiva una emergente generación de máquinas acompañantes artificialmente inteligentes".

El XAI pretende crear un conjunto de técnicas de aprendizaje de máquinas que contribuya a crear modelos más explicables manteniendo un alto de nivel de rendimiento. Así, se esperaría que se pudieran desarrollar técnicas que solucionen el

conflicto entre explicabilidad-contra-desempeño de los sistemas. En diversas áreas, como el diagnóstico asistido computarizado, existe una necesidad de que los sistemas sean transparentes, entendibles y explicables para ganar la confianza de los médicos, reguladores e, incluso, de los pacientes.

Idealmente, como afirma Singh [9], un sistema de diagnóstico médico 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 propuestos. Esta validación requiere, 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 que lo determinan.

Para el equipo de Singh, la explicabilidad es un elemento necesario para una utilización segura, ética, justa y confiable de la Inteligencia Artificial en el mundo real; por tanto, el XAI podría fungir como un método para desmitificar el carácter de "cajas negras" de estos sistemas. En el área médica se han implementado diversas estrategias para reducir el nivel de incertidumbre que provoca la operatividad de estos sistemas.

Tal es el caso de los métodos de interpretabilidad sensible, como los *saliency maps* y los métodos de detección de señal, que crean mapas de características detectadas por el sistema; en el caso médico estas herramientas permiten, por ejemplo, visualizar mapas de regiones salientes en imágenes radiológicas que han servido a las tecnologías para identificar tumores o signos de una enfermedad y que justifican su diagnóstico.

Asimismo, otros recursos como la difusión exitosa de significados capa a capa o los métodos contrafactuales permiten identificar la relevancia de un valor de entrada en la obtención de un resultado. 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: el XAI 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 del sistema, probar sus comportamientos o, incluso, encontrar fallas dentro de él. Sin embargo, los sistemas de XAI entran en un círculo de opacidad 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.Por lo anterior, podemos afirmar que los métodos de Inteligencia Artificial Explicable o bien aumentan la opacidad epistémica del sistema en su conjunto o bien permiten solamente disminuir el índice de incertidumbre sobre la operatividad de partes del sistema mismo.

3.2. Análisis de robustez

De mano con los métodos propuestos por el XAI, los análisis de robustez han sido asumidos como medios para verificar el correcto funcionamiento de los sistemas 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. Podemos afirmar que el análisis en la robustez es aquello

que permite aprender sobre los resultados de un modelo dado, para saber si son artefactos 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.

Solo 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. Solo después de este proceso el sistema es puesto a prueba con un pequeño número de imágenes que confirman el nivel de efectividad de este 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. Así, la robustez partiría de una generalización inductiva sobre el funcionamiento del sistema. A mayor cantidad de casos en los que el sistema se comporta de manera correcta en situaciones imprevistas, tendría un carácter más robusto. A pesar de que este método de falsación es común en la evaluación de tecnologías de todo tipo, podría cuestionarse cómo se pueden establecer estos parámetros de éxito estadístico y si los resultados de estas evaluaciones proveen suficiente seguridad en el empleo de los sistemas. Hacia el final de este artículo trataremos algunas posibilidades que permiten aprovechar estas evaluaciones de robustez en relación con la gestión de riesgos.

3.3. Historia de las implementaciones exitosas/no exitosas

Sería un error suponer 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 lo inmediato, podemos comprometernos con que el desarrollo tecnoevolutivo de los sistemas de IA/AP ha estado determinado, en gran medida, por la aspiración de superación de otras tecnologías en competencia.

Por ejemplo, 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 confiar en el sistema y, dentro del marco de nuestra discusión, para asumir que sabemos cómo funciona la tecnología.

El problema fundamental con este recurso es que asume que los individuos conocen y comprenden la historia de las tecnologías. Si embargo, esta información generalmente no forma parte del saber colectivo. A pesar de que esto no es una fuente directa de la opacidad, sí implica una desventaja epistémica para ciertos actores. Esto nos conecta con la cuarta fuente de transparencia que suele asumirse como medio para asegurar el funcionamiento de las tecnologías y, en muchos casos, la fiabilidad en su uso: el conocimiento experto.

3.4. Conocimiento experto

De acuerdo con Durán y Formanek, el conocimiento experto es la cuarta fuente que provee justificaciones suficientes para confiar en un sistema computacional y en muchos casos ha sido explorado como medio para asegurar la transparencia epistémica. Es común pensar que la existencia de expertos en materia computacional, o en los principios que sirven de base a cualquier sistema opaco, demuestra con ciertos conocimientos es posible reducir la opacidad epistémica respecto a un sistema.

Un problema de esta taxonomía 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 aplicar métodos de XAI resulta necesario contar con un alto nivel de conocimiento que permita crear, implementar e interpretar los resultados del análisis. Asimismo, para hacer un estudio de robustez se requiere 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 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 asumen que solo para los expertos en la materia no existe la opacidad epistémica respecto a estas tecnologías. Sin embargo, podríamos afirmar 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 no tiene un nivel de pericia y conocimientos suficiente sobre la tecnología para que pueda tomar una decisión informada respecto a su salud.

En general, los usuarios finales (médicos y pacientes) no tienen, por diversos motivos que trataremos más adelante, este conjunto de saberes. El uso de estas tecnologías con fines médicos implica que, potencialmente, un amplio número de usuarios (radiólogos, médicos, enfermeros, 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 viéramos en la inevitable necesidad de utilizar estos sistemas para la procuración de nuestra salud.

Si siguiéramos la propuesta del dúo de filósofos tendríamos que apelar a que todos deberían tener los conocimientos suficientes sobre los opacos y crípticos principios y mecanismos de los sistemas o bien que debemos renunciar a esta transparencia y apelar a que los expertos evalúan y toman decisiones acertadas respecto al uso de las tecnologías para el manejo de la salud de las personas. Esta suposición, por demás utópica, resultaría, en lo inmediato, imposible de realizar.

4. La opacidad epistémica como producto del analfabetismo tecnológico

A pesar de que hoy en día vivimos rodeados de tecnologías digitales es importante reconocer 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, 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 [10] 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 altos 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. En este sentido, 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 analfabetismo digital. Como afirma Kate Crawford, 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 [11]. 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. Cuestionar el papel de la política, la industria y el desarrollo tecnológico en los marcos económicos actuales resulta por tanto necesario para comprender las dimensiones que se ven afectadas por el uso de sistemas inteligentes.

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 [12]. 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.

Así, el usuario final-paciente que desconoce el funcionamiento del sistema se ve comprometido a asumir una posición de 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 [13] 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. De este modo la vía de solución al conflicto del desconocimiento general de la operatividad de los sistemas radica no solo 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.

El problema radica, en paralelo a la limitante del analfabetismo técnico, 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). Un exhaustivo estudio desarrollado por Gómez-González [14] ha demostrado 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.

Por ello, la reducción de las brechas digitales y la inclusión tecnológica deben ser considerados como factores necesarios para asegurar la disminución de la opacidad epistémica de estas tecnologías y, por tanto, para la toma de decisiones informadas por parte de médicos y pacientes.

5. La opacidad epistémica como producto de la gobernanza digital

En lo inmediato, podemos afirmar que las tecnologías de IA/AP pueden impulsar en gran medida el aumento en el valor de la industria de la salud. Esto se debe a las contribuciones que las tecnologías tienen para aumentar la velocidad de ejecución de las tareas asignadas, así como la reducción de los costos y la complejidad de muchos procesos médicos y administrativos. En el área de la salud, 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) [15].

A pesar de estas grandes promesas que ofrece la IA/AP un problema que limita su uso es que actualmente no existen normativas claras para su regulación. Por ello es necesario llevar a cabo un esfuerzo para establecer criterios de buena gobernanza tecnológica que no solo beneficien a las industrias, sino que proteja a los usuarios finales. En su plan de acción para el uso de software de IA/AP como dispositivo médico, 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.

De acuerdo con el diálogo público sostenido previo a la publicación del 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 [16].

A partir de estas exigencias, la FDA ha expresado su interés por promover un tipo de transparencia 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*).

Paralelamente, un problema que representa el uso de estas tecnologías y la existencia de una amplia multiplicidad de herramientas inteligentes aplicadas en el área de la salud radica en que, en muchas ocasiones, los pacientes y usuarios finales no tienen conocimiento de si los dispositivos cuentan o no con arquitecturas de IA/AP o si han sido utilizadas para el diagnóstico de sus enfermedades, en su tratamiento, o en el seguimiento de sus biométricos.

Por ello, resulta necesario, en primer lugar, visibilizar el uso y disponibilidad de tecnologías de IA/AP en las instituciones de salud y fuera de ellas, regularizando la obligatoriedad en el aviso al usuario de que las herramientas que utiliza cuentan con tecnologías inteligentes. Ante este panorama, una pregunta que debe ocupar las agendas de los desarrolladores, reguladores y de la industria en general es ¿qué información debe estar disponible para los usuarios finales de las tecnologías y cómo debe ser presentada?

6. La gestión de riesgos como alternativa a la transparencia epistémica

A lo largo de esta discusión hemos afirmado que existe un alto grado de opacidad epistémica que implica la imposibilidad por asegurar la transparencia, explicabilidad e inteligibilidad de las tecnologías de IA/AP. Esta conclusión ha partido de una definición de la opacidad epistémica que consiste en afirmar la existencia de elementos epistémicamente relevantes que son desconocidos para los agentes cognitivos. Una crítica que se podría hacer a esta argumentación es que, en realidad, no existe un conjunto de elementos que sean epistémicamente relevantes de forma universal.

Mientras que a un técnico radiólogo puede parecerle necesario conocer cómo opera un sistema antes de utilizarlo para sus diagnósticos, para un paciente puede ser suficiente saber quién lo ha producido para asumir que la tecnología funciona debido a algún sesgo personal. De acuerdo con el grupo de investigación de Microsoft liderado por Vaughan y Wallach, una estrategia centrada en los intereses humanos que promueva la inteligibilidad de las tecnologías debe comenzar definiendo las necesidades de partes interesadas relevantes [17].

En este sentido, la búsqueda de inteligibilidad debe responder a las particularidades de los científicos de datos, desarrolladores, diseñadores, administradores de programas, reguladores, usuarios y de la gente que es afectada por los sistemas en general y no a partir de un supuesto bioético universalizante. Pero ¿cómo asegurar un acceso a la información mínimo para tomar decisiones informadas de acuerdo con las necesidades

de las partes interesadas? Las recomendaciones de la FDA tienen por objetivo advertir sobre la necesidad por pensar los futuros riesgos en el comportamiento de los dispositivos médicos.

Esta es una posible vía para asegurar la toma de decisiones informadas. En inicio, podemos reconocer que un elemento epistémicamente relevante a considerar al someterse a una evaluación, diagnóstico, tratamiento o seguimiento médico en el que se ve involucrado el uso de un sistema de IA/AP es cuál es el índice de éxito que ha demostrado tener el sistema, si el sistema es robusto, cuáles son los riesgos que implica su uso y qué acciones deben ser tomadas en caso de un fallo.

Proveer esta información a los usuarios finales de la tecnología a través de una buena gestión de riesgos puedes ser una respuesta satisfactoria que permita llevar a cabo una toma de decisión informada por parte de los pacientes y médicos a pesar de la existencia de las limitantes cognitivas y sociales que condicionan la existencia de la opacidad epistémica. Así, promover estándares para mantener informado al paciente y al usuario a través de la disponibilidad de manuales y reportes de seguridad funcionaría como un medio para que estos puedan decidir de forma autónoma respecto a su salud.

Una normativa de este tipo no se comprometería con que el paciente tenga que conocer las funcionalidades de los dispositivos ni tener acceso a la información que la alimenta (lo cual promueve, además, el derecho a la privacidad de la información y la protección de la propiedad intelectual). Además, la FDA sugiere que la presentación y acceso a la información sea prescrita por regulaciones locales, lo que permite adaptar la presentación de la información a las necesidades de los públicos que forman parte de cada conjunto sociocultural.

En conclusión, una consideración epistemológica para la creación de regulaciones en el uso de estos sistemas en el área de la salud radicaría en no apostar por la total transparencia operativa de las tecnologías, sino de vías que promuevan la inteligibilidad de los sistemas en casos y para agentes específicos. Como como afirma Marda [4] la transparencia no es necesariamente útil ni posible cuando se tratan sistemas de aprendizaje computacional. Esta búsqueda por la transparencia absoluta ha caído en el equívoco de una aspiración por hacer íntegramente comprensibles a los aspectos técnicos de los sistemas a todo agente humano.

De acuerdo con el planteamiento presentado en este artículo, resultaría necesario tener una mayor precisión en el discurso legislativo y una claridad conceptual que responda a las necesidades efectivas de las comunidades tecnológicas en consideración ya que la aspiración por volver transparentes a los sistemas no es, en muchos casos, sino la búsqueda de hacerlos meramente inteligibles a ciertos públicos.

7. Conclusiones y trabajo a futuro

A lo largo de este trabajo 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 los métodos convencionales 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 son insuficientes.

Esta perspectiva me ha llevado a señalar un gran peligro: alcanzar la absoluta transparencia de los sistemas se vuelve un objetivo imposible. En realidad, podríamos

afirmar que, aunque las consideraciones bioéticas promuevan la reducción en la opacidad de los sistemas, los agentes que interactuamos con ellos siempre nos encontraríamos con un límite para saber cómo funcionan y cómo utilizarlos; en este sentido, no habría transparencia porque seguiríamos estando imposibilitados para ver dentro de los sistemas.

Si las agendas en materia de regulación de la IA/AP continúan comprometiéndose con la transparencia, explicabilidad e inteligibilidad en el uso de estas tecnologías deben también comprometerse con una inclusividad tecnológica que socave los niveles de analfabetismo tecnológico y, como afirma la División de Desarrollo Social de la Comisión Económica para América Latina y el Caribe (CEPAL), alcanzar umbrales de competencia digital para la inclusión social [18].

A pesar de los beneficios sociales que puede tener este enfoque, las condiciones materiales de las sociedades contemporáneas pueden encontrar severas limitantes para cumplir con esta misión. Como alternativa a este planteamiento, podría sugerirse que la toma de decisión informada en materia de IA/AP a través de un conocimiento básico de los riesgos que implica el uso de la tecnología funciona como deriva de solución al problema de la "medicina de caja negra", a pesar de que siempre haya un límite en la transparencia de los sistemas.

En resumen, el anhelo de transparencia debe ser abandonado para considerar derivas más realistas en el tratamiento de la IA/AP en el área de la salud. La IA debe dejar de ser entendida como una caja negra que es necesario abrir: no hay un secreto que exponer sobre su operatividad ni su naturaleza. Un trabajo futuro consistirá en definir cómo se debe presentar esta información a los usuarios finales para hacerla comprensible. Esto requerirá poner especial atención en los niveles de alfabetización poblacional, en las fórmulas de interacción paciente-médico y en las estructuras legislativas del contexto en estudio.

Naturalmente esta tarea se vuelve de especial importancia en contextos tecnológicamente menos desarrollados y que no cuentan con iniciativas claras de regulación tecnológica, como es el caso de México y otros países latinoamericanos. Solo en tanto reconozcamos la complejidad de estas tecnologías es que podremos comenzar a entender la influencia que tienen en la vida humana y obtener una mejor comprensión de su papel en el mundo.

Referencias

1. Organización Mundial de la Salud: Ethics and governance of artificial intelligence for health: WHO guidance (2021)
2. Bjerring, J., Busch, J.: Artificial Intelligence and patient-centered decision-making. Philosophy and Technology, vol. 34, pp. 349–371 (2020)
3. Article 19: Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence. Article 19 (2019)
4. Marda, V.: Machine Learning and Transparency: A Scoping Exercise. SRRN (2017)
5. Sztompka, P.: Trust: A sociological theory. Cambridge University Press (1999)
6. Humphreys, P.: The philosophical novelty of computer simulation methods. Synthese, vol. 169, no. 3, pp. 615–626 (2009) doi: 10.1007/s11229-008-9435-2.
7. Burrell, J.: How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, vol. 3 (2016)

8. Durán, J. M., Formanek, N.: Grounds for Trust: Essential Epistemic Opacity and Computational Reliabilism. Minds and Machines, vol. 28, pp. 645–666 (2018) doi: 10.1007/s11023-018-9481-6.
9. Singh, A., Sengupta, S., Lakshminarayanan, V.: Explainable Deep Learning Models in Medical Image Analysis. Journal of Imaging, vol. 6, pp. 52 (2020) doi: 10.3390/jimaging6060052.
10. Young, T, Cole, J., Denton, D.: Improving Technological Literacy. Issues in Science and Technology, vol. 18, no. 4, 73–79 (2002)
11. Crawford, K.: Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press (2021)
12. Dragomir, A.: Luke, I´m NOT Your Father: Beyond Technological Paternalism, towards Mutual Cooperation between Patients, Medical Staff and AI. In: CEPE/IACAP Joint Conference 2021: The Philosophy and Ethics of Artificial Intelligence (2021)
13. Carel, H., Kidd, I.: Epistemic injustice in healthcare: a philosophical analysis. Medicine. Health Care and Philosophy, vol. 17, pp. 529–540 (2014) doi: 10.1007/s11019-014-9560-2.
14. 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., Capitán-Morales, L.: Artificial intelligence in medicine and healthcare: a review and classification of current and near-future applications and their ethical and social Impact. (2020) doi: 10.48550/arXiv.2001.09778.
15. Miralles, F.: Sector salud y bienestar: Pruebas de concepto de referencia. AI & Big Data Congress (2021)
16. FDA: Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan (2021)
17. Vaughan, J., Wallach, H.: A Human-Centered Agenda for Intelligible Machine Learning. Microsoft (2022)
18. Martínez, R., Trucco, D., Palma, A.: El analfabetismo funcional en América Latina y el Caribe: Panorama y principales desafíos de política. CEPAL (2014)