

Gaceta



CONBIOÉTICA

Publicación trimestral de la Comisión Nacional de Bioética
Año XII Núm. 48 Junio de 2023

Bioética e inteligencia artificial

Inteligencia artificial
¿un sustituto del amor?

Visión general de la inteligencia
artificial en medicina

Reconocimiento facial:
una herramienta
de alto riesgo



Encuentro bioético

Inteligencia Artificial y ética. Conversación con Gabriela Ramos,
Subdirectora General de Ciencias Sociales y Humanas de la UNESCO



SALUD
SECRETARÍA DE SALUD



Integridad
Científica



gob.mx/salud/conbioetica

La CONBIOÉTICA te invita a su biblioteca presencial,
la cual cuenta con libros e información especializada
en Bioética y áreas afines



Ubicación: Calzada Arenal No. 134 P.B., Col. Arenal Tepepan,
Alcaldía Tlalpan, 14610 CDMX, México
Horario: Lunes a viernes, 9:00 a 14:00 y de 15:00 a 18:00 horas



SALUD
SECRETARÍA DE SALUD



Integridad
Científica



gob.mx/salud/conbioetica

- 2 **Editorial**
Bioética e inteligencia artificial
- 3 **A fondo**
 - *Inteligencia artificial, ¿un sustituto del amor?*
Miguel Ángel Torres Quiroga
- 8 • *Visión general de la inteligencia artificial en medicina*
José Luis García Vígil
- 14 • *Humanos y máquinas inteligentes: la responsabilidad en el uso de la inteligencia artificial en la atención médica*
Elisa C. Calleja Sordo
- 19 • *Reconocimiento facial: una herramienta de alto riesgo*
Juan Espíndola Mata
- 24 • *Implicaciones éticas de la implementación de la inteligencia artificial en la práctica médica*
Luis Muñoz Fernández
- 30 • *Implicaciones bioéticas en la IA*
Luis Ángel Lara Pereda
- 35 • *Integridad académica y recursos digitales: acotaciones para una reflexión de vista a los recursos derivados de la inteligencia artificial*
Geovany Meza Chavero
- 39 **Escenario nacional**
 - *Desafíos para la regulación en torno al desarrollo y aplicación de la IA*
Garbiñe Saruwatari Zavala
- 45 **Escenario internacional**
 - *Los desafíos éticos y científicos de ChatGPT en salud: utopismo, tecnofobia y pragmatismo*
Luis García Valiña e Ignacio Mastroleo
- 50 **Tareas y perspectivas institucionales**
 - *Resultados de las encuestas sobre la percepción de sesiones conjuntas entre el CEI, CI, CB o CICUAL aplicadas a los Comités de Ética en Investigación y a las Comisiones Estatales de Bioética*
Areli Cerón Sánchez, Gabriela Pineda Hernández, Karla Alejandra Tovar López, Karla Gabriela Sánchez Villanueva y Flor de María Cruz Estrada
- 57 **Encuentro bioético**
 - *Inteligencia Artificial y ética. Conversación con Gabriela Ramos, Subdirectora General de Ciencias Sociales y Humanas de la UNESCO*
- 62 **Cultura y bioética**
Sugerencias en medios digitales de la CONBIOÉTICA
Ariana Leticia Landín López
- 65 **Rincón bibliográfico**
Sugerencias editoriales de la CONBIOÉTICA
Karla Nallely Rosas Chelius

La inteligencia artificial (IA) ha sido descrita como la “cuarta revolución industrial” y ha ido desarrollándose vertiginosamente en múltiples esferas de las actividades humanas: la industria, la ciencia, la tecnología, la salud y la educación, por mencionar algunas.

Tecnologías como el aprendizaje automático y el procesamiento del lenguaje natural son parte del panorama y aplicación de la IA, evolucionando a lo largo de su propio camino y, cuando se aplican en combinación con el análisis de datos y la automatización, pueden ayudar a lograr objetivos en los diferentes ámbitos en que se utilizan, creando un cambio de paradigma en nuestros estilos de vida.

Por otro lado, la IA ha tenido implicaciones en el cuidado de la salud, en las esferas individual y colectiva, toda vez que el horizonte de posibilidades que brindan estas tecnologías aumenta constantemente. Actualmente, las funciones más comunes de la IA en entornos médicos, son el apoyo a las decisiones clínicas y el análisis de imágenes orientadas a mejores diagnósticos. En este sentido, las herramientas que existen auxilian a los médicos en la toma de decisiones sobre tratamientos, fármacos, análisis de datos y otros requerimientos de los pacientes, brindando un acceso rápido a la información. En el mismo orden de ideas, la IA se utiliza para analizar tomografías computarizadas, rayos

X, resonancias magnéticas y otras imágenes, en busca de hallazgos que podrían pasarse por alto, es así que la IA se convierte rápidamente en parte integral de la atención médica moderna.

Esta edición número 48 de la Gaceta CONBIOÉTICA, que versa sobre el tema “Inteligencia artificial y bioética”, parte de la consideración de las implicaciones bioéticas que derivan del apresurado desarrollo tecnológico, las preocupaciones acerca de sus aplicaciones y los posibles sesgos sobre la protección de datos personales y la regulación de dichas tecnologías, entre otros aspectos.

En virtud de lo anterior, es importante recordar lo propuesto por el filósofo alemán Hans Jonas acerca del compromiso ético que pide prudencia y equilibrio, preservando el interés de la existencia de la humanidad y tomando conciencia de todos los eventuales daños causados por nuestras acciones.

En espera de que el contenido de esta publicación sea de su interés y estimule el intercambio de reflexiones sobre las promesas que ofrece la IA, sus posibles riesgos, así como los límites éticos, sociales y legales que implica su aplicación, particularmente en actividades que tienen un impacto directo en nuestra vida cotidiana, lo invitamos a continuar con su lectura.

Inteligencia artificial, ¿un sustituto del amor?

Miguel Ángel Torres Quiroga*

En este ensayo de divulgación, planteo la conexión entre la Inteligencia Artificial (IA) y el amor como experiencia social significativa, desde las amistades y relaciones sexo-afectivas que la IA está modificando y modificará aún más, hasta llegar al transhumanismo. El cine imagina visiones estéticas sobre dilemas éticos de nuestro tiempo, y retoma conflictos sociales que transforma en textos audiovisuales complejos, narrativas sobre el amor, la amistad, el erotismo, la intimidad, la ruptura. Un probable objetivo es comunicar con audiencias actuales y futuras las inquietudes emocionales, malestares y experiencias de muchísimas personas en una época determinada. *Her* (Spike Jonze, 2013), *I.A. Inteligencia artificial* (Spielberg, 2001) y un episodio de la serie británica *Black mirror* titulado *San Junípero* (Owen Harris, 2016) son obras sobresalientes que analizan cuestiones de la identidad y naturaleza humana en relación a las tecnologías con IA. ¿Qué respeto nos merece una IA con la que podríamos mantener una relación íntima como con cualquier otra persona? ¿Es positiva o negativa la incorporación de dispositivos de IA para enriquecer experiencias afectivas, o suplirlas cuando desaparecen o no existen? ¿Una experiencia mental sobre el amor y otros sentimientos profundos producida artificialmente por dispositivos tecnológicos de realidad virtual es tan real como una relación mantenida con otra persona? Escapa de mi objetivo dar una respuesta justificada a esas preguntas. Mi propósito es suscitar el interés en la audiencia sobre las posibilidades de algunos dispositivos de IA que procuran estimular nuestras emociones amorosas, especialmente cuando se suscitan problemáticas dolorosas, como pérdidas, duelos, soledad, decepción amorosa, etcétera.

Proferir emociones profundas, empáticas, y positivas hacia otras personas y animales no humanos es una evidencia de vida inteligente, de compleja actividad e interacción mente-cerebro. Seres humanos, y algunos animales como antropoides, cetáceos y elefantes, demuestran que el amor y los afectos mantienen una gran relación con la in-

teligencia y la solución de problemas. No obstante, ¿es congruente incluir robots y sistemas operativos como agentes suficientemente inteligentes para amar? El cine ha dedicado gran interés por los tópicos relacionados con las formas de vida artificial, desde su nacimiento: *Viaje a la luna* de Georges Méliès (1902); *Fausto* (1924), de Murnau; *El golem* (1920), de Wagener y Boese; *El doctor Frankenstein* (1931) y *La novia de Frankenstein* (1935) de James Whale. Crear vida; Vencer la muerte; alcanzar capacidades sobrehumanas; y alterar el poder de la suerte y los accidentes en la felicidad humana, han sido los cometidos más codiciados de científicos y tecnólogos identificados con movimientos trans y posthumanos. En *Blade runner* (Ridley Scott, 1982) se expone la transición de la tecnología como transformación y progreso de la vida cotidiana, a la transformación radical de nuestra naturaleza de especie; la vieja naturaleza se reemplaza por otra que quiere alejarse de cualquier funcionamiento defectuoso, de enfermedades y padecimientos que ocasionan sufrimientos.

Paradójicamente, la nueva naturaleza humana organiza una reformulación de las viejas clases sociales, con características indeseables, distópicas, o malvadas, que la hacen aún más perversa que las antiguas que parecían involuntarias.

Las películas *Inteligencia artificial*, *Her* y el episodio de *Black Mirror* *San Junípero* ejemplifican estupendamente el rompimiento con la realidad amorosa entendida exclusivamente entre humanos. En la cinta de Spielberg, una humanidad agónica sobrevive en un planeta vencido por el calentamiento global que ha convertido a Nueva York en una ciudad perdida, que a la vez funciona para un pequeño niño robot meca llamado David, como su destino final. David fue diseñado por un científico que perdió a su único hijo; ahora, el único propósito del pequeño robot es amar sin límites a quien

*Doctor en Filosofía, Universidad Autónoma de Madrid; Investigador posdoctoral, Universidad Autónoma Metropolitana, Unidad Iztapalapa. Autor de libros y artículos.

lo programe para recibir ese afecto. Nada puede revertir el programa una vez iniciado. David no es un ser vivo: no crece, no se reproduce (siempre será un robot-niño) y nunca morirá, pero demuestra la intensidad con la que los seres vivos más complejos se aferran a un cierto tipo de vínculo significativo moralmente hablando, en el que no basta con cumplir una función básica de supervivencia. David sólo anhela el amor de “su madre” Mónica, la mujer que lo programó, y que encuentra en él un insuficiente placebo por el sufrimiento de perder a su hijo, congelado en un laboratorio de criogenia. Cuando éste despierta, entendemos que el amor de David no es humano: no es capaz de infringir daño voluntariamente a todo aquél que amenace la unidad con su madre, aunque las actitudes del hermano resucitado son demasiado humanas, lo cual incluye cierto placer por destruir. Tras varios eventos desafortunados, David es abandonado por la única persona que le da sentido a su consciencia de vida. Su odisea reconecta con muchas travesías clásicas: debe alcanzar la arcadia donde vive quien promete la felicidad eterna, el hada azul que convirtió a Pinocho en un niño de verdad, y que hará lo propio con él.

En un futuro más próximo, el film de Jonze comienza con un plano ya conocido en *El apartamento* (Wilder, 1960) y previamente en *Y el mundo sigue* (Vidor, 1928), con la enajenación y la soledad de la vida urbana. Es una empresa que escribe cartas de amor a quienes lo solicitan, y Theodor es uno de sus empleados más especiales por su vena poética. En la realidad, vive deprimido por su separación con la mujer que pensó lo conocía mejor que nadie. Un día, un sistema operativo avanzado con “entidad femenina” llamado Samantha es capaz de introducirle en una espiral emocional que le devuelve el sentido de la vida. Aunque Samantha no posea cuerpo, pasado, familia, ni contacto material y efectivo con la vida real, sus palabras e intervenciones conmueven profundamente Theodor: sonrisas, orgasmos, alegría, placer, satisfacción personal, júbilo. Como nada dura para siempre (desconocemos cuánto

puedan durar las inteligencias artificiales), Theodor cae en cuenta que su relación con Samantha no es un milagro, sino un comportamiento para el cual el sistema operativo fue diseñado y que comparte con muchísimas personas aparte de él. La nueva vida de Theodor se desploma.

Finalmente, en una nostálgica discoteca llamada San Junípero, Kelly y Yorkie son dos jóvenes lesbianas que coinciden en un lejano 1987 varios fines de semana. Ambas se enamoran, hacen el amor, tienen riñas y se confiesan la una a la otra sus sentimientos muy pronto. Lamentablemente, un accidente trunca la vida de Yorkie, dejándola en estado vegetativo antes de concretar algo con Kelly, en parte por culpa de prejuicios sociales familiares. Sin embargo, algo extraño sucede con el tiempo de estos encuentros: lo que hemos visto no son hechos sucedidos linealmente, sino la elaboración mental de Yorkie de su último día de vida antes de sufrir el terrible accidente. Ahora, Yorkie es soportada por vida artificial, sin esperanza de recuperación, convertida en una mujer mayor que espera la muerte. Gracias a unos dispositivos, habita mentalmente un mundo en el que los moribundos y los viejos rememoran, encarnando cuerpos jóvenes, sus mejores momentos. En el hospital, Yorkie es visitada por Kelly muchos años después, a quien le produce una sensación de repulsión y tristeza la tragedia de su antigua amante. Un día, el personal médico aconseja un experimento: que el dispositivo de IA les sea reconectado a las dos, recreando en la vida mental de ambas un posible reencuentro. Para Kelly supone una renuncia, pues sí ha podido hacer una vida heterosexual, aunque no deseada en el fondo, pero mucho más compensadora que la de la pobre Yorkie.

La IA cobra formas innumerables, realizando tareas también difíciles de limitar, pero, ¿nos ofrecería amor? ¿es capaz de devolvernos aquello que la vida nos arrebató? ¿cura acaso la infelicidad y la tragedia o sólo las disfraza? Supongamos que una respuesta de todo lo anterior es permanecer abiertos a la po-

sibilidad de esta nueva ruta en la evolución social, en la transformación afectiva de la naturaleza humana. Invariablemente se ampliará muchos umbrales sobre los sujetos de afecto, más allá de animales y plantas. Que lo importante no es relacionarnos exclusivamente entre humanos mortales, y que es muy relevante lo que podemos descubrir sobre la complejidad de los afectos, de las emociones más significativas, que forzosamente son relacionales, en la recreación de consciencias con inteligencia suficiente para mantener un relato de su vida, una misión que cumplir, una compañía que ofrecernos, un futuro que modificar favorablemente.

Las películas citadas ofrecen una excelente demostración de un porvenir tecnológico aparentemente idílico. Como contraparte, presentan posibles argumentos en contra del avance ilimitado en el diseño de una IA que haga realidad el amor. El primer argumento, que una IA que provea amor incondicional con todos los matices existentes es incompatible con la libre voluntad, que permita establecer los alcances y los límites de la relación amorosa. El amor es muy significativo justamente en buena medida porque, algún día, se extinguirá o se transformará en un sentimiento distinto. El amor es perecedero como la vida misma, y por eso su presencia enriquece e ilumina un camino que, invariablemente, finaliza con la muerte. Una IA como el pequeño David parece contradecir lo anterior. El segundo argumento, dado que la IA carece de vida, tampoco puede superar el relato de los alcances de invención humana, que busca satisfacer propósitos específicos. Contrariamente, las vidas humanas poseen innumerables propósitos, tantos como las circunstancias hagan posibles. Vivimos varias vidas desde que nacemos hasta que morimos, y los fines de la vida son múltiples, pero, ¿podría la IA reproducir algo similar? Y el tercer argumento, que los humanos encontraremos serias dificultades para amar una IA sin sufrir alteraciones severas en nuestras capacidades emocionales. Parece improbable que haya compatibilidad entre un amor humano (imperfecto, frá-

gil, egoísta en cierta medida) con un amor absoluto e incondicional con paciencia infinita que solamente vive para recibir amor del objeto indicado (otra vez, el caso de Inteligencia artificial). Sería inconcebible para muchas personas amar un sistema operativo con IA que desafía los límites humanos del espacio y el tiempo, que se expresa en con diferentes consciencias y en múltiples dimensiones que la consciencia humana no es capaz de entender (como sucede en Her). Por último, encontraremos sumamente difícil experimentar un amor sin cuerpo, sin las características materiales de los seres vivos: nacen, crecen, duermen, comen, se sexualizan, se reproducen, envejecen, y finalmente mueren.

A pesar de las advertencias, las películas también revelan cuestiones sugerentes sobre nuestra evolución socio-afectiva, social y relacional. Frecuentemente, el cine presenta el lado distópico, inhumano y violento de la alteración de la naturaleza humana por parte de la tecnología. El mensaje es que está mal que la tecnología desplace las relaciones “verdaderas” y libres entre personas con agencia y autonomía. De suceder esto, se perdería la capacidad de reconocer y sobrellevar el sufrimiento; aceptar la finitud de la vida; contemplar la fuerza de las relaciones amorosas. Pero también debemos recordar otro aspecto, y es el de la política que dirige las tecnologías, particularmente la IA. Es problemático que éstas provengan de una élite opresora, tiránica, con aspiraciones totalitarias e inhumanas, que desea moldear la naturaleza según sus intereses ideológicos, económicos y supremacistas.

El día en que la IA pueda ayudarnos a superar momentos difíciles y hasta trágicos puede estar muy cerca. El cine, con todo y su observación distópica, también introduce que estos cambios son inevitables, en buena medida, porque la evolución humana no es sólo biológica, sino que está íntimamente relacionada con los valores y los códigos éticos empleados hacia las formas de vida no humanas y, ahora, hacia otras formas de

consciencia. En este sentido, el cine invita a observar que la IA posee sus contextos socio-políticos, los cuales influyen poderosamente en sus funciones y objetivos sociales. Así, se pueden plantear miradas más benévolas sobre el papel de la IA en aquellas cuestiones sumamente relevantes de la complicada vida afectiva contemporánea.

"Proferir emociones profundas, empáticas, y positivas hacia otras personas y animales no humanos es una evidencia de vida inteligente, de compleja actividad e interacción mente-cerebro".

Recordando Blade Runner, los replicantes poseen memoria, consciencia, emociones; son reservorios de la memoria humana que los propios humanos no son suficientemente capaces de recabar y preservar, y por eso mismo constituyen un peligro para ciertas élites. Los androides quieren o quisieron pertenecer a una sociedad que los rechaza y promueve su destrucción. De este modo, una función de la IA consiste también en conducir las emociones, como el amor, en contextos más extensos, derivándose así una suerte de búsqueda de justicia. En Inteligencia artificial, Spielberg revive las persecuciones y atrocidades perpetradas en su película sobre el holocausto 'La lista de Schindler', cuando los mecas son sistemáticamente destruidos en nombre de una defensa de la naturaleza humana en una fiesta llamada 'de la carne', motivada por un discurso sobre la pureza y la identidad que se compagina con la identidad de movimientos

sociales reaccionarios. Cientos de mecas destruidos son depositados por enormes grúas en bosques abandonados, como en su tiempo los nazis enterraron los cadáveres de miles de judíos en terrenos baldíos o en bosques. David no solamente es un pequeño robot único en su género -a pesar de haber sido construido en serie para ofrecer alivio a muchas personas devastadas por la pérdida y la soledad. También es un reservorio de memoria de las cosas grandiosas, buenas, grises y malvadas de la humanidad, testigo ocular de sus legados y miserias.

En *Her* encontramos otra aproximación al colapso civilizatorio en el que la IA parece ofrecer una luz ante la inconexión social y el embotamiento emocional, como epidemias normalizadas. Samantha conoce a profundidad a Theodor, y sabe del incómodo mal de su tiempo. Él padece la soledad urbana de las sociedades capitalistas avanzadas, donde incluso las cartas de amor más íntimas, poéticas y singulares son una mercancía, y no una obra personal. No queda tiempo que perder para pequeñeces en un entorno tan dirigido a la producción. Se vive en una sofisticada trampa que dificulta el despertar social sobre la miseria compartida y la incapacidad de reconectar con la carencia como constitución de nuestra naturaleza, de origen y nacimiento. Samantha descubre que el amor de Theodor no lo liberará de la alienación y el eterno sufrimiento, y opta por salir de su vida como último sacrificio amoroso. Así, abandona a su amante en el naufragio del desamor, pero esta vez no se encuentra completamente solo: otra amiga que trabaja en la compañía pasa por una situación similar, sin duda por una relación con otro sistema operativo que, como Samantha, tuvo que marcharse antes de provocar problemas más grandes.

Convivir con diferentes mecanismos de IA no tendría por qué significar solamente distopía y apocalipsis, o un gobierno de robots y máquinas vengativas. La imaginación debe servirnos para cimentar el camino que nos falta recorrer como especie, y aquí la IA puede ser

un aliado más que un enemigo. Las películas reseñadas ofrecen un horizonte imaginativo que fomenta la humanidad y el rescate de las relaciones más importantes. Igualmente, cabe admitir que muchos males atribuidos a la convivencia con máquinas son, realmente, demasiado humanos, y promovidos por humanos también. Los duelos nos arrebatan la respiración sin matarnos; probablemente, las nuevas inteligencias, máquinas, robots y demás, pueden acompañarnos en el tránsito de este valle de lágrimas.

David espera 2000 años bajo las ruinas de la vieja Nueva York hasta que el hada azul cumple su deseo de ser un niño real. Entonces, una inteligencia extraterrestre lo encuentra, convirtiéndose así en el único testigo de la desaparecida civilización humana y de sus aportaciones, logros y contradicciones. Como en *Blade Runner*, memorias extraordinarias sobre eventos irrepetibles se han perdido inevitablemente, como lágrimas en la lluvia, cuando los androides/replicantes desaparecieron tras la extinción humana. Pero en David se conserva a uno que sobrevivió varios holocaustos y catástrofes. En otro escenario, Theodor se reencuentra con el desconsuelo y la tristeza de las que quiso huir ahora con la despedida de Samantha. Finalmente, encontramos por lo menos un desenlace feliz en San Junípero, aunque sea fabricado artificialmente. Yorkie y Kelly pueden concretar su amor a pesar de los obstáculos en un universo inexistente fuera de las recreaciones mentales complejas pro-

ducidas por los dispositivos conectados a sus cerebros. Gracias a ella, llegan a amarse fuera del bucle del tiempo en el que el mismo día de 1987 se repetía una y otra vez: la IA alcanza una complejidad tal que resucita a los muertos, les regala vidas dignas de vivirse, y que las dos perciben como si fueran reales. Es una versión amable y reconfortante de la sórdida técnica empleada en *Soylent green* (conocida en México como “Cuando el destino nos alcance”) para ejecutar a los viejos y enfermos, para después convertirlos en el alimento de una humanidad moral y ecológicamente devastada. Claro, no sin antes regalarles un último viaje sensorial por un hermoso mundo virtual inexistente.

Visto lo anterior, las lecciones por descubrir sobre la IA y el amor son múltiples: ¿es mejor vivir en la miseria social antes que recurrir a las tecnologías que impulsen la conexión emocional con las necesidades afectivas? ¿Deberían despreciarse o minusvalorarse las relaciones con dispositivos de IA cuando las relaciones con personas atraviesan momentos complicados? Y, si la respuesta es afirmativa, ¿por qué? ¿Acaso la sexualidad puede ser igualmente significativa cuando un humano la practica con dispositivos de IA por motivos diversos (discapacidades, dificultades personales serias, encarcelamiento o privación de libertades)? ¿Pueden la IA y otros robots convertirse en embajadores de las civilizaciones humanas y heredar las memorias de nuestra cultura, o quizás incluso conformar una cultura con identidad propia?

Obras consultadas

- Harris, Owen (2016). *Black Mirror: San Junípero*. Netflix-House of Tomorrow.
- Jonze, Spike (2013). *Her*. Annapurna Pictures, Stage 6 films.
- Méliès, Georges (1902). *Le Voyage dans la Lune*. Star-film.
- Murnau, F.W. (1926). *Fausto*. UFA.
- Scott, Ridley (1982). *Blade runner*. Warner Bros.
- Spielberg, Steven (1993). *Schindler's list*. Universal Pictures.
- Spielberg, Steven (2001). *A.I. Artificial Intelligence*. Dreamworks SKG, Warner Bros.
- Vidor, King (1928). *The Crowd*. MGM.
- Wagener, Paul y Boese, Carl (1920). *Der golem*. PAGU.
- Whale, James (1931). *Doctor Frankenstein*. Universal Pictures.
- Whale, James (1935). *Bride of Frankenstein*. Universal Pictures.
- Wilder, Billy (1960). *The apartment*. United Artists.

Visión general de la inteligencia artificial en medicina

José Luis García Vigil*

Introducción

Aún existe controversia en la definición más adecuada de inteligencia, con mayor razón esta controversia se generaliza cuando se trata de aplicar al campo de los sistemas de información o a la informática computacional. En el humano se acepta que la inteligencia es aquella cualidad que permite a nuestros congéneres entender el mundo en que vivimos y aprender a relacionarnos con nuestros semejantes; todo ello en afán de tener una vida plena y consciente llena de satisfactores tanto físicos como emocionales.

La inteligencia general del humano (IH), es en realidad una cualidad emergente y singular formada por un conjunto armonioso interrelacionado de cualidades específicas o competencias particulares que se diferencian fenomenológicamente por ser conscientes y enfocadas a objetivos sensibles, emocionalmente importantes y que al conocerlos y conceptualizarlos como problemas, se pueden resolver para llegar a un estado de homeostasis compatible con la supervivencia de la especie, calidad de vida y satisfacción del resultado que lleva a la sensación de logro físico, biológico, fisiológico o espiritual.

Estas competencias o cualidades conscientes y singulares se han dado a conocer y difundido por el psicólogo Howard Gardner en su obra "*Estructuras de la mente. Teoría de las Inteligencias Múltiples*"⁽¹⁾; las que se expresan en el contexto de tareas, disciplinas y ámbitos específicos, las que se pueden concebir en términos neurobiológicos, a saber:

1. Inteligencia Lingüística.
2. Inteligencia Musical.
3. Inteligencia Lógico-matemática.
4. Inteligencia Espacial.
5. Inteligencia Cinestésico-corporal.
6. Inteligencias Personales (Conocimiento del YO, intrapersonales, introspección).
7. Inteligencias Interpersonales (socialización, establecimiento de vínculos).

Sin ser el motivo específico del tema que nos ocupa, sólo menciono con el propósito de partir con base en conceptos y argumentos con algún grado de evidencia respecto de la inteligencia humana; y esto, con el fin de extrapolarlo a lo que ahora se conoce como Inteligencia Artificial (IA) o inteligencia de las máquinas (robots, androides, "bots").

El término inteligencia artificial (IA) se empezó a utilizar en la década de los años 50s del siglo pasado, eran dispositivos inicialmente analógicos y después digitales con la elaboración y aplicación de programas de tipo determinista; es decir, escritura de instrucciones para realizar alguna acción (resultado) siempre y cuando se cumpliera alguna condición, Programación tipo IF, THEN, ELSE: Si se da la situación de poder avanzar en la instrucción, entonces identificar si se cumple la condición deseada para tener el resultado esperado; o en dado caso, detener el proceso. Volver a empezar o reprogramar las instrucciones.⁽²⁾

Como se puede observar, este tipo de programación tenía muchas limitaciones dada la gran cantidad de variables que se podrían presentar en escenarios complejos de la vida diaria y, por tanto, las posibilidades son infinitas ante situaciones no previstas. Esta situación se pudo superar con el desarrollo de códigos y algoritmos que, con la instrucción inicial y recursividad, el programa por sí mismo se reforzaba en algunos códigos o desechara los algoritmos defectuosos a partir de ejemplos tipo y se elaboraban de manera automática, las correcciones necesarias; esto llevó a un adelanto más de las máquinas de cómputo (*machine learning*). Inteligencia artificial (IA) de aprendizaje profundo.

*Especialista en Medicina Interna; Certificado por el CMMI; Maestría en Ciencias Médicas. Investigación clínica y educativa; Educación a distancia; Profesor, Fac. de Med., UNAM.

Ahora nos encontramos en la frontera del conocimiento, donde sin temor a equivocarnos, el futuro ya nos alcanzó. Estamos ante una nueva realidad, un gran desarrollo de la IA de aprendizaje profundo, el cual se ha desarrollado a partir del modelo de las redes neuronales, inicialmente en paralelo y después en segunda y tercera dimensión, que hacen factible se establezcan redes complejas (redes de redes) tal y como si fueran redes de neuronas del sistema nervioso central interrelacionadas con base en sinapsis. Los nuevos programas y códigos de IA son ya de tipo convolucional (en capas) y genéticas (generativas).

Los dispositivos de IA más evolucionados son los capaces de compilar una gran cantidad de conceptos, términos, frases, párrafos completos respecto de una temática o un campo semántico y disciplinar específico; expresando sus resultados en textos fluidos, didácticos y atractivos, pero carentes de criterio y reflexión. No están mediados por sentimientos ni emociones, solo compilación exitosa de textos creíbles, pero susceptibles de tener errores de criterio (porque las máquinas no piensan y reflexionan como los humanos). El ejemplo actual es la tecnología digital llamada ChatGPT.

A esta semblanza introductoria, tenemos en primer lugar los conceptos de IH y un símil muy alejado de ella que es la IA, pero indudablemente en las siguientes décadas, tendremos con evolución del Internet 2.0, el Internet de las Cosas (IoT) y robots humanoides con sensores periféricos similares al del humano (video, audio, tacto, cinestesia, olfato y gusto digitales); el escenario casi completo del Ciberespacio con IA multiconectada a terminales inteligentes (computadoras macro tipo servidores, computadores de escritorio, tabletas, teléfonos celulares) y comunicación IoT merced a redes de fibra óptica con los robots humanoides (androides) en convivencia íntima e inteligente con los humanos.⁽²⁾

Dentro de todo este escenario presentaré una visión general de la IA el ejercicio y desarrollo de la Medicina y de sus personajes

interactuantes; llámense pacientes, estudiantes de medicina y médicos en ejercicio profesional en el campo de la salud, desarrollando particularmente el antecedente de la IA en forma de Sistemas Expertos, después, la aplicación de la IA en el proceso salud enfermedad en sus tres niveles de prevención (1ª, 2ª y 3ª; con énfasis en la Atención Primaria) y con foco en el sujeto objetivo de todos nuestros afanes, el paciente. Por último, unos párrafos relacionados con la ética(bioética) que se expresa o está implicada en la interacción e integración de la IH con la IA.

Aplicación actual y a futuro de la IA en Medicina

Sistemas Expertos:

Cuando los programas de IA estaban suficientemente evolucionados, con Internet, redes locales y de área (LAN, GAN), digitalización de gran cantidad de información con programas y bases de datos relacionales, emanadas del proceso salud enfermedad en la Atención Médica, surgen los Sistemas Expertos (SE), El objetivo de los SE es el de apoyar en el trabajo de los médicos de ciertas especialidades; por ejemplo, sistemas de ayuda para realizar diagnósticos tempranos y eficaces, a fin de tomar decisiones en cualquiera de los niveles de prevención en el proceso de atención a la salud. Estos SE fueron utilizados más en las labores de diagnóstico oportuno y tratamiento médico adecuado cuando en virtud de los signos y síntomas (síndromes, enfermedades) de los pacientes, el médico llegaba a un diagnóstico de alta probabilidad (hipótesis diagnóstica) y a un nivel crítico en la toma de decisión terapéutica.⁽²⁻⁴⁾

Atención Médica:

La atención médica inicia cuando se confrontan y ponen en contacto los personajes importantes e interactuantes del acto médico: Pacientes, médicos y estudiantes de medicina; tal como sucede en los modelos de comunicación humana (entrevista médica, relación médico-paciente, historia clínica) donde existe diálogo bidireccional

(juego de roles entre preceptor y emisor de información), comprensión, interpretación, aceptación vs negación para finalmente, tras negociación a ganar-ganar; se llega al establecimiento de compromisos respecto del tema en salud de que se trate, en forma de la toma de varias decisiones como la necesidad de solicitar estudios clínicos de laboratorio y gabinete, a fin de establecer con alto grado de probabilidad la hipótesis diagnóstica con la cual llegar al nivel crítico en la toma de decisiones, tanto diagnósticas como terapéuticas con medidas farmacológicas (medicamentos) y no farmacológico (estilos de vida, control de peso, vida laboral, calidad de vida, actividad física, recreación, descanso, nutrición).

En todo el proceso del acto médico se genera información y gran cantidad de datos. Mucha información fluye verbalmente en la relación médico-paciente, otra se recaba al realizar el examen médico según el arte de la clínica y de acuerdo con la propedéutica, nosología y anatomía topográfica. Otro tipo de información y datos se obtienen cuando se miden las constantes vitales (peso, talla, presión arterial, temperatura) y los resultados de exámenes de laboratorio (exámenes de sangre básicos y especiales, orina, heces, otros líquidos corporales normales o anormales, histopatología de biopsias) y gabinete (imagenología simple y especial (radiografías, tomografías computarizadas, resonancias magnéticas, ultrasonidos), electrocardiograma en reposo o de esfuerzo, electroencefalograma, etcétera.

La información y datos así obtenidos son susceptibles de expresarse en lenguaje analógico y digital y visualizarse para su análisis e integración por el personal del equipo de salud, en especial el médico experto tratante de un paciente en particular. La expresión gráfica y digitalizada en la actualidad se genera en gran parte de manera automatizada (equipo analizador automatizado con códigos y algoritmos de IA). Esta información simple o parcialmente integrada se obtiene en formatos específicos que el médico tra-

tante lo descifra según su preparación e integra con la similar; obtenida de la relación médico-paciente, el examen físico y la historia clínica, traduciéndola en diagnósticos de alta probabilidad a la manera de síndromes específicos de diversas enfermedades y estableciendo las medidas de estudio adicionales o terapéuticas según el caso.^(3, 4)

"Los dispositivos de IA más evolucionados son los capaces de compilar una gran cantidad de conceptos, términos, frases, párrafos completos respecto de una temática o un campo semántico..."

Esta vasta información que se obtiene durante la atención médica generalmente termina impresa de manera digital (computadoras) en un formato estructurado llamado historia clínica. Historia clínica digital a la cual se agrega toda la información de laboratorio y gabinete para formar un expediente. En la actualidad ya se tienen por paciente bases de datos relacionales y programas de IA que tienen como salida la información que el médico revisa y solicita de cada paciente. Esta gran base de datos relacionales y programas están estructurados en un gran formato llamado Expediente Clínico Electrónico Estructurado (ECEE)⁽²⁾, elaborado según Norma Oficial Mexicana (NOM). Entonces tenemos un cúmulo enorme de información obtenida por paciente y susceptible de organizarse por institución de salud (IMSS, ISSSTE, SEMAR, SEDENA, INSTITUTOS NACIONALES DE SALUD, etcétera), por Unidad de Atención Médica (Medicina Familiar,

Hospitales de Zona o Área, Hospitales Regionales, Hospitales de Especialidades); y finalmente, por delegaciones, municipios, ciudades, estados, regiones del país y nación.

Avalancha de información ya estructurada y organizada según se menciona anteriormente y susceptible de analizarse por medio de códigos y algoritmos de IA. Realizar cálculos simples o estadísticos complejos a diferente escala partiendo de lo micro o individual del paciente, hasta el macro de unidades, instituciones de salud, localidades geográficas, grandes poblaciones y de índole nacional. Factible de realizar con el diseño de grandes bases de datos (Big Data) ⁽²⁾ y programas específicos para conocer de una información aparentemente caótica; tendencias claras en el rumbo, dirección y tipo de información. Así conoceremos más mediante este tipo de investigación (elemental, de servicios de salud, epidemiológica y clínica), de que se enferman los mexicanos, cual es el grado de deterioro en la salud y calidad de vida de nuestra población.

Aún más y con enfoque de las ciencias de la complejidad, conocer y atender mejor a la población mexicana en todo el proceso de salud-enfermedad con medidas de prevención pertinentes y oportunas y que esta atención primaria, sea más eficiente y costo-efectiva, en especial con foco en la tercera edad blanco de padecimientos crónicos degenerativos y neoplásicos. ⁽⁵⁾ Papel relevante jugó y sigue aportando la IA en la atención médica y los servicios de salud a la población durante la pandemia de COVID-19; desde la investigación epidemiológica y clínica (ensayos clínicos y cuasiexperimentales, metaanálisis y revisiones sistemáticas) para encontrar la fórmula adecuada de medicamentos antivirales, esquemas terapéuticos probando antivirales conocidos y medidas farmacológicas y no farmacológicas de soporte (distanciamiento social, uso de cubrebocas, lavado de manos, uso frecuente de antisépticos antivirales, uso suplementario de oxígeno) ⁽⁶⁾, anticuerpos monoclonales anticovid-19 y vacunas, culminando en vacu-

nas anticovid-19 tipo mRNA (Pfizer y Moderna) y antivirales específicos anticovid-19.

Impacto de la IA en la medicina compleja y biotecnológica

En la actualidad no se concibe una medicina, servicios de salud y atención médica, donde no se aborde de manera integral el padecer de nuestros semejantes; esto es, cuando se ubican en alguna de las partes del gran continuo que es el proceso salud-enfermedad. El organismo humano como un todo (órganos, aparatos y sistemas, así como tejidos, células y red de redes interrelacionadas de funciones, procesos biológicos moleculares y bioquímicos) sigue siendo un misterio. Las áreas y funciones menos conocidas son las de escala microscópica intracelular y otras a nivel molecular y en escala nano. Algunas funciones y mecanismos bioquímicos poco explorados lo son no por escaso interés de académicos e investigadores, sino porque todavía no se disponía de las herramientas necesarias para tal fin. Esto ha sido factible con el gran desarrollo reciente de dispositivos digitales y microprocesadores alimentados con programas de IA de aprendizaje profundo, convolucional y con códigos y algoritmos genéticos (generativos).

Estas innovaciones tecnológicas en el área de la medicina y que han recibido más atención por futura aplicación en el campo de la salud y enfermedad son:

- **Biología sintética.** ⁽⁷⁾

Fabricación de células, tejidos y órganos programados para que lleven a cabo procesos controlados por los seres humanos", un mundo completamente nuevo con muchas ventajas. Fabricación a demanda y por necesidad, dada la carencia o mala función de un órgano: Riñones, hígados, pulmones, suprarrenales, tiroides, piel, músculos, huesos; por solo mencionar algunos, y que aliviarían la carencia actual de órganos para trasplante.

- **Edición de genes.** ⁽⁸⁾

El sistema *CRISPR/Cas* es parte de un sistema inmune adaptativo que los organismos procariontes desarrollaron para defenderse de la incorporación de material genético exógeno. Este sistema de inmunidad está mediado por una nucleasa específica que degrada al DNA invasor y posteriormente algunos fragmentos de la molécula degradada se almacena para reconocer y eliminar secuencias similares en el futuro. Recientemente fue posible reprogramar este sistema para reconocer cualquier secuencia de DNA y realizar ediciones genéticas en una gran cantidad de organismos de manera altamente específica. El sistema *CRISPR/Cas* ha sido adaptado para el desarrollo de una estrategia altamente específica en el tratamiento de infecciones producidas por bacterias resistentes a antimicrobianos.

También tendría aplicación en gran cantidad de padecimientos genéticos por deficiencia de alguna sustancia indispensable para la vida, o por exceso de ella dada su nula degradación y consiguiente acumulación tisular con disfunción orgánica (glucoproteínas, enzimas, proteínas estructurales de huesos, tejido pulmonar, óseo, etcétera).

- **La medicina regenerativa.** ⁽⁹⁾

Inducción de células madre pluripotentes: novedades en medicina regenerativa. “*Stem cells*” (células tallo). La medicina regenerativa busca contrarrestar el envejecimiento y los daños celulares del cuerpo humano. Un nuevo método de obtención de células pluripotentes descubierto por el Dr. Yamanaka en 2012, nombrado reprogramación celular, ha permitido la inducción de células madre pluripotentes (iPSC) a partir de células de la piel. La investigación liderada por el doctor Michael Edel de la Unidad de Anatomía y Embriología de la UAB ha mejorado este método a través de la sustitución del gen c-Myc (oncogen) por

Ciclina D1. Esto podría sentar las bases para futuras aplicaciones clínicas.

Las aplicaciones en medicina clínica son varias, pero principalmente en la regeneración de tejidos dañados por alguna enfermedad adquirida, ya sea crónica degenerativa, inmunológica o por cáncer. Podría en algunos casos mejorar la función de órganos atrofiados o invadidos por tejido no funcional (fibrosis) o cáncer metastásico; o aumentar la sobrevivencia antes de algún procedimiento de trasplante de órganos. Ejemplos: Cardiopatía isquémica crónica congestiva, insuficiencia renal, insuficiencia hepática, leucemias, etcétera.

Aspectos éticos en interacción de la Inteligencia Humana (médicos, medicina) vs Inteligencia Artificial:

Gran parte del personal médico que atiende a los pacientes, tanto en medicina privada como en instituciones públicas de salud, son estudiantes de medicina general y residentes en formación en alguna especialidad médica; además de los médicos ya formados y recién graduado que están en constante proceso de actualización y educación continua. En todos ellos es indispensable una gran vocación de servicio, compasión y empatía, a fin de que los servicios médicos sean proporcionados con alto sentido humanitario. La readecuación curricular efectuada en 2010 en la licenciatura de la Facultad de Medicina de la UNAM ya contempló en los planes y programas de estudio a la ética médica y bioética como base deontológica de su formación; asimismo, se emitió también en dicha fecha, la declaración de principios éticos del educador en medicina. ⁽¹⁰⁾ Por otro lado, todos los médicos cursan la asignatura de informática médica, donde un papel relevante lo juega el conocimiento de la IA.

En internet ha crecido la inteligencia artificial hasta convertirse en un programa con códigos y algoritmos que aprenden y se reprograman para efectuar tareas preestablecidas con mayor eficiencia; si bien lo ante-

rior se traduce en mejoría, el programador desconoce los alcances de los resultados y de la reprogramación. Ante el riesgo de desviación de los objetivos preestablecidos y de los reglamentos éticos, se tienen que implementar filtros al inicio, durante y al final del proceso, como alarmas cuando existan desviaciones con implicación bioética. La interacción de la inteligencia humana con la

inteligencia artificial ha tenido desencuentros negativos y positivos. Al principio, bastó con adecuar normas, leyes laborales y derechos humanos; ahora se requiere establecer normas éticas, como las formuladas en la Declaración de Barcelona para el Adecuado Desarrollo y Uso de la Inteligencia Artificial en Europa.⁽¹¹⁾

Referencias

1. Gardner H. *Estructuras de la mente. La teoría de las inteligencias múltiples*. Segunda edición. México: Fondo de Cultura Económica; 1994.
2. Huesca E, Juárez J, Cicero P; Coordinadores. *Mi Vecino Es Un Robot. Los Retos De Convivir Con La Inteligencia Artificial*. Primera edición. México. Debate; diciembre 2022.
3. Lifshitz GA, Ramiro-Hernández M, García-Vigil JL, Mendoza-Salas K, Jinich-Brook H. *Síntomas y signos cardinales de las enfermedades*. México: Octava edición. Manual Moderno; noviembre 2022.
4. García-Vigil JL. *Principios generales de prescripción en el proceso de razonamiento diagnóstico*: Research Gate 08/2010.Consulta 24/05/2023. https://www.researchgate.net/publication/311106313_PRINCIPIOS_GENERALES_DE_PRESCRIPCION_PROCESO_DE_RAZONAMIENTO_DIAGNOSTICO.
5. García-Vigil JL. Las enfermedades crónicas y neoplásicas, desde las ciencias de la complejidad y la atención primaria. *Rev Med Inst Mex Seguro Soc* 2010; 48 (1): 7-12.
6. García-Álvarez JL, García-Vigil JL. Guidelines for clinical management of SARS-CoV-2 infection. *Gac Med Mex*. 2020; 156:576-583.
7. Perspectivas de Salud Pública. Biología Sintética. *Salud Pública de México*. Consulta 26-05-2023. <https://www.scielosp.org/pdf/spm/v52n3/12.pdf>
8. Chávez-Jacobo V. El sistema de edición genética CRISPR/Cas y su uso como antimicrobiano específico. *Revista Especializada en Ciencias Químico-Biológicas*; 2018; 21(2): 116-123, 2018. DOI: 10.22201/fesz.23958723e.2018.2.5 <https://www.scielo.org.mx/pdf/tip/v21n2/1405-888X-tip-21-02-e201825.pdf>
9. Universidad Autónoma de Barcelona (UAB Divulga). *Inducción de células madre pluripotentes: novedades en medicina regenerativa*. Consulta 26-05-2023. <https://www.uab.cat/web/detalle-noticia/inducccion-de-celulas-madre-pluripotentes-novedades-en-medicina-regenerativa-1345680342040.html?noticiaid=1345854825813>
10. García-Vigil JL, García-Mangas JA, Ocampo-Martínez J, Martínez-González A. Declaración de principios éticos del educador en medicina. *Rev Med Inst Mex Seguro Soc*. 2011; 49:571-574.
11. García-Vigil JL. Reflexiones en torno a la ética, la inteligencia humana y la inteligencia artificial. *Gac Med Mex*. 2021; 157:311-314.



Imagen: freepik.com

Humanos y máquinas inteligentes: la responsabilidad en el uso de inteligencia artificial en la atención médica

Elisa C. Calleja Sordo*

"Nosotros somos responsables de todo y de todos, ante todos, y yo más que todos los demás"

Dostoyevski

1. Introducción

A través de los años la atención médica ha cambiado, y gracias a los avances tecnológicos actualmente encontramos el uso de la inteligencia artificial jugando un papel cada vez más importante en este campo.

Se utiliza para diagnósticos por imagen, genética, electrodiagnósticos, seguimientos fisiológicos, evaluaciones de discapacidad y tamizajes masivos, entre otros, y en los últimos diez años su aplicación ha aumentado (Jiang et al., 2017), lo que nos permite afirmar que es una tecnología que no solo continuará desarrollándose, sino que cada vez más el proceso de atención médica se apoyará en ella.

Indudablemente esto conlleva grandes ventajas: diagnósticos de manera oportuna, reducción de errores médicos, mayor precisión en algunos casos y eficiencia, por mencionar algunos ejemplos. Sin embargo, es importante cuestionar el papel que la responsabilidad tiene en el uso de la inteligencia artificial en este campo de aplicación, buscando que los pacientes tengan una atención médica integral, en donde tanto la inteligencia artificial como el personal médico involucrado funcionen de manera conjunta.

2. ¿Qué es la inteligencia artificial?

Una forma para establecer a qué nos referimos cuando se habla de inteligencia artificial, es observando sus componentes. Por un lado, la palabra artificial indica que la entidad es el producto de seres humanos, y que no puede existir de forma natural sin que estos estén involucrados. Un artefacto es un objeto creado por el hombre que no se presenta naturalmente, sino que ocurre como resultado de preparativos o investigación realizados por los mismos. Por otro lado, inteligencia es usualmente definida como la habilidad de adquirir y aplicar conocimiento;

una definición más comprensible se refiere a ella como el uso hábil de la razón, el acto de comprender y la habilidad de pensar de manera abstracta medida por criterios objetivos. Es así como la inteligencia artificial se refiere a una entidad creada por seres humanos que posee la habilidad de entender y comprender conocimiento, de razonar utilizando dicho conocimiento e incluso actuar debido al mismo (Ekmekci & Arda, 2020).

John McCarthy la define como la ciencia y la ingeniería de hacer máquinas inteligentes, especialmente programas informáticos inteligentes, y está relacionada con la tarea similar de utilizar computadoras para comprender la inteligencia humana, pero la inteligencia artificial no tiene que limitarse a métodos biológicamente observables (2011).

Asimismo, es una tecnología informática impulsada por algoritmos, programada para autoaprender a partir de datos, hacer predicciones inteligentes, y tomar decisiones en tiempo real utilizando redes neuronales artificiales, aprendizaje automático, automatización de procesos robóticos y minería de datos. En el área de la atención médica es el uso de tecnologías inteligentes basadas en datos que aprovechan los recursos e información de la práctica médica de forma efectiva para respaldar y agilizar la toma de decisiones y, en consecuencia, brindar un mejor cuidado que se adapte a las necesidades individuales. (Siala & Wang, 2022, p. 1).

Para informar de manera efectiva la toma de decisiones en el cuidado de la salud, las tecnologías de inteligencia artificial usualmente utilizan algoritmos de aprendizaje

* Licenciada en Filosofía, U. del Claustro de Sor Juana; Maestra y Doctora en Ciencias, UNAM; Doctora en Filosofía, U. de Salamanca; Miembro de *Cefiloe*, U. de Valparaíso, Chile.

automático para realizar actividades analíticas e inferencias que pueden ser llamadas inteligentes, que se utilizan para detectar y predecir pandemias, enfermedades, diagnosticar y manejar condiciones crónicas y neurológicas, interpretar escáneres médicos e imágenes de radiología, dar servicios y tratamientos médicos, investigación de fármacos, y emparejar pacientes a ensayos clínicos de acuerdo a sus características (Siala & Wang, 2022, p. 1).

3. Problemas derivados del uso de inteligencia artificial en la atención médica

Indudablemente el uso de la inteligencia artificial en la atención médica tiene grandes beneficios para los pacientes, así como para el personal médico, pero es importante también presentar los problemas que se pueden derivar del uso de esta tecnología.

Si bien a continuación no se agotarán las problemáticas que pueden llegar a derivarse del uso de esta tecnología, son cinco puntos importantes que deben abordarse conforme su uso va ganando terreno (Sand et al., 2022, p. 163):

- 1) Privacidad y protección de datos de los pacientes: La inteligencia artificial recopila y analiza una gran cantidad de información personal, lo cual plantea preocupaciones éticas entorno a la privacidad, confidencialidad y seguridad de los datos y de los pacientes. No podemos continuar hablando de confidencialidad en sentido tradicional, como bien lo señaló Mark Siegler (1982), pero es importante responder a la pregunta por cómo se protegerá la privacidad de los pacientes al integrar inteligencia artificial en el proceso médico.
- 2) Toma de decisiones automatizadas: Esta tecnología puede tomar decisiones clínicas, médicas y éticas, pero ello nos enfrenta con preguntas sobre la equidad y transparencia del proceso que lleva a dicha decisión, dado que el proceso deliberativo es interno. Es lo que se se-

ñala como la naturaleza de *caja negra* propia de estos sistemas, en los que no existe propiamente una explicación sobre cómo producen un determinado resultado -output- a partir de información específica que se le proporciona -input- (Holm, 2021, p. 2; Sand et al., 2022, p. 167).

- 3) Sesgos y discriminación por parte de los algoritmos: Los algoritmos propios de la inteligencia artificial pueden verse influenciados por sesgos inherentes en los datos de entrenamiento, lo que puede llevar a resultados injustos o discriminatorios (Holm, 2021, p. 7).
- 4) Autonomía y consentimiento informado: El uso de la inteligencia artificial nos enfrenta con la pregunta por la autonomía de los pacientes y el consentimiento informado, ¿cómo se puede garantizar que los pacientes comprendan el proceso y puedan participar activamente en las decisiones sobre su atención médica estando la inteligencia artificial de por medio?
- 5) Rendición de cuentas: A medida que la inteligencia artificial desempeña un papel más importante en la toma de decisiones y la atención médica, cabe preguntar quién responde en caso de que se presenten errores o daños causados por los sistemas de inteligencia artificial, dado que no podemos señalar a un sistema operativo o máquina como culpable (Jasanoff, 2016, p. 41).

Si bien esto no agota los problemas que se pueden derivar del uso de la inteligencia artificial dentro de la atención médica, evidencian el riesgo que puede implicar el uso de esta tecnología, por lo que propondremos la responsabilidad como elemento que debe acompañar el proceso médico cuando está complementado o sustituido por la inteligencia artificial.

4. Responsabilidad

Veamos ahora a qué nos referimos al hablar de responsabilidad: se enfoca en actos que

se han llevado a cabo y que exigen desde fuera que el autor de la acción asuma las consecuencias.

De acuerdo con Hans Jonas (...) la «responsabilidad» se refiere a actos hechos y que adquiere realidad en el *hacer* desde fuera responsable a alguien por lo que ha hecho. El *sentimiento* que aquí acaso acompañe al agente, sentimiento por el cual asume interiormente la responsabilidad (sentimiento de culpa, arrepentimiento, disposición a expiar, pero también obstinado orgullo), es igual de retroactivo que el objetivo «tener que» responder de lo hecho (Jonas, 2015, p. 162).

En este sentido, ser responsable de los actos propios significa tener capacidades volitivas y cognoscitivas suficientes para imputar a uno el acto particular en cuestión, “pero no prejuzga en nada la responsabilidad en relación con ese acto” (Canto-Sperber, 2011, p. 1398).

Al preguntar por las condiciones en las que una persona *responsable de sus actos* lo es de un acto en particular, el cual puede ser condenable o meritorio a primera vista, podemos responder que

Una natural sugestión consiste en decir que éste tiene algo que ver en todo acto en el que ha habido un compromiso volitivo y cognoscitivo mínimo de su parte, en el que el compromiso volitivo mínimo ha sido un acto (es decir, un movimiento corporal intencional y no un movimiento simplemente sufrido, efecto de algún mecanismo fisiológico o de una causa externa, como la fuerza física ejercida por algún otro) y, el compromiso cognoscitivo mínimo, que el agente sea consciente de lo que hace (Canto-Sperber, 2011, p. 1398).

Sin embargo, ¿qué pasa cuando estamos tratando con inteligencia artificial en la atención médica? Por un lado, podemos señalar la responsabilidad de los actores involucrados en el proceso, pero no podemos hablar de la responsabilidad de la tecnología informática, dado que no se refiere a un agente que tenga capacidades volitivas y cognoscitivas que impulsen los actos.

Es por esto importante que la responsabilidad recaiga en los actores involucrados, y que no se derive parte de la misma a la inteligencia artificial. En el caso de la privacidad y protección de los datos de los pacientes, la UNESCO recomienda respetar, proteger y promover la privacidad a lo largo del ciclo de vida de los sistemas de Inteligencia artificial, asegurando que los datos se recopilen, utilicen, compartan, archiven y supriman de acuerdo con el derecho internacional y los valores presentes en documentos emitido por las instituciones (UNESCO, 2022, p. 22).

Al hablar de la toma de decisiones de forma automatizada, es importante que el papel del médico no se limite a la prestación de cuidados de la salud, sino que se incluya la responsabilidad de interpretar, mediar y salvaguardar entre el paciente y la inteligencia artificial, esto con el fin de evitar que se deshumanice la medicina, dado que se trata de una tecnología que no promueve la interacción entre las personas, el respeto por la identidad, ni sirve a fines humano (Jotterand & Bosco, 2022). De igual manera, los médicos deben poder evaluar la información proporcionada a la tecnología informática y saber el tipo de información que se utiliza en cada procedimiento (Sand et al., 2022, p. 167), esto con el fin de entender los límites de esta tecnología, lo cual contribuye al asumir la responsabilidad al momento de la toma de decisiones.

Lo anterior permite abordar el problema de sesgos y discriminación presentes en los al-

goritmos utilizados partiendo del mismo principio, la supervisión de la información proporcionada a la inteligencia artificial, para salvaguardar la dignidad de los pacientes.

En el caso de la autonomía de los pacientes y el consentimiento informado, se asume la responsabilidad al integrar a éstos en la toma de decisiones, es decir, explicando a los pacientes cuál es la información con la que se cuenta y que utilizará la inteligencia artificial, así como haciéndolos partícipes a lo largo del proceso. La UNESCO lo plantea como una recomendación, “La transparencia y la explicabilidad están estrechamente relacionadas con las medidas adecuadas de responsabilidad y rendición de cuentas, así como con la fiabilidad de los sistemas de IA” (UNESCO, 2022, p. 23).

Por último, cuando hablamos de la rendición de cuentas, se recomiendan mecanismos de supervisión y control, es decir, una responsabilidad directa, la responsabilidad por la vigilancia de la información proporcionada a la inteligencia artificial, el proceso de análisis y el resultado del mismo, la presencia de personal capacitado que asuman las consecuencias, tanto positivas como negativas, del proceso médico.

Responsabilidad conjunta

Es importante establecer el tipo de responsabilidad de la que se está hablando, dado que al abordar la atención a la salud no tratamos con una sola persona, es decir, el médico tratante y directamente responsable del paciente, sino que nos encontramos con una red de agentes involucrados, y aún más al haber un sistema de inteligencia artificial involucrado en la atención médica.

El médico tratante es el primer agente que podemos señalar como responsable, sin embargo, dada la naturaleza de la atención médica y la tecnología involucrada, es importante considerar a todos los actores que juegan un papel: el fabricante del equipo que se utiliza, el especialista en programar-

lo, el personal médico, y el paciente. Esto nos habla de una responsabilidad conjunta, en donde se tiene un objeto común que a su vez permite una actividad basada en la corresponsabilidad de todos los participantes, se trata de una responsabilidad genuina y que no se limita al éxito de la participación, sino que también acepta las consecuencias de un posible fracaso, lo cual no se refiere a una catástrofe paralizante, sino que puede motivar los próximos esfuerzos conjuntos que den como resultado el éxito deseado (Kwiatkowski, 2018, p. 201).

Es lo que Michael Bratman señala como responsabilidad colectiva, la cual se basa en una actividad cooperativa compartida, donde destaca un tipo específico de interacción interpersonal y está formado por tres rasgos característicos: (i) capacidad de respuesta mutua, (ii) compromiso con la actividad conjunta y (iii) compromiso de apoyo mutuo (1992, p. 328).

La capacidad de respuesta mutua se refiere a la intención de los actores involucrados por responder a las acciones e intenciones de los otros. En el caso del uso de la inteligencia artificial en la atención médica, la encontramos presente cuando cada participante orienta su comportamiento de acuerdo con los de los demás, y sabiendo que es recíproco, es decir, cuando los agentes buscan que la atención médica sea más precisa y eficiente con el uso de inteligencia artificial y sus esfuerzos se enfocan en lograrlo.

El compromiso con la actividad conjunta es cuando cada participante se compromete con la actividad en cuestión, y la capacidad de respuesta mutua se encuentra en función de lograr ese compromiso, en este caso, desarrollar la tecnología que permita ofrecer una atención médica más eficiente.

Por último, el compromiso de apoyo mutuo nos habla del compromiso que cada participante tiene de apoyar el esfuerzo de los otros para que cada uno pueda desempeñar su papel en la actividad conjunta, es decir,

en la interacción y colaboración que existe entre los involucrados, complementando así el papel que cada uno desempeña.

Se trata de una colaboración en el sentido en que los diferentes actores presentes trabajan con la misma finalidad en mente: el desarrollo e implementación de la inteligencia artificial para la atención médica.

Conclusión

Podemos ver que tanto la inteligencia artificial como los actores involucrados, participan de la actividad, la atención médica, teniendo como fin mejorar el estado de salud del paciente, lo que nos lleva a afirmar la necesidad de una responsabilidad conjunta, en donde ésta no se diluya entre los diferentes actores, sino que cada uno de ellos reconozca el papel que juega y el impacto que tiene en la salud y en la vida de los pacientes.

Se debe priorizar la presencia humana en el proceso de atención médica, lo cual no debe impedir que se implemente la tecnología en el contexto clínico, sino que la complemente, teniendo como meta la salud de los pacientes.

Es decir, podemos confiar en la inteligencia artificial como medio para dar una mejor atención médica, pero no debemos ceder el control, sino que este debe al final recaer en los seres humanos "(...) ya que estos pueden recurrir a los sistemas de IA en la adopción de decisiones y en la ejecución de tareas, pero un sistema de IA nunca podrá reemplazar la responsabilidad final de los seres humanos y su obligación de rendir cuentas" (UNESCO, 2022, p. 22).

Referencias

- Bratman, M. E. (1992). Shared Cooperative Activity. *The Philosophical Review*, 101(2), 327–341. Doi: 10.2307/2185537
- Canto-Sperber, M. (2011). *Diccionario de ética y de filosofía moral* (II). Fondo de Cultura Económica.
- Ekmekci, P. E., & Arda, B. (2020). *Artificial Intelligence and Bioethics (1º)*. Springer Briefs in Ethics. Springer International Publishing.
- Holm, S. (2021). Handle with care: Assessing performance measures of medical AI for shared clinical decision-making. *Bioethics*, 36(2). <https://doi.org/10.1111/bioe.12930>
- Jasanoff, S. (2016). *The Ethics of Invention: Technology and the Human Future* (1º edición). W.W. Norton & Company, Inc.
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y [Yilong], Dong, Q., Shen, H., & Wang, Y [Yongjun]. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4), 230–243. <https://doi.org/10.1136/svn-2017-000101>
- Jonas, H. (2015). *El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica*. Herder. <https://doi.org/10.2307/j.ctvt9k2sz>
- Jotterand, F., & Bosco, C. (2022). Artificial Intelligence in Medicine: A Sword of Damocles? *Journal of Medical Systems*, 46(9). <https://doi.org/10.1007/s10916-021-01796-7>
- Kwiatkowski, W. (2018). Medicine and technology. Remarks on the notion of responsibility in the technology-assisted health care. *Medicine, Health Care, and Philosophy*, 21(2), 197–205. <https://doi.org/10.1007/s11019-017-9788-8>
- McCarthy, J. (2011). What is AI? / *Basic Questions*. <http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html>
- Sand, M., Durán, J. M., & Jongsma, K. R. (2022). Responsibility beyond design: Physicians' requirements for ethical medical AI. *Bioethics*, 36, 162–169. <https://doi.org/10.1111/bioe.12887>
- Siala, H., & Wang, Y [Yichuan] (2022). Shifting artificial intelligence to be responsible in healthcare: A systematic review. *Social Science & Medicine* (1982), 296, 1–15. <https://doi.org/10.1016/j.socscimed.2022.114782>
- Siegler, M. (1982). Sounding Boards. Confidentiality in medicine - a decrepit concept. *The New England Journal of Medicine*, 307(24), 1518–1521. <https://doi.org/10.1056/NEJM198212093072411>
- UNESCO. (2022). *Recomendaciones sobre la ética de la inteligencia artificial*. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

Reconocimiento facial: una herramienta de alto riesgo

Juan Espíndola Mata*

El reconocimiento facial es una forma de inteligencia artificial que implica la extracción, digitalización y comparación automatizadas de la distribución espacial y geométrica de los rasgos faciales para identificar a las personas. Son cuatro sus variedades, según el tipo de recolección de datos que lleven a cabo, cada una con sus implicaciones éticas. De la más básica hasta la más compleja, estas funciones son: la “detección”, que no recopila información de identificación personal; la “caracterización”, que recopila información como género, rango de edad e indicadores emocionales, pero no recopila ni retiene plantillas faciales de identificación personal; la “verificación”, que recopila y retiene plantillas faciales, y luego utiliza un sistema de coincidencia “uno a uno” en el que el software determina si la persona es quien dice ser (por ejemplo, un teléfono inteligente escanea el rostro de la persona para determinar si coincide con la plantilla almacenada en el teléfono); y, de mayor relevancia para el propósito de este ensayo, la “identificación”, que también crea una plantilla identificable de una persona única pero la usa en un sistema de correspondencia de “uno a muchos,” donde el sistema compara una imagen recopilada con una base de datos existente (Selinger, Evan y Leong 2021).

El uso de la tecnología de reconocimiento facial ha ido ganando terreno en las sociedades contemporáneas. Es común su utilización en las actividades policiales, y distintas compañías como PimEyes, FindClone o, lo más controvertido de todos, Clearview AI, ponen el servicio al alcance de agentes estatales y no estatales. Con el avance del reconocimiento facial, las críticas dirigidas a ella han aumentado de manera paralela. El despliegue de esta tecnología en regímenes autoritarios y democráticos liberales por igual para perseguir a grupos étnicos, reprimir a disidentes políticos o llevar a cabo una vigilancia generalizada e injustificada se puede describir, acertadamente, como una amenaza social y política. Incluso cuando se utiliza para fines legítimos, el reconocimiento facial ha sido criticado por promover y solidificar prejuicios

sociales, que perjudican desproporcionadamente a grupos minoritarios. Para explorar los problemas éticos que rodean al reconocimiento facial, este ensayo comienza por reconstruir la mejor justificación posible para su uso. Posteriormente, el ensayo explora algunas de las preocupaciones con respecto a la tecnología, sobre todo aquellas asociadas a los casos en que ésta se combina con otras tecnologías, como los sistemas de circuito cerrado de televisión o las redes sociales. Más concretamente, y para ordenar la discusión, las objeciones que se considerarán, en este orden, son tres: la objeción de la privacidad; la objeción de la pendiente resbaladiza; y la objeción de guerra. Las dos primeras son objeciones que pueden formularse contra la tecnología en cualquier contexto donde se despliegue la tecnología.

La objeción final aborda preocupaciones que surgen exclusivamente en circunstancias de guerra. Argumentaré que la suma de estas objeciones —de entre muchas otras posibles que no consideraré en este ensayo— hace que el uso de la tecnología de reconocimiento facial sea inaceptable.

La objeción de la privacidad

Acaso la preocupación más obvia y citada con respecto al reconocimiento facial es que vulnera la privacidad informativa. Al alimentarse de cantidades ingentes de datos personales (en concreto, imágenes) extraídas de sitios de Internet, generalmente sin el consentimiento de los usuarios, el reconocimiento facial entraña una forma de vigilancia masiva que viola la privacidad de dichos usuarios. La privacidad informativa garantiza a las personas controlar el acceso a su información, incluidas las imágenes personales, excluyendo a otras personas u organizaciones, como el Estado, de dicho acceso si así lo desean. La privacidad informativa es valiosa por estar estrechamente relacionada con el valor fundamental de la autonomía, que puede definirse, en términos generales, como

*Doctorado, U. de Michigan; especializado en Filosofía Política, interesado en Ética Aplicada; Investigador y miembro del SNI 1; Ha realizado estancias post-doctorales o de investigación/docencia en el extranjero.

la libertad de pensar y hacer lo que uno elija. Violentar la privacidad informativa es moralmente objetable porque, desde una perspectiva individual, puede interferir con los planes que las personas elaboran para sus vidas. Desde una perspectiva colectiva, puede socavar la democracia liberal al inhibir el deseo de los ciudadanos de manifestarse o expresarse en contra de las autoridades por temor a las represalias, cuya ejecución se vuelva menos costosa con herramientas como el reconocimiento facial (Smith y Miller 2022).

La objeción de la pendiente resbaladiza

Evan Selinger y Brenda Leong (2021) han desarrollado una crítica del reconocimiento facial a partir del conocido argumento de la pendiente resbaladiza (*slippery slope*). La objeción de la pendiente resbaladiza es un argumento en contra de una acción, A, que no es objetable en sí misma, pero cuya realización conducirá a la realización de una cadena de acciones que al final conducirá a la acción Z, que sí es objetable. Dicho de otro modo, en un argumento de pendiente resbaladiza, no es permitir una acción inocua o inmediata lo que nos preocupa, sino la posibilidad de que permitir esta acción conduzca a una serie de acciones indeseable.

Ahora bien, el argumento de la pendiente resbaladiza es falaz si no identifica ni los pasos causales que nos llevan de las acciones o escenarios inocuos e inobjetables a las acciones o escenarios moralmente indeseables ni la probabilidad de que se materialicen. En contraste, las versiones razonables del argumento de la pendiente resbaladiza especifican explícitamente los mecanismos que podrían conducir al deslizamiento de un paso a otro. Algunos de los mecanismos que pueden de-sencadenar la cadena causal que nos lleva de lo inobjetable a lo objetable son, por ejemplo, el de “reducción de costos” (adoptar una tecnología para llevar a cabo el propósito A, que es benigno, tendrá como resultado que se reduzcan los costos de perseguir el propósito Z, que es maligno), el del “cambio de actitudes” (adoptar una tecnología para llevar a cabo el propósito A, que es benigno, tendrá como resultado un cambio

de actitudes generalizado, el cual tendrá como resultado que se acepte su uso para el propósito Z, que es maligno) o el de la “creación del impulso político” (adoptar una tecnología para llevar a cabo el propósito A, que es benigno, detonará un movimiento político que llevará inevitablemente a la adopción del propósito Z, que es maligno).

"La privacidad informativa garantiza a las personas controlar el acceso a su información, incluidas las imágenes personales, excluyendo a otras personas u organizaciones, como el Estado, de dicho acceso si así lo desean".

En el caso del uso del reconocimiento facial, la preocupación es que alentar o tolerar su uso para fines permitidos en ciertos contextos legítimos, como identificar a sujetos que han cometido actos criminales, podría abrir la puerta a objetivos reprobables, como su despliegue para realizar acciones injustificadas de inteligencia en contra de ciudadanos inocentes o su uso por actores no estatales locales o transnacionales cuyos objetivos, aunque no necesariamente injustos, son opacos para nosotros. Me centraré en estos dos casos, aunque algunos autores prevén peores desenlaces para la tecnología de reconocimiento facial, como su incorporación a sistemas de decisiones automatizada, como los que se utilizan en las llamadas armas letales autónomas.

¿Cuáles son los mecanismos que podrían llevarnos por la pendiente resbaladiza en el

caso del reconocimiento facial? Según los detractores de la tecnología, las empresas que promueven el uso de tecnología de reconocimiento facial buscan normalizar su uso. Este objetivo es un ejemplo del mecanismo de modificación de actitudes, descrito antes. La normalización se alcanza mediante la representación positiva de la tecnología, al caracterizarla como una herramienta “divertida” que promueve la eficiencia. Al habituarse a los entornos divertidos y “eficientes” que crea, los ciudadanos se inclinarán con mayor facilidad a aceptar su uso en un número cada vez mayor de ámbitos de la vida social. Las cámaras de circuito cerrado de televisión, que regularmente operan en conjunto con el reconocimiento facial, son otro ejemplo de este tipo de habituación pernicioso: la pérdida duradera de privacidad que conllevan se percibe como menos problemática si se considera que son efectivas para prevenir el crimen, y con el paso del tiempo los ciudadanos habrán perdido de vista el dilema entre la privacidad y la efectividad contra el crimen.

Otro impulsor causal de la pendiente resbaladiza es la probable renuencia de las autoridades a retirar la tecnología una vez que se ha resuelto la problemática que le dio origen (por ejemplo, un repunte en la criminalidad) y, en cambio, su inclinación a utilizar la tecnología en un número cada vez mayor de actividades, incluyendo algunas ilegítimas, como vigilar a los ciudadanos sin justificación adecuada. La renuencia es particularmente preocupante en contextos de baja institucionalidad democrática. En su discusión sobre la recopilación de metadatos con fines de inteligencia, Michael Skerker (2016) arguye que, en los estados con un Estado de Derecho robusto y altos estándares éticos de servicio público, es probable que los riesgos de abuso de metadatos sean menores que los de un ataque terrorista o un ataque militar. Lo contrario es cierto cuando faltan estas cualidades institucionales. La misma lógica se aplica con respecto al reconocimiento facial: que los riesgos asociados con la tecnología se materialicen o no, y la medida en que se materialicen, depende de la presencia o ausencia de estas cualidades

institucionales. En muchas naciones donde el reconocimiento facial ha sido utilizado, las capacidades institucionales que podrían mantenerla bajo control están ausentes.

Podría argumentarse, para mitigar esta preocupación, que las herramientas de reconocimiento facial pueden implementarse con salvaguardas tecnológicas para eliminar o reducir el potencial de abuso gubernamental. Clearview AI, por ejemplo, puede imponer restricciones en el uso de sus servicios y así asegurarse de que sólo los funcionarios gubernamentales autorizados tengan acceso a ella. Dichas salvaguardas pueden ser relativamente simples, como la llamada autenticación de dos factores, de modo que, al menos, los funcionarios públicos de las ramas del gobierno que no tienen funciones de seguridad no tengan acceso a la tecnología. Además, dado que servicios de reconocimiento facial como los de Clearview AI funcionan a través de un servicio en la nube, la compañía conserva la capacidad de revocar el acceso en casos de abuso de la tecnología.

Por varias razones, estas soluciones tecnológicas a posibles abusos gubernamentales son inadecuadas. Como lo muestra el caso del Grupo NSO (creador del software espía Pegasus), las garantías corporativas son, en el mejor de los casos, frágiles. Más importante aún, confiar en este tipo de garantías corporativas simplemente desplaza la preocupación por las barreras contra el abuso del sector público al privado. La pregunta sería ahora si existen garantías institucionales para evitar que las corporaciones privadas reutilicen o vendan, con fines de lucro, los datos personales extraídos por la tecnología. Transferir la responsabilidad de frenar el abuso del sector público al privado podría incluso agravar los riesgos asociados con el reconocimiento facial. Por un lado, la capacidad de la sociedad para regular a los actores corporativos, en particular a las empresas de tecnología, puede ser menor que la capacidad para controlar a los agentes del gobierno, como lo muestra el auge de Google y Facebook (Zuboff 2019, Véliz 2020). Por otro lado, al dejar que los actores corporativos tomen decisiones sobre lo

que constituye un uso inadecuado de la tecnología, los ciudadanos terminarían otorgando a los actores corporativos, que carecen de legitimidad política, la autoridad para tomar decisiones que afectan sus vidas. Las corporaciones tecnológicas vendrían a establecer estándares comunes para la regulación digital, una tarea que debería ser un asunto determinado por decisiones colectiva. Se crearía, en otras palabras, un déficit de legitimidad (Benn y Lazar 2022).

La objeción de la guerra

La objeción al que sea quizás uno de los usos más controvertido del reconocimiento facial se refiere a hechos recientes, donde la tecnología ha sido utilizada, de manera inédita, para identificar los cuerpos de los soldados muertos y difundir sus imágenes en las redes sociales (Espindola s.a.). Esto fue lo que ocurrió en las etapas iniciales de la invasión rusa a Ucrania, cuando las autoridades ucranianas utilizaron los servicios de Clearview AI, cuya base de datos contiene millones de imágenes extraídas de distintos sitios de internet, para tratar de revelar la identidad de soldados rusos caídos en combates y diseminar su identidad a través de redes sociales.

Las autoridades ucranianas justificaron la estrategia alegando que contribuía al deber humanitario de devolver los cuerpos de los fallecidos a sus familiares en Rusia. Según este alegato, entregar los cuerpos de los difuntos a sus familiares les permitiría a éstos realizar un luto apropiado, evitando que experimenten lo que Pauline Boss llama “pérdida ambigua”. La pérdida ambigua se produce cuando la pérdida de un familiar o ser querido está rodeada de incertidumbre. La pérdida ambigua conduce a un luto no resuelto, congela el proceso de duelo y contribuye al deterioro de la salud mental de los dolientes. Estos factores no están presentes, o no en la misma magnitud, en casos que no involucran pérdida ambigua. La sintomatología asociada a la pérdida ambigua incluye sentimientos de soledad o de distanciamiento de otras personas e

incluso de la comunidad y un sentido disminuido de la propia identidad. Estos efectos socavan las capacidades (*capabilities*) de las personas en el sentido que Martha Nussbaum da a ese término (su capacidad para entablar relaciones sociales significativas y enmarcar una concepción de la buena vida), y por lo tanto erosiona sus derechos fundamentales. Así, en consonancia con estas ideas, si la tecnología de reconocimiento facial pudiese ayudar a las personas a superar la pérdida ambigua, acelerando o ayudando a identificar cadáveres, entonces sería fundamental para proteger los derechos fundamentales y, por lo tanto, estaría en principio justificada.

La moralidad del uso del reconocimiento facial con el propósito antes mencionado es sumamente cuestionable. Primero, la tecnología puede cometer errores. Tales errores no derivan necesariamente, como es común en contextos no bélicos, de sesgos raciales o étnicos incorporados al reconocimiento facial, con los correspondientes riesgos de generar discriminación injustificable, sino más bien del hecho de la descomposición *post mortem*, que puede afectar la eficacia del reconocimiento facial. Esta es una preocupación particularmente relevante en el contexto de la guerra, donde los cuerpos pueden estar sujetos no solo a descomposición avanzada sino también a desfiguración o mutilación. Cuando las identificaciones erróneas *post mortem* superan un umbral razonable, la gran cantidad de falsos positivos puede socavar las funciones humanitarias que se supone que el reconocimiento fácil debería facilitar.

Por otra parte, el mero hecho de difundir imágenes de soldados caídos, independientemente de que su identidad haya sido revelada mediante reconocimiento facial o no, podría interpretarse como una falta de respeto a la dignidad de los muertos, entrando así en tensión con las Convenciones de Ginebra. Es cierto que las Convenciones sólo hacen escasas referencias al tratamiento de los muertos, pero su orientación es cla-

ra. Una de sus disposiciones insta a las partes beligerantes a “proteger [a los muertos y heridos] contra . . . malos tratos” cuando “las consideraciones militares lo permitan”, y otra disposición pide a las partes que respeten “los restos de las personas que hayan muerto por motivos relacionados con la ocupación o que estén detenidas como resultado de la ocupación o las hostilidades”. Una interpretación amplia de estas disposiciones equipararía la difusión de imágenes de soldados muertos con un trato degradante en vista de la expectativa generalizada de que la exhibición pública de cadáveres con fines distintos a los ceremoniales es, con algunas excepciones, indigna. Esta es una posición controvertida porque invoca un derecho póstumo contra el trato indigno, que puede rebatirse aduciendo que los muertos no experimentan humillación, no son afectados por ella y, por lo tanto, no hay justificación para defender tal derecho. Incluso si ello fuera así, sin embargo, los familiares de los difuntos sí pueden verse humillados y afectados por el trato indigno que éstos reciban,

y es en virtud de tales daños que puede sustentarse el rechazo del reconocimiento facial en este ámbito.

Conclusión

En este ensayo he discutido algunas de las razones que ponen en entredicho la moralidad del uso del reconocimiento facial. A la luz de esta crítica, no es ninguna sorpresa que, en días recientes, el Parlamento Europeo, una de las principales ramas legislativas de la Unión Europea, haya aprobado un proyecto de ley conocido como *AI Act*, que impondrá nuevas restricciones a los usos más riesgosos de la inteligencia artificial, dentro de las cuales se incluye la utilización de software de reconocimiento facial. El Parlamento Europeo votó a favor de prohibir el uso del reconocimiento facial en tiempo real, aunque quedan dudas sobre si se deben permitir exenciones para la seguridad nacional y otros fines de aplicación de la ley. Otra disposición prohibiría a empresas como Clearview AI extraer datos biométricos de las redes sociales para crear bases de datos.

Bibliografía

- Almeida, Denise, Konstantin Shmarko, and Elizabeth Lomas, “The Ethics of Facial Recognition Technologies, Surveillance, and Accountability in an Age of Artificial Intelligence: A Comparative Analysis of US, EU, and UK Regulatory Frameworks,” *AI and Ethics* 2 (August 2022), pp. 377–87.
- Ashbridge, Sarah, “Digital Dignity in Death: Are the Geneva Conventions Fit for Purpose in the Age of Social Media?,” Royal United Services Institute, March 29, 2022, rusi.org/explore-our-research/publications/commentary/digital-dignity-death-are-geneva-conventions-fit-purpose-age-social-media.
- Benn, Claire and Seth Lazar, “What’s Wrong with Automated Influence,” *Canadian Journal of Philosophy* 52, no. 1 (January 2022), pp. 125–48.
- Boss, Pauline, *Ambiguous Loss: Learning to Live with Unresolved Grief* (Cambridge, Mass.: Harvard University Press, 1999).
- Espindola, Juan, “Facial Recognition in War Contexts: Mass Surveillance and Mass Atrocity,” *Ethics and International Affairs* (volumen no asignado).
- Nussbaum, Martha C., *Creating Capabilities: The Human Development Approach* (Cambridge, Mass.: Belknap, 2013).
- Selinger, Evan y Brenda Leong, “The Ethics of Facial Recognition Technology,” en Carissa Véliz (ed.), *The Oxford Handbook of Digital Ethics* (Oxford: Oxford University Press, 2021), p. 31.
- Skerker, Michael, “Moral Concerns with Cyberespionage: Automated Keyword Searches and Data Mining,” en Fritz Allhoff, Adam Henschke, and Bradley Jay Strawser (eds.), *Binary Bullets: The Ethics of Cyberwarfare* (Oxford: Oxford University Press, 2016).
- Smith, Marcus and Seumas Miller, “The Ethical Application of Biometric Facial Recognition Technology,” *AI & Society* 37 (March 2022), pp.167–75.
- Véliz, Carissa. *Privacy Is Power: Why and How You Should Take Back Control of Your Data* (London: Random House, 2020).
- Zuboff, Shoshana, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019).

Implicaciones éticas de la implementación de la inteligencia artificial en la práctica médica

Luis Muñoz Fernández*

Introducción

En el libro *Vida 3.0. Qué significa ser humano en la era de la inteligencia artificial*, su autor Max Tegmark seleccionó muy bien el título del primer capítulo: “Bienvenidos a la conversación más importante de nuestro tiempo” (Tegmark, 2018).

La inteligencia artificial (IA), sus expectativas, beneficios y riesgos, es un tema cotidiano en los medios de comunicación desde finales de 2022, cuando fue lanzado ChatGPT, un sistema de respuestas automáticas (*chatbot*) basado en la inteligencia artificial cuya capacidad para elaborar textos ha asombrado a sus usuarios y ha generado también numerosas controversias (<https://es.wikipedia.org/wiki/ChatGPT>).

Así como no existe una definición universal de la inteligencia humana, tampoco la hay para la IA. Margaret A. Boden, profesora investigadora de Ciencias Cognitivas de la Universidad de Sussex, Inglaterra, señala lo siguiente:

“La inteligencia no es una dimensión única, sino un espacio profusamente estructurado de capacidades diversas para procesar la información. Del mismo modo, la IA utiliza muchas técnicas diferentes para resolver una gran variedad de tareas” (Boden, 2017).

El Consejo de Europa ha preparado un glosario sobre la IA, a la que define como “un conjunto de ciencias, teorías y técnicas cuyo fin es reproducir mediante una máquina las capacidades cognitivas de un ser humano” (<https://www.coe.int/en/web/artificial-intelligence/glossary>).

En un tema como este, que crece y cambia con tanta rapidez, es difícil asimilar toda su extensión y reflexionar detenidamente sobre sus implicaciones éticas. Una primera aproximación útil es la *Recomendación sobre la ética de la inteligencia artificial* de la Unesco, adoptada por este organismo el 23 de noviembre de 2021, en el que reconoce que “Considerando las tecnologías de la IA

pueden ser de gran utilidad para la humanidad y que todos los países pueden beneficiarse de ellas, pero que también suscitan preocupaciones éticas fundamentales, por ejemplo, en relación con los sesgos que pueden incorporar y exacerbar, lo que puede llegar a provocar discriminación, desigualdad, brechas digitales y exclusión y suponer una amenaza para la diversidad cultural, social y biológica, así como generar divisiones sociales o económicas; la necesidad de transparencia e inteligibilidad del funcionamiento de los algoritmos y los datos con los que han sido entrenados; y su posible impacto en, entre otros, la dignidad humana, los derechos humanos y las libertades fundamentales, la igualdad de género, la democracia, los procesos sociales, económicos, políticos y culturales, las prácticas científicas y de ingeniería, el bienestar animal y el medio ambiente y los ecosistemas” (Unesco, 2021).

Asimismo, el documento de la Unesco describe las repercusiones éticas en los ámbitos de aplicación de la IA que le competen y que son el de la educación, la identidad y diversidad culturales, la comunicación y la información y también la ciencia. Sobre esta última dice lo siguiente:

“La ciencia, en el sentido más amplio, que abarca todos los ámbitos académicos desde las ciencias exactas y naturales y las ciencias médicas hasta las ciencias sociales y humanas, ya que las tecnologías de la IA aportan nuevas capacidades y enfoques de investigación, influyen en nuestra concepción de la comprensión y la explicación científicas y crean una nueva base para la adopción de decisiones” (Unesco, 2021). Este último punto relativo a las decisiones diagnósticas y terapéuticas es especialmente relevante en la práctica médica.

Aplicaciones de la IA en la práctica médica

En marzo de 2023, la prestigiosa revista médica *The New England Journal of Medicine*

*Médico cirujano; Especialista en Anatomía Patológica; Máster en Bioética y Derecho; Miembro del Colegio de Bioética A.C.; Miembro del Consejo Consultivo de la CONBIOÉTICA.

publicó el primero de una serie de editoriales y artículos bajo el título *Artificial Intelligence in Medicine (Inteligencia artificial en Medicina)*, en el que se lee lo siguiente:

“Recientemente, la inteligencia artificial (IA) ha adquirido relevancia pública con el lanzamiento de modelos de aprendizaje profundo (*deep-learning models*) que generan todo tipo de contenidos con una intervención humana mínima, desde arte hasta ensayos de fin de curso. Este desarrollo ha revitalizado del papel actual y potencial de la IA en todos los aspectos de la vida. Sin embargo, de todos los campos en los que puede aplicarse la IA, la medicina destaca tanto por su gran potencial como por los retos que implica” (Beam, A.L. y cols., 2023).

De ahí que la revista decidiese tratar este tema, primero con la serie de textos a los que nos hemos referido, pero, además, publicando a partir del 2024 una revista nueva, *NEJM AI* (ai.nejm.org). Esto nos da una idea de la importancia que una revista como esta le da al papel de la IA en la Medicina.

En este editorial se señala que en la actualidad no hay prácticamente ninguna área de la Medicina y del cuidado de la salud que no haya sido tocado por la IA:

“Por ejemplo, las aplicaciones dirigidas por IA para el dictado y la captura de notas médicas; muchas de estas aplicaciones están intentando sintetizar las entrevistas a los pacientes y los resultados de laboratorio para escribir directamente las notas sin que intervenga el médico clínico. [...] Ya hemos visto muchos artículos que informan sobre el uso de la IA en la interpretación de imágenes radiográficas, histopatológicas y oftalmológicas (fondo de ojo). Se ha incrementado el uso de instrumentos con IA que analizan e interpretan grandes bases de datos de investigación que contienen información que va desde los datos de laboratorio hasta los datos clínicos” (Beam, A.L. y cols., 2023).

A partir de este nuevo contexto de la práctica médica, el editorial no oculta su preo-

cupación y señala que todos estos métodos basados en la IA no son necesariamente la panacea. Expone que son frágiles, que su campo de acción es reducido y que pueden haber sido contruidos con sesgos provenientes de sus desarrolladores, de modo que pueden llegar a afectar de manera negativa a grupos marginados. Sobre este último punto, uno de los ejemplos fuera de la medicina más citados es el algoritmo COMPAS, un sistema policial basado en IA que predijo erróneamente que la mayoría de los delincuentes que iban a reincidir eran primordialmente personas de piel negra (Coeckelbergh, 2021).

Implicaciones éticas de la IA

La Fundación Víctor Grífols i Lucas, un referente de la bioética española, publica periódicamente cuadernos sobre diversos temas relacionados con esta disciplina. En 2023 publicó su cuaderno número 63, titulado *Inteligencia artificial en Salud. Retos éticos y científicos*, en el que Pere Ibern, miembro del Patronato de la Fundación e investigador del Centro de Investigación en Economía y Salud de la Universidad Pompeu Fabra de Barcelona, presenta en el cuaderno citado la respuesta que dio el famoso algoritmo de inteligencia artificial ChatGPT-4 (versión del 23 de marzo de 2023) cuando se le preguntó sobre la propia IA que, según Bill Gates, es “la tecnología más revolucionaria desde hace décadas” (Ibern. Fundación Víctor Grífols i Lucas, 2023).

La respuesta de Chat GPT-4 es coherente con lo que se ha publicado sobre este tema, lo que es completamente lógico, pues el algoritmo ha sido entrenado en base a la amplia información disponible, pero, además, señala también los problemas bioéticos de la IA que deben abordarse y que a continuación se transcriben:

1. *Privacidad y seguridad de los datos: la IA implica la recopilación, el almacenamiento y el uso de datos masivos. Es importante garantizar la privacidad y la seguridad de estos datos para evitar*

el robo de identidad, el ciberataque y la vulneración de la privacidad.

- 2. Discriminación: la IA puede ser programada para tomar decisiones que pueden ser discriminatorias por razones de género, raza, edad, orientación sexual, etc. Es importante asegurar que la IA sea imparcial y que no reproduzca prejuicios existentes en la sociedad.*
- 3. Transparencia y explicabilidad: la IA puede tomar decisiones sin dejar claro cómo se ha llegado a ellas. Esto puede provocar que sea difícil comprender por qué se toman determinadas decisiones y dificultar la responsabilidad en caso de error.*
- 4. Responsabilidad: la IA puede tomar decisiones que tienen consecuencias importantes para las personas y la sociedad en general. Es importante establecer quién es responsable de las decisiones que toma la IA y cómo se puede responsabilizar a los desarrolladores o a los propietarios de esta tecnología.*
- 5. Autonomía: la IA puede llegar a ser autónoma en la toma de decisiones, lo que puede ser un problema cuando las decisiones que toma afectan a la vida de las personas. En este sentido, es importante establecer límites a la IA y asegurar que siempre se dé una supervisión humana.*
- 6. Ética de los vehículos autónomos: la IA está siendo utilizada para desarrollar vehículos autónomos. Es importante garantizar que estos vehículos se comporten de manera ética en situaciones límite y que no pongan en peligro la vida de las personas.*

A continuación, ChatGPT-4 enlista los criterios que se pueden tener en cuenta para que una IA sea explicable, es decir, que siga el principio de explicabilidad: asegurar la transparencia y responsabilidad en las decisiones tomadas por las Tecnologías de la Información y la Comunicación que afectan a las personas.

- 1. Transparencia: la IA debe ser transparente en su funcionamiento, de manera que se pueda entender cómo se toman las decisiones. Esto incluye la capacidad de mostrar los datos utilizados para entrenar el modelo, así como los algoritmos y las reglas que utiliza para tomar decisiones.*
- 2. Interpretabilidad: la IA debe ser interpretada por las personas, de manera que se pueda comprender lo que está haciendo y por qué está tomando las decisiones que toma. Esto implica una interacción humana y la capacidad de proporcionar explicaciones claras y comprensibles para los usuarios.*
- 3. Consistencia: la IA debe ser consistente en su comportamiento, de forma que pueda ser confiable y previsible en su conducta. Esto implica la capacidad de proporcionar respuestas coherentes a situaciones similares.*
- 4. Robustez: la IA debe ser robusta, es decir, debe ser capaz de manejar situaciones inesperadas o excepcionales sin caer en errores graves. Esto incluye la capacidad de gestionar datos no previstos o situaciones no programadas.*
- 5. Accesibilidad: la IA debe ser accesible para todos los usuarios, independientemente de su nivel de conocimiento tecnológico. Esto implica la capacidad de proporcionar una interfaz de usuario clara e intuitiva, así como la capacidad de dar apoyo en diferentes idiomas y formatos.*

En resumen, ChatGPT-4 señala que para conseguir que una IA sea explicable, es importante que sea transparente, interpretativa, consistente, robusta y accesible para todos. Esto ayudará a garantizar que la IA sea ética y responsable en su funcionamiento.

El principio de explicabilidad al que hace alusión se relaciona con otros principios bioé-

ticos como el respeto a la privacidad, la seguridad y la libertad de las personas (Ibern. Fundación Víctor Grífols i Lucas, 2023).

No está de más volver a decir que tanto la lista de los problemas bioéticos de la IA, como los criterios para que la IA sea explicable fueron elaborados por ChatGPT-4, no por un ser humano. Eso da una idea del alcance y potencial de estas tecnologías.

Alicia de Manuel, investigadora del Observatorio de Ética en Inteligencia Artificial de Catalunya, es una de las coautoras del Cuaderno de la Fundación Víctor Grífols i Lucas al que estamos haciendo referencia y ha escrito un capítulo el que describe los principios éticos de la IA a través del concepto de la IA confiable, con siete puntos clave para que los sistemas de IA puedan ser considerados confiables:

1. *Agencia humana y supervisión: los sistemas deben empoderar a los seres humanos, así como garantizar los mecanismos de supervisión adecuados.*
2. *Robustez técnica y seguridad: los sistemas deben ser seguros, precisos y fiables y asegurar planes alternativos en caso de error para minimizar y prevenir los daños no intencionales.*
3. *Privacidad y gobernanza de datos: además de garantizar la privacidad y la protección de datos, los sistemas de IA deben tener en cuenta la calidad e integridad de los datos asegurando un acceso legitimado a ellos.*
4. *Transparencia: los modelos de negocio de datos y sistemas de IA deben ser transparentes.*
5. *Diversidad, no discriminación y equidad: en los sistemas de IA deben evitarse los prejuicios y sesgos injustos.*
6. *Bienestar social y medioambiental: los sistemas de IA tienen que ser sostenibles y respetuosos con el medioambien-*

te, ya que deben poder beneficiar a los seres vivos y las generaciones futuras.

7. *Rendición de cuentas: deben establecerse mecanismos que garanticen la auditabilidad y la rendición de cuentas tanto de los sistemas de IA como de sus resultados* (de Manuel. Fundación Víctor Grífols i Lucas, 2023).

Implicaciones éticas de la IA en la Medicina y las Ciencias de la Salud

La editorial de *The New England Journal of Medicine* antes citada señala que la pericia en el campo de la IA y el aprendizaje de máquinas (*machine learning*) están vinculados estrechamente a las aplicaciones comerciales, que se trata de una tecnología cambiante y que, en muchos casos, la producen compañías e investigadores con intereses financieros en sus productos. Dada esta situación, las características operativas de estos modelos de IA están solamente en manos de sus desarrolladores, impidiendo el acceso a esta información a muchos otros interesados (Beam, A.L. y cols., 2023).

Todo ello hace pensar que, pese al gran potencial de estas herramientas en la práctica médica y, en general, en todas las ciencias de la salud, si deseamos que presten un verdadero servicio todos los seres humanos que lo requieran, independientemente de su raza, nivel socioeconómico y características socioculturales, es indispensable que su desarrollo se lleve a cabo con apego a normas éticas y bajo un marco regulatorio internacional vinculante, tal como lo señala el pronunciamiento de la Comisión Nacional de Bioética publicado en 13 de junio de 2023 (Comisión Nacional de Bioética, 2023).

En este mismo documento, la Comisión Nacional de Bioética define las pautas necesarias para fortalecer el marco regulatorio a nivel federal:

- *A fin de constituirse como un quehacer con responsabilidad social, el desarrollo de tecnologías de IA en salud evitará el*

aumento en las brechas de inequidad social y económica entre poblaciones, además de no interferir en procesos de participación democrática.

- En el desarrollo de tecnologías de IA en salud constituye un imperativo bioético la protección de los derechos de propiedad, integridad y privacidad de la información personal, especialmente cuando se trate de información genómica.
- El uso de la IA no diluye la responsabilidad de los profesionales de salud, quienes son responsables en última instancia por operación, seguimiento y tecnovigilancia de las aplicaciones de IA.
- La relación médico-paciente es una entre dos personas humanas; en este sentido, el uso de la IA solamente debería constituir un mecanismo de mediación, pero que no implique obstáculo alguno en la relación entre las personas usuarias y el equipo de salud.
- Considerando que afecta a un gran número de personas, se requiere 1) generar condiciones para un diálogo horizontal entre los desarrolladores de algoritmos, tomadores de decisiones, instancias de la sociedad civil con representación de todos los sectores de la sociedad, especialmente las poblaciones en condiciones de vulnerabilidad, considerando estrategias y programas de capacitación y divulgación, a fin de generar entendimiento común sobre la amplia gama de riesgos, impactos socioeconómicos y oportunidades inherentes al diseño, desarrollo e implementación de tecnologías de IA en el cuidado de la salud; 2) establecer medidas adicionales para asegurar el acceso equitativo a los beneficios de las tecnologías de IA; 3) promover estándares internacionales de seguridad, eficacia y eficiencia en el desarrollo de aplicaciones de IA –como requeriría cualquier dispositivo médico–, como

también 4) asegurar la participación de profesionales de la salud en el desarrollo de aplicaciones de IA en salud.

- Ante la posibilidad de daño a la integridad física, social y moral de las personas, el desarrollo de tecnologías de IA requiere a) incluir preceptos éticos y estándares de derechos humanos -a la manera de una gramática moral artificial- directamente en la configuración de los algoritmos de procesamiento, lo cual considerará b) la participación de los comités de ética en investigación (CEI), además de c) un mecanismo permanente de monitoreo a lo largo de su implementación.
- Considerando los casos en los que se deriven consecuencias negativas de la aplicación de la IA, debe haber una cadena de responsabilidad clara.
- En el uso de aplicaciones de IA en salud se contará con la supervisión de un profesional sanitario certificado, quien será responsable por su adecuado funcionamiento.
- En cuanto al marco normativo, se requiere de una revisión permanente a fin de mantener actualizadas las disposiciones relativas a la responsabilidad de los desarrolladores en el diseño, implementación y mantenimiento de los sistemas de IA en salud.
- Específicamente en los procesos de machine learning, es insoslayable establecer un mecanismo de curaduría de los datos sometidos a fin de asegurar su calidad.
- En el proceso de desarrollo de aplicaciones de IA es necesario evitar en la medida de lo posible el uso de algoritmos de caja negra (black box), o bien establecer pautas para hacer más inteligible su procesamiento.
- A fin de evitar la explotación de comunidades desprotegidas, se debe promo-

ver el acceso equitativo a las tecnologías para desarrollo de la IA, lo cual incluye las bases de datos que se empleen para entrenar los algoritmos de las diversas aplicaciones.

- Se promoverá la alfabetización en IA entre los planes de estudios de las instituciones de educación superior, con énfasis en los principios bioéticos de la IA y los usos responsables de las aplicaciones de IA.
- Considerando el enorme gasto energético que implican ciertas tecnologías de IA, los desarrolladores buscarán reducir la huella ecológica de las mismas (Comisión Nacional de Bioética, 2023).

Comentarios finales

Para el doctor Abraham Verghese, médico y escritor, “las dos culturas –la computadora y el médico– debe trabajar juntas. [...] Un modelo de la caja negra puede ayudar a que

los médicos tomen buenas decisiones, pero sólo si mantienen a la inteligencia humana en el circuito, incluyendo los contextos social, clínico y personal de cada caso. Además, sólo el cerebro humano con adiestramiento clínico puede generar nuevas ideas, vislumbrar nuevas aplicaciones y uso de la IA y el aprendizaje de máquinas y conectar estas tecnologías con las humanidades y las ciencias sociales en formas que las actuales computadoras no pueden hacer” (por lo menos no todavía).

Y agrega: “En el cuidado del enfermo, hay una función clave que los médicos llevan a cabo, a la Tinsley Harrison llama ‘la función sacerdotal del médico’. La inteligencia humana trabajando con la inteligencia artificial –un médico empático y bien informado, armado con buenas herramientas predictivas y sin la carga del trabajo burocrático– pueden hacer que los médicos cumplan la máxima de Peabody que decía que el secreto del cuidado del paciente es hacerse cargo de él”.

Referencias

- Beam, A.L., Drazen J. M., Kohane, I.S., y cols. (2023). Artificial Intelligence in Medicine. *N Engl J Med*, 38 (13), 1220-1221.
- Boden, M. A. (2017). *Inteligencia artificial*. Turner Noema.
- Coeckelbergh, M. (2021). Ética de la inteligencia artificial. Cátedra.
- Comisión Nacional de Bioética. (2023). *Bioética de la inteligencia artificial en salud*. Disponible en: https://www.gob.mx/cms/uploads/attachment/file/832082/Bio_tica_de_la_inteligencia_artificial_Junio2023.pdf
- Consejo de Europa. *Glosario*. Disponible en: <https://www.coe.int/en/web/artificial-intelligence/glossary>. Consultado el 18 de junio de 2023.
- de Manuel, A. (2023). Inteligencia artificial ética y su aplicación en el ámbito de la salud sistemas aplicados a la COVID-19 en Inteligencia artificial en salud. Retos éticos y científicos. *Cuadernos de la Fundación Víctor Grífols i Lucas*, 63, 46-59.
- Ibern, P. (2023). Presentación en Inteligencia artificial en salud. Retos éticos y científicos. *Cuadernos de la Fundación Víctor Grífols i Lucas*, 63, 8-11.
- Tegmark, M. (2018). *Vida 3.0. Qué significa ser humano en la era de la inteligencia artificial*. Taurus.
- Unesco. (2021). *Recomendación sobre la ética de la inteligencia artificial*. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa.
- Verghese A., Shah, N.H., Harrington R.A. (2017). What this computer needs is a physician. *Humanism and Artificial Intelligence*. *JAMA*, 319 (1), 19-20.
- Wikipedia. *ChatGPT*. Disponible en: <https://es.wikipedia.org/wiki/ChatGPT>. Consultado el 18 de junio de 2023.

Implicaciones bioéticas en la IA

Luis Ángel Lara Pereda*

Introducción

Nuestra vida cotidiana está permeada por diferentes aplicaciones que utilizan inteligencia artificial. Desde hace algunos años, en la cultura popular la Inteligencia Artificial ha adquirido una imagen que no le es muy favorable. Generalmente se le asocia con eventos de corte apocalíptico, al mero estilo de una rebelión de las máquinas *versus* los seres humanos. Las más recientes noticias en torno a la Inteligencia Artificial parecen reavivar ese viejo temor frente a ella. Sin embargo, si atendemos a la literatura científica disponible, parece que estamos muy lejos de esos escenarios catastrofistas que hemos visto en las series y películas. No obstante, esto no implica que la IA esté exenta de algunos riesgos. En este sentido, el objetivo de esta reflexión es ponderar los riesgos efectivos, o reales, que implica la IA desde una perspectiva bioética más allá de las visiones apocalípticas que ya permean en el imaginario colectivo.

Inteligencia Artificial: entre el sueño y la realidad

Máquinas que se revelan cuyo pensamiento es superior al de los seres humanos, es la imagen de la IA que permea el imaginario colectivo. El temor a ser esclavizados por los robots o ser sustituidos por ellos en todas las tareas que hasta hoy realizamos los seres humanos han sido alimentados por la literatura y el cine de ciencia ficción. En la actualidad no hay nada más alejado de la realidad que esa imagen. Definir la IA es un ejercicio complicado, ya que ni siquiera los especialistas en el área han llegado a un consenso sobre lo que se debe entender por ella. De acuerdo con la Oficina Ejecutiva de la Presidencia de los Estados Unidos (2016), no hay una definición unívoca de lo que es la IA. No obstante, la variedad de definiciones existentes tiene puntos en común, los cuales nos permiten ofrecer una definición operativa en la cual enmarcar el análisis bioético en torno a esta tecnociencia. Las concepciones más aceptadas, según el texto de la Oficina Ejecutiva, señalan que la IA es un sistema capaz de resolver racionalmente un problema complejo, o alcanzar una determinada meta tal como lo haría un ser humano. (p. 6)

De esta manera, estamos en posibilidad de afirmar que la IA, a través de la programación de los algoritmos, constituye una inteligencia alterna a la del ser humano. Como tal, su meta es resolver problemas o llevar a cabo tareas que requieren el procesamiento de información. En conversación con una de las IA más recientes, Chat GPT, obtenemos que la IA es “la ciencia y la ingeniería de hacer máquinas inteligentes, especialmente programas informáticos. En términos simples, se refiere a sistemas o máquinas que imitan la inteligencia humana para realizar tareas y pueden mejorar iterativamente a partir de la información que recopilan” (Open AI, 2023).

Uno de los puntos problemáticos para definir a ciencia cierta lo que es la IA es el concepto de inteligencia. El nombre mismo de la tecnología parece indicar que es, o una extensión o algo contrapuesto a algo que se llamaría inteligencia natural. ¿Qué entendemos entonces por inteligencia? Si partimos de la definición ofrecida por la Oficina Ejecutiva Presidencia, o por la misma aplicación de IA, parece que la inteligencia se caracteriza por ser la capacidad de resolver problemas, procesar datos o crear estrategias para obtener ciertos fines y prever determinadas consecuencias. Si este es el caso, como señala Ramón López (2017) el ser humano no es el único ser vivo que posee inteligencia, evidentemente hay otros seres sintientes, como los animales de compañía, que pueden clasificar como seres inteligentes.

En este sentido, la IA estaría partiendo de una especie de imitación de una inteligencia orgánica. De esta forma, la mimesis de ciertas capacidades de los seres vivos, como la visión, el razonamiento, el lenguaje y el aprendizaje serían parte de los componentes esenciales de la IA, aunado al asunto de la autoconsciencia, el cual se supone el culmen de la inteligencia y el objetivo principal de este desarrollo tecnocientífico.

*Candidato a doctor en Filosofía de la Ciencia, UNAM; Grado de maestro con tesis sobre implicaciones bioéticas de la biología sintética; Líneas de investigación: Bioética, Ética y Filosofía de la Tecnología; Profesor.

Si bien, la IA es algo que ha estado en boga recientemente gracias a las plataformas que hacen uso de esta tecnología para crear textos o imágenes, no es una tecnología nueva. Su surgimiento puede datarse de mediados de la década del siglo pasado. En principio, los científicos encargados de crear y programar los algoritmos utilizados en las primeras variantes de la IA, creían que no sería tan complejo tener súper máquinas que pudieran superar, o al menos igualar, la inteligencia del ser humano. No obstante, con el paso del tiempo, el desconocimiento de cómo funciona de facto el cerebro humano y otras dificultades y problemas filosóficos de fondo han mostrado que el desarrollo de una IA similar o superior a la del ser humano es algo sumamente complejo. En esta línea argumentativa, Margaret Boden afirma que uno de los grandes aportes del desarrollo de la IA radica en que “nos ha enseñado que la mente humana es enormemente más rica y más sutil de lo que los psicólogos se imaginaban en un principio. De hecho, esta es la lección más importante que hay que aprender de la IA”. (2022, p. 48).

Durante su desarrollo la IA se ha diversificado. Existen dos grandes tipos de inteligencia artificial, los cuales son aplicados en diferentes aspectos de nuestra realidad. Se trata de las llamadas Inteligencia Artificial Débil (IAD) y de la Inteligencia Artificial Fuerte (IAF). La IAD es aquella que se encarga de realizar ciertas tareas específicas que realizamos los seres humanos, es decir, son una imitación de una mente humana, por ello, pueden jugar ajedrez, buscar información, crear o analizar imágenes

y procesar enormes cantidades de información. Si bien, indudablemente lo hacen de manera más eficaz y eficiente que varios seres humanos, no pueden llevar a cabo una tarea distinta para la que fueron programadas, es decir, mientras que una persona puede jugar ajedrez y a su vez ejecutar una receta de cocina, una IA, como *Deep Blue*, sólo puede derrotar a Kaspárov, pero es incapaz de realizar una receta de cocina además de derrotarlo. La IAD es la más extendida hoy en día, y es utilizada en diferentes sectores y aspectos de nuestra vida cotidiana, como en las aplicaciones de redes sociales que utilizamos diariamente, o los denominados asistentes personales.

Por su lado, la IAF es el sueño de los científicos. Este tipo de inteligencia sería capaz de llevar a cabo varias tareas de una manera similar o superior a lo que lo haría un ser humano. Según Ramón López (2022) una IAF no sería la imitación de una mente, sino que por ella misma sería una mente. En este sentido, la IAF tendría no sólo la capacidad de mejora y aprendizaje, sino que también sería capaz de desarrollar autoconsciencia. Este tipo de inteligencia artificial es el que vemos representado en películas como *Terminator* o *ExMachina* o en series como *Black Mirror*. Si bien, este tipo de IA es la que ha despertado el temor de la población debido a su capacidad de autoconsciencia, de acuerdo con la Oficina Ejecutiva de la Presidencia (2016), hoy en día no hay bases científicas sólidas que nos permitan afirmar que este tipo de IA está próxima a ser desarrollada, al menos durante las siguientes décadas. (p.7)

Cuadro comparativo sobre las diferencias de las IA's

IA débil	IA fuerte
Diseñada para realizar tareas específicas y predictibles.	Capaz de aprender y adaptarse a nuevas situaciones.
Programada para seguir instrucciones. Sólo puede llevar a cabo una tarea.	Su programación le otorgaría la capacidad de razonar y tomar decisiones.
Algunos ejemplos son los asistentes virtuales, Chat GPT, videojuegos, automóviles autónomos.	No se ha desarrollado una IA fuerte en la actualidad.

El hecho de que la llamada IA fuerte no exista aún tiene sus raíces en el hecho de que no tenemos del todo la comprensión de eso que llamamos conocimiento. De todo lo que conocemos, existen cosas que no sabemos cómo las sabemos, por ejemplo, el manejo de una bicicleta y las maniobras que se deben hacer para mantenerse en equilibrio, es decir, no hay un manual con reglas escritas que nos diga cómo mantener el equilibrio mientras pedaleamos, simplemente aprendemos a andar en bici andando en bici. De acuerdo con el físico y filósofo Michael Polanyi (1983) sabemos más de lo que podemos decir. Esto implica que todo el conocimiento que tenemos es imposible de ser transmitido o codificado en algún medio para ser transferido a una máquina. Como consecuencia de ello, la IAF pensada como una mente artificial, capaz de tomar decisiones, aprender, reflexionar y demás acciones propias del ser humano no ha sido posible de ser elaborada. La IA existente parte de la programación del conocimiento explícito, es decir, de aquellos saberes que sí podemos transmitir y codificar en manuales o dispositivos electrónicos.

De lo que sí debemos preocuparnos: las aplicaciones de la IA débil

Los escenarios donde los seres humanos son sustituidos por máquinas de IA aún están lejanos. Sin embargo, eso no implica que en nuestras actividades cotidianas no tratemos con la IA. Cuando vemos una película en *streaming*, hacemos compras en línea, o usamos nuestro *smartphone*, nos beneficiamos de distintas aplicaciones que hacen uso de la IA para ofrecernos una experiencia más amena en sus sitios o plataformas. La IA débil es muy utilizada en diversos sectores, Coeckelbergh (2022) señala que está presente en cosas tan comunes como el marketing, redes sociales, motores de búsqueda, aplicaciones de transporte, vehículos autónomos, hasta en los servicios sanitarios, seguridad o de impartición de justicia. Justo, porque la IA está en aplicaciones tan comunes es que se vuelve urgente reflexionar en torno al uso de esta tecnología y el acceso a

la misma. Siguiendo a Coeckelbergh, es necesario señalar que el uso de IA en ámbitos como el sector sanitario o de impartición de justicia puede ampliar las brechas de desigualdad ya existentes en el acceso a los servicios de salud o de justicia, y en esta última, puede reproducir estereotipos que afecten a poblaciones vulnerables en la impartición de justicia.

"Uno de los puntos problemáticos para definir a ciencia cierta lo que es la IA es el concepto de inteligencia. El nombre mismo de la tecnología parece indicar que es, o una extensión o algo contrapuesto a algo que se llamaría inteligencia natural".

Las variadas aplicaciones de la IA pueden ser muy útiles. Con su apoyo los médicos pueden lograr diagnósticos más precisos y oportunos, por ejemplo, con ella es posible interpretar mejor los estudios de imagenología o hacer cirugías a distancia. En otros casos, por ejemplo, en EUA, la IA ya se ha utilizado como herramienta de apoyo para tomar decisiones judiciales. Esto nos lleva a preguntarnos ¿quiénes deberían acceder a la tecnología de IA?, ¿es éticamente correcto que una IA tome una decisión judicial? Más aún, debemos partir del hecho de que las IA trabajan con una enorme cantidad de datos que procesan, ¿cómo adquieren esos datos?, ¿qué hacen con ellos?

Estas interrogantes nos llevan a pensar en determinados problemas éticos. En principio, debe pensarse en el asunto de la transparencia. Como señala Coeckelbergh (2022), nuestras identidades pueden estar en riesgo, toda nuestra información personal es manejada por los algoritmos de la IA que utilizan las empresas que nos proveen servicios todos los días. Ninguno de nosotros sabemos dónde van a parar esos datos, quiénes y para qué tienen acceso a ellos. En muchas ocasiones, utilizamos aplicaciones en nuestros móviles sin leer los términos y condiciones, en donde muchas veces otorgamos permiso para el manejo de nuestros datos como mercancía a intercambiar entre diversas empresas.

Por otro lado, las aplicaciones de la IA en otros ámbitos propician otras interrogantes éticas. Si pensamos en los vehículos autónomos, debemos cuestionarnos ¿por qué el vehículo debe tomar una decisión y no otra?, ¿bajo qué parámetros éticos lo hará?, ¿qué enfoque será mejor para ello? Uno de los problemas más profundos que surgen aquí va de la mano con ciertos valores epistemológicos que influyen en la ética. En muchas ocasiones, los seres humanos tomamos decisiones con base en lo que creemos. Las creencias conforman un plexo de valores éticos que orientan nuestras decisiones. Estas creencias no las adquirimos solamente por los libros, los manuales o las enseñanzas escolares, sino, sobre todo, por las experiencias vitales. Como señala Ramón López (2017), esos saberes de sentido común conforman nuestra inteligencia y son imposibles de transcribir en codificación alguna para ser insertada en una máquina. (p. 36)

La IA también trastoca valores epistemológicos, lo cual puede propiciar problemas que afecten a la toma de decisiones en bioética. El uso de plataformas que usan IA para generar imágenes o textos puede generar un problema en relación con lo que consideramos como verdadero. Si partimos del hecho de que en la bioética se deben tomar decisiones con base en la información

disponible, el hecho de no tener fiabilidad en la producción de la información puede generar sesgos en la toma de decisiones. Si bien es cierto, ya existen problemas con la integridad académica asociados con prácticas poco éticas como el plagio, este tipo de prácticas pueden agravarse con un uso inadecuado de las aplicaciones de IA.

Por último, es importante mencionar que el uso de la IA también conlleva problemas ambientales. Si bien, en apariencia la IA no tiene un respaldo material, deben considerarse los factores materiales que hacen posible la existencia de la IA, como el consumo de energía eléctrica y la infraestructura que se encuentra detrás de la producción de la misma. De acuerdo con Pengfei Li y demás (2023) el uso de Chat GPT tiene una huella ecológica importante, sobre todo en el caso del uso de recursos hídricos. De acuerdo con su investigación, cada consulta que realizamos en esta plataforma implica un consumo de aproximadamente 500 mililitros de agua, lo cual se traduce en casi 4 millones de litros anuales. Para darnos una idea de la cantidad de agua que consume esta plataforma, podemos mencionar que, según estimaciones de la ONU (2021) en promedio un habitante de la Ciudad de México consume alrededor de 366 litros de agua por día, en este caso la plataforma Chat GPT consume en un año el equivalente al consumo de 28 habitantes de la Ciudad de México.

Como es posible apreciar, el uso de IA no está exento de problemas éticos. De acuerdo con Onora O'Neil (2009), asumimos que la bioética es un cruce interdisciplinario en el que convergen diversos saberes que tienen una preocupación en común; la vida. Así, saberes como el derecho, la medicina y la filosofía procuran dar soluciones viables a los problemas que surgen con las recientes implementaciones tecnocientíficas. En este sentido, la IA puede y debe ser analizada desde una perspectiva multidisciplinaria, propia de la bioética, en función de los problemas que ya se manifiestan con su uso. Por ello, es necesario pensar desde la bioética formas

de regular los usos de la IA. No se trata de condenar a esta tecnología al destierro, sería ilógico, cuando menos. Sin embargo, dado los usos que tiene y los potenciales daños que se observan sí debe gestionarse y regularse su uso de manera tal que los grupos vulnerables, no se vean más afectados por sus aplicaciones y el acceso a este tipo de implementaciones no genere más brechas de desigualdad. En este sentido, debe pensarse en regulaciones que tengan en cuenta la dignidad de las personas, de forma tal que no se vea socavada con los usos de este tipo de tecnociencia.

Conclusiones

Sin duda alguna la Inteligencia artificial nos ha facilitado varias actividades cotidianas. Incluso nos ha auxiliado en labores más importantes, como el sistema de salud o de seguridad. No obstante, su uso no está libre de algunas controversias axiológicas que pueden ser analizadas en términos bioéticos. Más allá de las consecuencias apocalípticas que conocemos sobre la IA, debemos pensar

en aquellas consecuencias que en términos reales son susceptibles de presentarse. De acuerdo con los estudios científicos recientes, hoy en día es imposible que una IA sustituya por completo a un ser humano, si bien puede hacer tareas específicas mejor de lo que humanos especializados lo hacen, la IA no tendrá en las décadas próxima un desarrollo que nos lleve a tal escenario.

Por otro lado, sí debemos pensar en las consecuencias que el uso de la IA existente conlleva. Por ejemplo, la transparencia en el uso y manejo de los datos que se procesan en las aplicaciones que utilizan algoritmos de IA, los sesgos que podrían presentarse en el sistema de impartición de justicia o de acceso a la salud. El análisis de esta compleja problemática se debe realizar desde una perspectiva bioética, ya que nos garantiza una pluralidad de enfoques que nos permitirá robustecer el análisis de los riesgos y de los beneficios de esta tecnociencia, así como la mejor manera de regularla y tener un uso óptimo de la misma.

Referencias

- Boden, M. (2022). *Inteligencia Artificial*. Turner.
- Coeckelbergh, M. (2022). *Ética de la inteligencia artificial*. Cátedra.
- Executive Office of the President. (2016). *Preparing for the future of Artificial Intelligence*. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
- Li, P., Yiang J., Islam, M., Ren, S. (2023) Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. <https://arxiv.org/pdf/2304.03271.pdf>
- López, R. Meseguer, P. (2017). *Inteligencia Artificial*. Los libros de la Catarata.
- O'Neil, O. (2009). *Autonomy and Trust in Bioethics*. Cambridge.
- ONU. (2021). Comprender las dimensiones del problema del agua. *ONU-Hábitat*. https://onuhabitat.org.mx/index.php/comprender-las-dimensiones-del-problema-del-agua?fb_comment_id=1919706488040991_2396617700349865#:~:text=Por%20ejemplo%2C%20el%20consumo%20promedio,promedio%20por%20habitante%20al%20d%C3%ADa.
- Polanyi, M. (1983). *The tacit dimensión*. Peter Smith.



Imagen: freepik.com

Integridad académica y recursos digitales: acotaciones para una reflexión de vista a los recursos derivados de la inteligencia artificial

Geovany Meza Chavero*

De manera reciente la cuestión de la integridad académica se discutió ampliamente en los ámbitos universitarios del país. Ello gracias a que el tema cobró un renovado interés a la luz de ciertos cuestionamientos, provenientes de la opinión pública, en cuanto a los procesos de evaluación académica por los que atraviesan miles de alumnos cada año, y mediante los cuales se logra comprobar la formación de los futuros profesionistas del país, incluidos aquellos por los que pasaron profesionistas que se desempeñan como servidores públicos (Santos, 2023). Asimismo, la emergencia de tecnología que se auxilia del desarrollo de inteligencia artificial (IA) tiene uno de sus campos de aplicación en los programas denominados anti “plagio”, mediante las cuales se examinan los trabajos escritos, publicaciones y proyectos que se presentan a evaluaciones académicas, esto con el propósito de encontrar coincidencias con bases de datos y repositorios institucionales.

No obstante, las ventajas evidentes de las que hoy se dispone como resultado de la IA, tanto en medios profesionales como de enseñanza, resulta útil contextualizar su inclusión en la cuestión de la integridad académica, así como vislumbrar los probables dilemas que se presentaran durante su incorporación al ámbito académico. En el presente texto me propongo señalar precisamente algunos de los ámbitos de discusión desde donde se han aportado reflexiones a la integridad académica, así como contribuir en la discusión de ¿Cuáles serían los hitos que tendrá la incorporación de IA como auxiliar de enseñanza e investigación?

Integridad, academia y plagio

Cuando hablamos de ética académica nos referimos al conjunto de principios o valores que regulan las actividades de las personas que participan en comunidades dedicadas a la enseñanza y la investigación, para el cumplimiento de las funciones de dicha comunidad. En este sentido, la integridad académica hace referencia del actuar de los individuos pertenecientes a dicho colectivo, en cuanto a

sus acciones, si estas se apegan o por el contrario se alejan de dichos valores. De ahí su importancia y necesidad de discusión.

Se puede decir que la parcialidad en evaluaciones académicas, la autoría ficticia, el sesgo en publicaciones el uso indebido de los recursos materiales y humanos provenientes de una institución, son ejemplos de lo que podemos entender por faltas a la integridad académica, esto debido a que su efecto vulnera la correcta realización de actividades encaminadas a cumplir la misión de enseñanza-aprendizaje, investigación y difusión de la cultura (Vera, 2016a).

Desde luego la cuestión que ha sido abordada más ampliamente es el llamado “plagio” entendido como “el acto de ofrecer o presentar como propia, en su totalidad o en parte, la *obra de otra persona” (OMPI, 1980, p. 188). La preocupación por el tema ha llevado a instituciones como el Instituto Tecnológico de Massachusetts a implementar un manual que guía la toma de decisión de sus alumnos ante situaciones donde se ha detectado faltas al derecho de autor (Brennecke, 2020). No obstante, podemos ponderar dicha situación al pensar que no toda mala práctica académica se reduce al plagio. Como se ha reflexionado, principalmente mediante la noción de “academic misconduct” (University of Cambridge, 2019), en realidad se cuenta con un abanico de situaciones, que al igual que el “plagio”, perjudican la función de una institución de educación superior (Gantús, 2016). Por tal motivo, se vuelve necesario recordar que las medidas que se toman para sancionar, impedir o prevenir un tipo de falta a la integridad académica, no implica necesariamente un efecto “domino” en otras, como ya lo han advertido otros autores (Gantús, 2016), es necesario analizar el propio funcionamiento, en particular de la institución para proponer una política educativa integral.

*Egresado de la Licenciatura en Estudios Latinoamericanos, FFyL, UNAM; Participó en el proyecto *Ética Académica UNAM* (PUB); Se desempeña en el área de Comisiones Dictaminadoras de la FFyL, UNAM.

Más allá del debate que se suscitó a inicios de año, el tema del “plagio”, en específico aquel que se lleva a cabo en trabajos académicos, y que *a posteriori* fuera detectado en trabajos para la obtención de grado, que presentaron en su momento profesores, investigadores y funcionarios públicos, lleva por lo menos una década como tema recurrente en medios de comunicación nacionales (Piñón, 2013). Al respecto es pertinente indicar, como lo han hecho distintos especialistas (Inclán, 2016), y como se puede inferir al observar los años de los trabajos en cuestión, la situación es de más larga data e incluso se puede rastrear incluso a niveles anteriores de escolaridad. El desarrollo histórico de la cuestión, en la perspectiva de los propios integrantes de la comunidad académica, es más larga todavía. Diversos estudios han dado aportes al respecto, como el del historiador Anthony Grafton en su obra hoy ampliamente recurrida *Los orígenes trágicos de la erudición*, la cual nos presenta las vicisitudes de las imprecisiones de referencias en textos académicos, que ocurrían incluso desde los albores del siglo XIX (Grafton, 2015). Como se ha señalado (Koepsell y Ruiz, 2015), podemos considerar que la mala práctica científica tuvo sus inicios a la par de su propio quehacer, y que el cuestionamiento ético de su praxis fue posterior.

Para establecer una justa relación entre lo que se dice acerca de la integridad académica y sus aportes o desventajas para las instituciones, es necesario indicar brevemente las áreas que han aportado iniciativas y debates sobre el tema, lo cual es un ejercicio disciplinar siempre útil.

Algunos enfoques de la cuestión

En primer lugar, la disciplina que quizá más tiempo ha puesto en reflexión la virtud, el deber y la naturaleza de los actos humanos, tanto de manera individual como en su colectivo, es la Filosofía, en particular la rama de la Ética. En un sentido aplicado, como en su momento se propuso para el estudio del *ethos* de la ciencia (Koepsell y Ruiz, 2015), ubica los valores que son comunes a los

miembros de las comunidades académicas, los cuales suelen ser expresados a través de códigos de ética que orientan el proceder de los académicos y alumnos, de manera particular en casos donde la reglamentación previa no es lo suficiente explícita. Además de identificar ciertos valores, proporciona corrientes, teorías desde las que se pueden abordar los dilemas que se presentan (deontología, utilitarismo, entre otros).

"Cuando hablamos de ética académica nos referimos al conjunto de principios o valores que regulan las actividades de las personas que participan en comunidades dedicadas a la enseñanza y la investigación..."

El segundo es el ámbito normativo, el cual desde una perspectiva nacional nos remite a los derechos de autor y la propiedad industrial, a los cuales recurren las universidades para el registro y protección de sus publicaciones e invenciones. Asimismo, las instituciones de educación superior generan reglamentos internos con el propósito de prevenir y sancionar las violaciones a estos derechos, además de las faltas a la integridad académica. Al respecto la Universidad Nacional Autónoma de México ha emitido su propio código de ética (UNAM, 2015), así como autorizó la integración de comités de ética institucionales (UNAM, 2019) y de manera más reciente las adecuaciones a su estatuto general para prever sanciones en el caso específico del plagio (Hernández, 2023). En principio puede resultar abrumador conciliar conceptos como autor, obra, derechos patrimoniales y derechos

morales, así como sus derivados, aún más al momento de distinguir sus aplicaciones en el mundo académico, no obstante, podemos pensarlos como un complemento, ya que ni el INDAUTOR puede emitir un grado académico, ni una universidad una cédula de derechos sobre una obra. Como ya han acotado otros autores (Vera, 2016b), cuando al interior de una normatividad académica existía un vacío al respecto, se recurría a la instancia federal para conciliar el conflicto de derechos, sin que en ocasiones tuviera una repercusión dentro de la institución. En este último caso, la inclusión de procedimientos disciplinarios que delimiten claramente el tipo de falta y su procedimiento correspondiente robustece las opciones a las cuales pueden recurrir instituciones e individuos que se consideren afectados.

El tercer ámbito se relaciona con la perfectibilidad de la formación profesional desde el medio educativo. En él se discuten las políticas que adoptan las instituciones en cuanto a las faltas a la integridad académica, por ejemplo, si se opta por una tolerancia cero o si es necesario establecer una actuación específica para estudiantes. Asimismo, la actualización curricular en cuanto a la formación metodológica y la inclusión de la ética académica en planes y programas de estudios, así como la incorporación de nuevas tecnologías para el apoyo docente y la evaluación de alumnos. En este último punto se ha estudiado cómo la disposición de información a través de las TIC puede contribuir a propiciar situaciones como el plagio (Inclán, 2016). No resulta nimio que una bibliografía actualizada contribuya o ponga en perspectiva la cuestión.

Al respecto, se han retomado las recomendaciones del prolífico escritor Humberto Eco en cuanto a la elaboración de una tesis (como se citó en Vera, 2016 a). Entre estas se reconoce el empleo de la autoría ficticia o por encargo e incluso la traducción como un auxiliar del plagio. Una postura sobre este texto, bastante conocido y socorrido en seminarios y cursos, puede ser su retiro de la

bibliografía. No obstante, también cabría la posibilidad de generar una reflexión adicional, se podría cuestionar qué llevó a un escritor prolífico como Eco a realizar estas recomendaciones, si es ello es aceptable hoy en día, o preguntarnos ¿cuáles eran las condiciones de la universidad en ese momento?, ¿los profesionistas contemporáneos de Eco comparten esta visión? Este último ejemplo pone en perspectiva la importancia de la formación integral y preventiva, misma que permanece como una alternativa vigente a mediano y largo plazo.

Del plagio artesanal, al software y la IA: el desarrollo tecnológico para el apoyo de las actividades académicas

Hoy en día se atisba distante el uso de la escritura no electrónica, sin nubes o bases de información asequibles las 24 horas del día, desde cualquier parte del mundo. Tener diversas fuentes de información a un solo clic. Ciertamente el acceso a la información con el que contamos en la actualidad difiere sobremanera con el que se contó en las universidades cuando Eco escribió *¿Cómo se hace una tesis?* (Eco, 2004). Ahora en una revisión un tutor puede hacer uso de iThenticate, seleccionar el idioma para refinar la búsqueda de “coincidencias” y obtener un listado de textos disponibles en la web con un porcentaje determinado de similitud, lo cual hasta hace pocos años todavía suponía un grado de complejidad en cierto sentido (Vera, 2016b). Asimismo, un alumno puede recurrir a programas y aplicaciones como Zotero, Mendeley o Evernote para gestionar las fuentes que hay recopilado y obtener la referencia bibliográfica en el formato de citación deseado.

No obstante, dicha inmersión tecnológica trajo consigo la simplificación en la disposición de información por medios pensados inicialmente como herramientas para el aprendizaje y la publicación, en lo cual resulta fundamental una formación ética en el tratamiento de las fuentes (Inclán, 2016; Gantús, 2016). En dicho orden de ideas no es extraño prever que el desarrollo de nuevas

tecnologías aplicadas a la simplificación de la lectura, escritura y síntesis de información traerá consigo nuevos retos para la revisión y dictaminación del trabajo académico. Aplicaciones como Chat GPT o Jasper ya cuentan con generación de texto mediante una selección temática, donde incluso se incorporan fuentes. Una de las primeras situaciones que nos llega a la mente con el arribo de aplicaciones, programas y sitios que proporcionen este tipo de recurso, incluso para la generación de imágenes o diapositivas, sería la automatización del trabajo académico.

Al respecto, se puede considerar que la cuestión radica en el uso de la herramienta, donde recordemos que toda técnica y procedimiento requiere de la capacitación del usuario para su uso profesional. Ciertamente la simplificación de interfaz para los mismos es un refinamiento técnico que lleva al empleo intuitivo hoy en día, lo cual buscaría

en primera instancia la simplificación de la actividad. Sin embargo, en el caso del trabajo académico, este tipo de producto visual o escrito puede tener como despropósito su presentación como una obra escrita de un alumno o académico, con lo que se recae en una falta a la integridad académica ya conocida, el fraude. La pretensión de autenticidad o la simulación de la autoría es un tema que ya se ha tratado anteriormente, así como las distintas acepciones que puede tener el plagio considerado como fraude, en caso de que no se integren las referencias correspondientes (Vera, 2016b; Yankelevich, 2016). De tal caso, la supresión del autor de una obra, por un producto generado por IA, por lo menos hasta el momento, y en el contexto académico, continua a disposición de quien lo presenta. Por ello resulta fundamental que la preparación ética acompañe la instrucción y uso de dichos recursos.

Referencias

- Brennecke, Patricia. (2020). *Academic integrity at MIT. A Handbook for students*. MIT Office of the Vice Chancellor. <https://integrity.mit.edu/handbook/academic-integrity-handbook>
- Eco, Humberto. (2004). *Cómo se hace una tesis: técnicas y procedimientos de estudio, investigación y escritura*. Gedisa.
- Gantús, Fausta. (2016). Conocimientos colectivos, obras particulares. Algunas reflexiones en torno al plagio académico. *Perfiles educativos*. Vol. 38, núm. 154.
- Grafton, Anthony. (2015). *Los orígenes trágicos de la erudición*, Fondo de Cultura Económica.
- Hernández, Mirtha. (2023, abril 10). Fortalece la Universidad integridad y honestidad académicas. *Gaceta UNAM*.
- Inclán, Catalina. (2016). Ctrl-C, Ctrl-V. La práctica escolar de copiar y pegar en el bachillerato. *Perfiles educativos*. Vol. 38, núm. 154.
- Koepseel, David R. y Ruiz de Chávez, Manuel H. (2015). Ética de la Investigación, *Integridad Científica*. Comisión Nacional de Bioética.
- Organización Mundial de la Propiedad Intelectual. (1980). *Glosario de derechos de autor y derechos conexos*. WIPO.
- Piñón, Alida. (2013, junio 5). La historia de un plagio serial. *El Universal*. <https://archivo.eluniversal.com.mx/notas/927579.html>
- Santos Cid, Alejandro. (2023, enero 11). La UNAM concluye que la tesis de 1987 de la ministra Yasmín Esquivel es un plagio. *El País*. <https://elpais.com/mexico/2023-01-12/la-unam-concluye-que-la-tesis-de-1987-de-la-ministrayasmín-esquivel-es-un-plagio.html>
- Universidad Nacional Autónoma de México. (2015, julio 30). Código de Ética de la Universidad Nacional Autónoma de México. *Gaceta UNAM*. https://dgapa.unam.mx/images/etica/2015_codigo-etica-unam.pdf
- Universidad Nacional Autónoma de México. (2019, agosto 29). Acuerdo por el que se Establecen los Lineamientos para la Integración, Conformación y Registro de los Comités de Ética en la Universidad Nacional Autónoma de México. *Gaceta UNAM*. <https://www.gaceta.unam.mx/wp-content/uploads/2019/10/190829-Convocatorias.pdf>
- University of Cambridge. (2019, noviembre 27). *What is academic misconduct?* <https://www.plagiarism.admin.cam.ac.uk/definition>
- Vera, Héctor. (2016). El plagio y la autonomía de las instituciones académicas. *Perfiles educativos*. vol. 38, núm. 154, 2016.
- Vera, Héctor. (2016). Introducción. El plagio nuestro de todos los días. *Perfiles educativos*. vol. 38, núm. 154, 2016.
- Yankelevich, Javier. (2016). Mapas prestados para entender el plagio académico. *Perfiles educativos*. vol. 38, núm. 154, 2016.

Desafíos para la regulación en torno al desarrollo y aplicación de la IA

Garbiñe Saruwatari Zavala*

El maquinismo de la Revolución Industrial dio origen a la dinámica de producción de bienes que conocemos hoy en día. Al sistematizarse los procesos de producción a través de la introducción de maquinaria, se dio origen al sistema fabril, en el que no sólo hubo mayor celeridad en la elaboración de productos, sino que se abarataron los costos y se uniformó la calidad de estos. Esta sistematización también contribuyó a un aumento en el consumo, en el que la oferta podía rápidamente satisfacer a la demanda.

Hubo varios fenómenos sociales como el crecimiento demográfico de las ciudades, la migración de la población del campo hacia los centros urbanos, el surgimiento del proletariado, los cuales revolucionaron la organización de las sociedades (Rafferty, s/f). Algunos pensaron que este rápido crecimiento urbano tuvo como consecuencia positiva el surgimiento de los derechos económicos, sociales y culturales, mediante los cuales se logró reconocer dentro del *nivel de vida adecuado* (HCHR, 2018), a los derechos a la alimentación, al agua, al saneamiento, vivienda digna, vestido, atención a la salud, protección y asistencia social; derechos que hoy en día, no sólo deben ser reconocidos, sino garantizados por los Estados, junto con los derechos de los trabajadores. Pero no se puede olvidar, que estos derechos surgieron por la lucha de los trabajadores para mejorar sus condiciones, en medio de las huelgas, que, en gran parte, se llevaron a cabo contra el maquinismo, debido al desplazamiento de la mano de obra.

Con la robótica en el siglo XX surgió una especie de nuevo “maquinismo”, en el que otro tipo de tareas aún más especializadas, se sistematizaron y perfeccionaron. En el imaginario colectivo se han instalado de forma positiva o negativa, imágenes de los robots surgidas del arte (cine, literatura de ciencia ficción, escultura), para cumplir la fantasía humana de una mejora en la calidad de vida, de la mano de una interacción virtuosa entre robots y humanos o, por el contrario, los robots como aquellos que encarnan la pe-

sadilla de la destrucción, aterrorizando a la especie humana. En los puntos intermedios de esta polarización, hay aspectos en los que fehacientemente puede comprobarse que la mecanización ha mejorado de manera sustantiva la relación del ser humano con su entorno; pero esto no debe darse por sentado para soslayar la reflexión sobre la repercusión del uso de la tecnología en todas las formas de vida.

Los claros beneficios aportados por el avance científico no pueden cegarnos ante los riesgos de que la maquinaria desplace al ser humano. No se trata de establecer de manera maniquea, una postura *ludita*¹ (Coll, 2020) por un lado o, una postura acrítica, por el otro, descalificando y etiquetando como retrógrados a quienes no están de acuerdo con la acelerada tecnologización. Desde hace décadas ronda el temor de que la automatización supla al trabajo humano, elimine puestos de trabajo y deje a miles de obreros y mano de obra calificada fuera del mercado laboral (Frey & Osborne, 2013). Ahora con la Inteligencia Artificial (IA) se extiende la preocupación hacia el ejercicio de distintas profesiones y artes, aplicaciones como ChatGPT han aumentado este temor; basta ver que, en la reciente huelga de guionistas de Hollywood, una de las peticiones sindicales es que su labor creativa en la industria del entretenimiento no sea reemplazada por la IA (Scheiber, 2023).

La tecnología como bien sabemos, *per se*, no es buena ni mala, sino son los usos que los humanos le damos, los que tienen una finalidad positiva o negativa y que deben ser examinados a través de sus implicacio-

*Licenciada en Derecho; Maestra en Bioética y Doctora en Cultura de Derechos Humanos; Jefa del Departamento de Estudios Jurídicos, Éticos y Sociales, y Presidenta del Comité de Ética en Investigación, INMEGEN.

¹ Ludismo: movimiento que surgió en Inglaterra, durante la Revolución Industrial, que promovía el rechazo de las máquinas y la automatización. Este fue promovido en el siglo XIX por los artesanos ingleses, que protestaron con el fin de no perder su empleo y, en casos extremos, destruyeron maquinaria.

nes éticas y jurídicas, así como la evaluación de su repercusión social. La *Recomendación sobre la Ética de la Inteligencia Artificial*, adoptada por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, 2021), señala que los aspectos éticos relativos de los sistemas de IA están vinculados a todas las etapas de su ciclo de vida, desde la investigación, concepción, desarrollo, despliegue, utilización, mantenimiento, funcionamiento, comercialización, financiación, seguimiento, evaluación, validación, desmontaje y hasta su terminación. Por lo que resulta fundamental identificar los distintos momentos, actores y partes interesadas, sean personas físicas o morales, públicas o privadas, para comprender el alcance de la responsabilidad jurídica de los involucrados durante todo el proceso o en una etapa de este; siendo lo primordial, considerar los derechos de las personas o grupos que pudieran ser vulnerados.

En este sentido, el documento de trabajo emitido por el Comité Ad hoc sobre Inteligencia Artificial del Consejo de Europa (CAHAI-COE, 2021), igualmente recomienda que deben existir medidas aplicables desde el desarrollo hasta la aplicación de los sistemas de IA, entre ellas, mecanismos de transparencia y de protección de datos, evaluaciones de riesgo y debates públicos basados en evidencia (numerales 22 y 25). Este documento sobre los *Elementos potenciales de un marco legal para el desarrollo, diseño y aplicación de la IA, basado en sus estándares sobre los Derechos Humanos, la Democracia y el Estado de Derecho*, hace énfasis en la clasificación de los riesgos para lo cual propone que, un instrumento transversal jurídicamente vinculante debe establecer una metodología de evaluación del impacto que sea concreta, clara, objetiva y con categorías de bajo o alto riesgo o de riesgo inaceptable. Estas dos últimas categorías permitirían marcar una moratoria o restricciones y condiciones para uso excepcional o incluso una prohibición, debido a que los riesgos inaceptables pueden interferir con el goce de los derechos humanos, el funcionamiento de la democracia o la observancia de la ley. A tra-

vés de este documento y otro titulado *Hacia la Regulación de los Sistemas de IA* (Ben-Israel, 2020), el mismo Comité (CAHAI) llama la atención sobre: (i) el uso indiscriminado por parte de instancias públicas o privadas, de los sistemas de reconocimiento facial y otras formas de identificación biométrica que permitan catalogar o inferir emociones o características de los individuos; (ii) la vigilancia masiva y formas de seguimiento a través de servicios de ubicación o de comportamiento en línea; (iii) el sesgo algorítmico o la discriminación provocada por los valores implicados en la recolección de datos usados para la codificación de los algoritmos de las aplicaciones; (iv) los sistemas de escalas sobre características físicas o mentales o de elaboración de perfiles que puntúen o califiquen el comportamiento de individuos para asignarles acceso a servicios esenciales; (v) los sistemas de IA encubiertos; (vi) programas con capacidad de generar imágenes, texto y audio sintéticos o falsificaciones profundas y con capacidad de distribuir este contenido a través de medios sociales; (vii) las interfaces humano-IA, entre otros.

Cabe mencionar que también, a través de programas o *software* malicioso que puede ser escalable, se puede tomar el control de vehículos, dispositivos, robots o infraestructura digital, para provocar ciberataques o apuntar hacia las vulnerabilidades de los sistemas (González Arencibia & Martínez Cardero, 2020). Desde hace años se han tratado de combatir los delitos relacionados con la piratería informática, como el *phishing* pero, desafortunadamente, aun no se ha logrado la erradicación ni sanción de estas conductas, por lo que se prevé que este aspecto obscuro siga acompañando al debate sobre la IA.

En marzo de 2023, la *Relatoría Especial para la Libertad de Expresión* (RELE) de la Comisión Interamericana de Derechos Humanos (CIDH) visitó diferentes empresas y organizaciones de internet radicadas en San Francisco, California, EE.UU., para dialogar y recopilar información sobre sus políticas y prácticas en materia de derechos humanos

y tecnología. Llama la atención que, derivado de la consulta se destacó como un desafío, el uso de la IA para los sistemas de moderación de contenido y su uso para elaborar contenido en las redes sociales. Existe el riesgo de que un uso inapropiado pueda exacerbar las tensiones sociales preexistentes y las disparidades de poder entre los diferentes actores involucrados (RELE-CIDH, 2023). Este riesgo, aunado al sesgo algorítmico, puede contraponer las libertades civiles y políticas del ciudadano, (como la de pensamiento, prensa, opinión), con el control y supervisión de los sistemas que debe ejercer el gobierno por motivos de seguridad o de monitoreo del funcionamiento.

"Desde hace décadas ronda el temor de que la automatización supla al trabajo humano, elimine puestos de trabajo y deje a miles de obreros y mano de obra calificada fuera del mercado laboral..."

Otra recomendación del CAHAI para los Estados miembros del Consejo de Europa, pero que bien puede ser adoptada por cualquier país o región, es el establecimiento de las llamadas "cajas de arena", que en este caso serían "cajas de arena regulatorias" [*regulatory sandboxes*], entendidas como espacios controlados para permitir pruebas a los sistemas de IA, lo que por un lado, estimularía la innovación responsable, pero por otro, permitiría evaluar el cumplimiento de las normas nacionales y los estándares de las normas multinacionales, como las del

Consejo de Europa. Estas cajas pueden ser mecanismos de menor escala y costo para llevar a cabo estudios de factibilidad y monitoreo, bajo riesgos calculados.

Por su parte, el gobierno de Estados Unidos de América publicó el *Plan para una Declaración de Derechos de IA* (The White House, 2022), que, a manera de catálogo de derechos humanos, consagra una serie de principios:

- (i) Sistemas seguros y efectivos: los usuarios deben ser protegidos contra sistemas inseguros e ineficaces. El documento de la Casa Blanca coincide con los instrumentos antes mencionados, en relación con la identificación y evaluación de riesgos, implementación de pruebas y de mecanismos de mitigación de resultados peligrosos, incluidos aquellos surgidos más allá del uso previsto de la tecnología en cuestión, así como el constante monitoreo sobre su seguridad, efectividad y eficacia.
- (ii) Protección contra la Discriminación Algorítmica: los usuarios deben ser protegidos contra la discriminación algorítmica y los sistemas deben ser diseñados de manera equitativa.
- (iii) Privacidad de la Información: los usuarios deben poder decidir sobre el uso de su información y estar protegidos contra prácticas abusivas de la utilización de sus datos.
- (iv) Aviso y Explicación: los usuarios deben saber que están utilizando un sistema automatizado y comprender a través de una clara explicación técnica, cómo es que esto contribuye en los resultados que lo impactan.
- (v) Alternativas humanas, consideración y respaldo: los usuarios deben poder optar por no interactuar con un sistema automatizado y tener acceso al trato de una persona que pueda solucionar el problema, de manera eficaz, eficiente y rápida, mediante una capacitación

adecuada. El trato humano no debería convertirse en una carga para el público usuario, sino en un apoyo y acompañamiento.

En el ámbito de la atención a la salud, también se han emitido instrumentos que sopesan los puntos favorables o desfavorables del uso de la IA, tales como la *Declaración sobre Inteligencia Aumentada en la Atención Médica* (AMM, 2019); el documento titulado *La Inteligencia Artificial en la salud pública* (OPS, 2021) y el *Reporte sobre el Impacto de la Inteligencia Artificial en la relación Médico Paciente* (COE, 2021). En ellos se hace énfasis en que la tecnología debe estar centrada en las personas, controlada y supervisada por seres humanos, así como fundamentada en la ética. La integridad científica es de suma relevancia para que los sistemas de IA sean confiables y accesibles, ya que las intervenciones de IA deben ceñirse a rigurosos estándares científicos y éticos, que permitan la reproducibilidad y apertura para poder ser compartidos y los beneficios puedan ser extendidos hacia un mayor número de personas.

Al igual que los demás documentos nombrados, la protección a las personas se centra en aspectos de transparencia, protección de datos, respeto a la privacidad, condiciones de equidad y no discriminación, pero en el contexto específico de los instrumentos centrados en la salud, es imperativo que la dinámica del consentimiento informado entre paciente y personal de salud, así como entre sujeto de investigación e investigadores, se actualice de acuerdo a los desafíos que plantea la IA. Usualmente, el personal médico debe asegurarse que la información que brinda al paciente sea comprensible, para que pueda tomar decisiones informadas; pero en este caso, la información también debe encauzarse hacia la comprensión del funcionamiento de la IA por parte del paciente. En esta tesitura, también es importante que el paciente conozca el riesgo del manejo de la información. El expediente clínico electrónico, por ejemplo, desde hace años ha planteado el reto sobre el control para

el acceso a las plataformas que contienen y almacenan la información. Incluso el manejo de bases de datos de salud, conformadas por información obtenida en el contexto de la investigación biomédica, ha generado una serie de reflexiones sobre las medidas de seguridad que deben adoptarse para evitar el acceso no autorizado o el intercambio inapropiado de datos (AMM, 2016). Los poderes legislativos en cada país deben seguir trabajando para la adecuada reglamentación del expediente clínico; de las bases de datos de salud, en especial las que almacenan información genética de las personas; de los biobancos y reservorios de material biológico humano y, de los acuerdos y condiciones para el intercambio de información de salud.

Una de las grandes virtudes de la IA es la automatización de tareas, pero existen los riesgos de que la responsabilidad con respecto al bienestar del paciente se diluya y, que se favorezca la falta de rendición de cuentas, debido a que los trabajadores de salud no serán los únicos, sino también los desarrolladores de sistemas de IA, que sean responsables de las imprecisiones, sesgos o incluso eventos adversos que resulten por un diagnóstico poco certero o interpretación errónea de la información. Asimismo, con relación a aplicaciones sobre bienestar físico o de salud mental, que estén al alcance del paciente o del consumidor, es crucial no automatizar ciertos procesos en las aplicaciones donde los factores psicológicos como la empatía (González Arencibia & Martínez Cordero, 2020), pueden ser determinantes tanto en la respuesta brindada por la aplicación como en la toma de decisiones por parte del paciente o usuario. Por ejemplo, *chatbots* especializados en diálogo, aplicados a líneas de ayuda a personas en peligro de suicidio (El-Atillah, 2023) o creados para brindar consejo a personas con trastornos alimenticios (Business Insider, 2022). Aunque sea el propio usuario quien accede voluntariamente a estas aplicaciones, se abre un abanico de diversas repercusiones legales ante el uso de la IA y la falta de compromiso de los desarrolladores para asumir responsabilidad por

su creación, desarrollo y cauce (Redacción RT, 2023). Por esta razón, resulta necesaria la evaluación constante de las tecnologías sanitarias que incorporan IA.

En síntesis, podemos concluir que, entre los principales desafíos para lograr la regulación del creciente desarrollo y uso de la IA en las diversas etapas de su ciclo de vida, se encuentran:

- Lograr el consenso, gobernanza y colaboración entre organizaciones internacionales y regionales, gobiernos, empresas, fabricantes y las múltiples partes interesadas para instaurar estándares, órganos regulatorios y códigos éticos que vigilen el desarrollo de la IA; para esto será imperativo implicar a la sociedad civil en el debate (González Arencibia & Martínez Cordero, 2020).
- Establecer la metodología y mecanismos para evaluar constantemente la proporcionalidad entre los riesgos y beneficios de las tecnologías, buscando como meta la seguridad e inocuidad en su implementación y su uso.
- Proveer a los ciudadanos, el acceso a los diversos servicios, sin discriminación, ni desigualdad, promoviendo la inclusión y diversidad. La *Declaración Universal sobre Bioética y Derechos Humanos* (UNESCO, 2005), constituye una guía definida, cuyos principios, como el respeto a la autonomía y a la integridad personal, la consideración de

la vulnerabilidad humana, la no estigmatización, el potenciamiento al máximo de los beneficios directos e indirectos, así como la reducción al máximo de los posibles efectos nocivos, pueden ser extrapolados al ámbito de la IA.

- Aminorar el impacto de la energía necesaria para mantener la tecnología en funcionamiento y disminuir su huella contaminante en el medio ambiente y en los ecosistemas (Strubell & Ganesh, 2019).
- Construir un marco regulatorio internacional, mediante el cual se busque que la IA no sea un elemento disruptor en la frágil construcción de la cultura de paz. Resulta innegable que el Internet y las redes sociales han logrado el surgimiento de un nuevo tipo de interconexión y, por lo tanto, de interacción social, que ha marcado el inicio del siglo XXI. En un mundo ideal, esta dinámica debería paliar algunas de las inequidades que frenan el desarrollo, ensanchan las brechas y excluyen a grandes grupos poblacionales (UNESCO, 2021). Lamentablemente, como humanidad no hemos logrado consolidar la aspiración de vivir en sociedades pacíficas y justas, por lo que sería preocupante que el inadecuado uso de la IA se sume a la crueldad ejercida por individuos o grupos escudados en el anonimato de las redes sociales o, a la incitación a la violencia o a la apología de delitos.



Imagen: freepik.com

Referencias

- AMM. (octubre de 2016). Declaración sobre las Consideraciones Éticas de las Bases de Datos de Salud y los Biobancos. *Asociación Médica Mundial*. Taipei, Taiwán: <https://www.wma.net/es/policias-post/declaracion-de-la-amm-sobre-las-consideraciones-eticas-de-las-bases-de-datos-de-salud-y-los-biobancos/>.
- AMM. (octubre de 2019). Declaración sobre Inteligencia Aumentada en la Atención Médica. *Asociación Médica Mundial*. Tiflis, Georgia: <https://www.wma.net/es/policias-post/declaracion-de-la-amm-sobre-inteligencia-aumentada-en-la-atencion-medica/>.
- Ben-Israel, I. (diciembre de 2020). Towards Regulation of AI Systems. *Hacia una Regulación de los Sistemas de IA (CAHAI-COE)*. Estrasburgo, Francia: <https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a>.
- Business Insider. (1 de noviembre de 2022). Chatbots: nuevos aliados en la prevención de trastornos alimenticios. *Business Insider México*, págs. s/p. <https://businessinsider.mx/chatbots-ayudan-prevenir-trastornos-alimenticios-estudio-ciencia/>.
- CAHAI-COE. (3 de diciembre de 2021). Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law. Estrasburgo, Francia: <https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d>.
- COE. (diciembre de 2021). Reporte sobre el Impacto de la Inteligencia Artificial en la relación Médico Paciente. *Comité de Derechos Humanos en los campos de la Biomedicina y la Salud del Consejo de Europa*. Estrasburgo, Francia: <https://rm.coe.int/inf-2022-5-report-impact-of-ai-on-doctor-patient-relations-e/1680a68859>.
- Coll, F. (8 de septiembre de 2020). *Ludismo*. Obtenido de Economipedia.com: <https://economipedia.com/definiciones/ludismo.html>
- El-Atillah, I. (4 de enero de 2023). Un hombre se suicida después de que un chat de IA le invitara a hacerlo. *Euronews*, págs. s/p. <https://es.euronews.com/next/2023/04/01/un-hombre-se-suicida-despues-de-que-un-chat-de-ia-le-invitar-a-hacerlo>.
- Frey, C. B., & Osborne, M. A. (17 de septiembre de 2013). The Future of Employment: How Susceptible are Jobs to Computerisation? Oxford, R.U.: https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf. Obtenido de Oxford University Engineering Sciences Department.
- González Arencibia, M., & Martínez Cardero, D. (2020). Dilemas éticos en el escenario de la inteligencia artificial. *Economía y Sociedad*, 93-109.
- HCHR. (2018). Artículo 25, Serie de artículos publicados en conmemoración del 70 aniversario de la Declaración Universal de Derechos Humanos. *Oficina del Alto Comisionado para los Derechos Humanos*. México: https://hchr.org.mx/wp/wp-content/themes/hchr/images/30acerca30/Art%C3%ADculo_25.pdf.
- OPS. (2021). La Inteligencia Artificial en la Salud Pública. *Organización Panamericana de la Salud*. Washington, D.C., EE.UU.: https://iris.paho.org/bitstream/handle/10665.2/53887/OPSEIHIS21011_spa.pdf?sequence=5&isAllowed=y.
- Rafferty, J. P. (s/f). The Rise of the Machines: Pros and Cons of the Industrial Revolution. *Encyclopedia Britannica*, <https://www.britannica.com/story/the-rise-of-the-machines-pros-and-cons-of-the-industrial-revolution>.
- Redacción RT. (4 de junio de 2023). Un 'chatbot' que reemplazó a trabajadores de una línea de ayuda contra los trastornos alimenticios da consejos dañinos y peligrosos. *RT*, págs. s/p. <https://actualidad.rt.com/actualidad/469152-chatbot-linea-trastornos-alimenticios-consejos-daninos-peligrosos>.
- RELE-CIDH. (31 de marzo de 2023). Comunicado de prensa R54/23. *Relatoría Especial para la Libertad de la Organización de Estados Americanos*, <https://www.oas.org/es/cidh/expresion/showarticle.asp?IID=2&artID=1271>. Washington D.C., EE.UU.: <https://www.oas.org/es/cidh/expresion/showarticle.asp?IID=2&artID=1271>.
- Scheiber, N. (4 de mayo de 2023). ¿La próxima 'Succession' será escrita por un chatbot? Obtenido de The New York Times: <https://www.nytimes.com/es/2023/05/04/espanol/guionistas-huelga-ia.html>
- Strubell, E., & Ganesh, A. (07 de 2019). Energy and Policy Considerations for Deep Learning in NLP. *57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: College of Information and Computer Sciences, University of Massachusetts. Obtenido de Cornell University: <https://doi.org/10.48550/arXiv.1906.02243>
- The White House. (octubre de 2022). Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People. *Plan para una Declaración de Derechos de IA*. Washington D.C., EE.UU.: <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>.
- UNESCO. (19 de octubre de 2005). Declaración Universal sobre Bioética y Derechos Humanos. París, Francia: https://unesdoc.unesco.org/ark:/48223/pf0000146180_spa.
- UNESCO. (23 de noviembre de 2021). Recomendación sobre la ética de la Inteligencia Artificial. París, Francia: https://unesdoc.unesco.org/ark:/48223/pf0000380455_spa.

Los desafíos éticos y científicos de ChatGPT en salud: utopismo, tecnofobia y pragmatismo

Luis García Valiña* e Ignacio Mastroleo**

Es posible que el 30 de noviembre de 2022 haya sido un punto de inflexión en el proceso de cambio tecnológico en curso. Ese fue el día en que se liberó al público en general la versión 3.5 del algoritmo conversacional ChatGPT. De acuerdo a sí mismo, “ChatGPT es un modelo de lenguaje desarrollado por OpenAI. Se basa en la arquitectura GPT-3.5 y ha sido entrenado con una amplia variedad de datos y textos para comprender y generar respuestas en lenguaje natural. ChatGPT está diseñado para interactuar con los usuarios a través de conversaciones en línea, proporcionando respuestas coherentes y contextualmente relevantes a las consultas y solicitudes de los usuarios. Es capaz de comprender preguntas, brindar información, ofrecer sugerencias y participar en conversaciones *de manera similar a como lo haría un humano*” (cursivas agregadas).¹

Aunque la tecnología en sí misma no es tan novedosa, ha sido justamente esa capacidad de generar respuestas similares a las humanas (que ahora le permite incluso trabajar con imágenes, en su versión 4.0) y su masividad las que han dado lugar a dos actitudes que hoy prevalecen en la conversación sobre IA y que podemos asociar, un poco laxamente, con el “bombo” (hype), por un lado; y con una suerte de *distopía pesimista*, que ha sido expresada en su versión extrema como un temor a que la inteligencia artificial acabe con la humanidad, por el otro.²

Creo que esas dos concepciones funcionan en el debate público acerca de los usos de modelos de lenguaje ampliado como ChatGPT más como distractores que como herramientas para ayudarnos a pensar cuál debe ser nuestro próximo paso. Cómo yo lo veo, ambas perspectivas están basadas en una concepción acerca de la tecnología que me gustaría cuestionar (una visión desmateriali-

zada y abstracta), para luego sugerir que sea reemplazada por otra, que llamaré “pragmatista”. Creo que este reemplazo tiene su importancia, dado que los pasos iniciales (y el temperamento intelectual con el que se los determina) con relación a una tecnología condicionan fuertemente los movimientos ulteriores. Habiendo tanto en juego, lo mejor sería que acertemos de entrada.

Empecemos con la posición “exageradamente entusiasmada”. Tres artículos recientes, uno general y dos específicos a los usos médicos, han reportado que la comunidad médica sigue con interés y expectativa la posibilidad de usar ChatGPT para la investigación (Deng, 2022; Sallam, 2023; Li, 2023). De acuerdo a estos relevamientos preliminares, aun cuando de momento no existen usos de ChatGPT suficientemente probados para su uso regular en la atención de la salud pública o la práctica médica clínica,³ los investigadores están interesados en usar ChatGPT para lograr más eficiencia en la escritura académica y la traducción de textos (en caso de investigadores de países no angloparlantes), en el procesamiento de datos y también

* Doctor en Filosofía; Profesor universitario; ha sido becario posdoctoral; colabora con el CRIION, Alemania; Investigador y Coordinador Académico, Programa de Bioética, FLACSO Argentina; trabajo interdisciplinario sobre las implicancias éticas, sociales y políticas de la Inteligencia Artificial.

** Investigador y Director Asociado, Programa de Bioética, FLACSO Argentina; Investigador, CONICET; Profesor universitario; recibió una beca de la Fundación Mertelsmann, con un proyecto de análisis ético-filosófico sobre el uso de la Inteligencia Artificial para la salud; Afiliado al CRIION, Alemania.

³ Por lo tanto, cualquier uso de ChatGPT en salud humana debería considerarse experimental o no probado en un sentido amplio, a saber, o bien ser un uso de investigación (clínica o de salud pública) para desarrollar conocimiento generalizable, o bien es un caso de práctica nueva no validada en beneficio de un paciente o población (popularmente conocido como innovación médica) (Mastroleo y Holzer 2020). En cualquier caso, su uso responsable debería cumplir con las protecciones y criterios éticos generales para los usos de intervenciones de salud no probadas dentro o fuera de ensayos clínicos (p.e. AMM 2013, CIOMS-OMS 2016, OPS 2022a, cap. 5, WHO 2022a), además de las consideraciones especiales para los usos de IA para la salud (p.e. WHO 2021).

¹ Consultado el 13 de junio de 2023.

² “Expert Says There’s a 50% Chance AI Will Wipe out Humanity” Daily Mail Online, <https://www.dailymail.co.uk/sciencetech/article-12151967/Expert-says-theres-50-chance-ai-wipe-humanity.html>, consultado el 13 de junio de 2013.

en la revisión de la literatura, entre otras cosas. Incluso algunos imaginan la posibilidad de que un algoritmo conversacional pueda ayudar a identificar nuevas hipótesis y métodos; es decir, no sólo como una “base de datos inteligente” sino como una verdadera herramienta de mejora cognitiva, en el espíritu de lo que William Schwartz imaginaba en un paper de 1970 (citado en Li, 2023): “los agentes conversacionales podrían servir como consultores en mejorar las funciones intelectuales de los médicos a través de las interacciones”. Con respecto al proceso clínico, los especialistas mencionan las ventajas para la medicina personalizada, la predicción de riesgos de enfermedades y resultados de tratamientos, mejorando el flujo clínico, los diagnósticos y la documentación, disminuyendo costos y mejorando la educación de los profesionales (op cit.).

El problema es que es fácil pasar de estas expectativas (bastante naturales y hasta justificadas, si atendemos a la verdadera explosión de aplicaciones ocurrida en apenas unos meses, especialmente aquellas destinadas a mejorar el proceso de investigación) al “hype” utópico.

Eric Topol, por ejemplo, ha sido uno de los más vigorosos defensores del uso de IA en medicina. Para Topol, el uso sistemático y generalizado de IA permitirá eliminar buena parte de las actividades administrativas que actualmente deben realizar los médicos, liberando tiempo para proporcionar un mejor tratamiento y restableciendo “la preciada y consagrada conexión y confianza -el contacto humano- entre pacientes y médicos.” (Topol, 2019: 32).

Hay un elemento de solucionismo tecnológico (Morozov, 2016) o “tecnochauvinismo” (Broussard, 2018) subyacente en estas palabras; es decir, la convicción de que los problemas éticos y políticos que enfrenta la humanidad (muchos de los cuales son de larga data) pueden ser interpretados en términos de un fallo de diseño y solucionados mediante la tecnología.

Sin embargo, el problema con esta perspectiva es su reduccionismo; es decir, el modo en que falla en considerar la complejidad de los problemas sociales, por un lado; y el carácter materializado de los procesos de automatización, su imbricación con el “mundo de la vida” y el resto de los procesos que intervienen en la experiencia humana, por el otro.

"La región de las Américas tiene un déficit de 600.000 profesionales (OPS, 2022b). A nivel mundial, se estimó antes del COVID-19 que serían necesarios 10 millones más de profesionales de la salud para 2030..".

A modo de ejemplo, podemos retomar la promesa de mejorar la investigación científica usando algoritmos como ChatGPT. Considerado sólo como un “modelo de procesamiento de lenguaje natural basado en la arquitectura GPT”; es decir, como un algoritmo informático o incluso un modelo matemático, ciertamente las expectativas no son infundadas. Ahora bien, considerar esta perspectiva de modo aislado e independiente de las arquitecturas materiales y sistemas de incentivos que operan en un sistema de investigación real, puede redundar en resultados inesperados. Por ejemplo, ¿es el financiamiento (siempre escaso) dependiente del número de publicaciones? ¿Cómo es ese sistema de publicaciones, de pago o abierto? Según la respuesta a estas preguntas (y otras), podría ocurrir que, por ejemplo, los investigadores, presionados por el dictum

de “publica o muere”, decidan relajar los estándares de control humano y depender en los resultados de ChatGPT, sin controlarlos o discutirlos. Existiendo ya numerosos casos de publicaciones apresuradas, plagio, etc., la introducción de ChatGPT puede de hecho terminar fomentando la ciencia irresponsable. Esto nos expone al riesgo de tener “más” investigación, pero de menor calidad.

Por otro lado, se ha puesto, y con razón, un énfasis especial en el modo en que los algoritmos pueden estar sesgados, pero no se habla mucho acerca del sesgo humano (de automatización, de autoridad, el efecto de *statu quo*, etc.) y de los elementos que pueden afectar su relación con los sistemas de IA (diferencias de temperamento y trasfondo cultural, educación, etc.). Como sostienen Kostick-Quenet y Gerke, tenemos que ampliar “nuestra noción de pruebas de usuario más allá de su enfoque actual en las métricas de rendimiento de la IA y la usabilidad próxima para examinar los factores humanos y sistémicos que dan forma a cómo los sistemas de IA son aplicados en la práctica por usuarios imperfectos en entornos imperfectos.” (Kostick-Quenet y Gerke, 2022, p. 2).

Otro ejemplo tiene que ver con el creciente desarrollo de aplicaciones conversacionales en el contexto de la salud mental. Según la Organización Mundial de la Salud, “los trastornos mentales son uno de los retos de salud pública más importantes en la Región Europea de la OMS, ya que son la principal causa de discapacidad y la tercera causa de carga global de morbilidad.” (OMS 2019, citado en Sedlakova, 2022). Sin embargo, un fenómeno creciente que plaga los sistemas de salud es la escasez de personal (Butryn, 2017). La región de las Américas tiene un déficit de 600.000 profesionales (OPS, 2022b). A nivel mundial, se estimó antes del COVID-19 que serían necesarios 10 millones más de profesionales de la salud para 2030 (WHO, 2022b). Los agentes conversacionales podrían ayudar en este contexto o en situaciones de pico de demanda (p.e. en una emergencia de salud pública), pero también podrían empeo-

rar la situación si no están suficientemente validados; o impedir a los Estados invertir en formación e infraestructura, perpetuando las inequidades en el acceso, la seguridad y la eficacia del tratamiento. Por ejemplo, entre quienes pueden acceder a profesionales humanos, quienes accederían a un chatbot “guiado” por profesionales y quienes deben conformarse con un chatbot “no guiado” (Graber-Stiehl, 2023).

Nuevamente, debemos considerar los efectos de la implementación de sistemas de automatización evitando asumir una posición reduccionista, que entiende la tecnología como una entidad inmaterial, un constructo matemático o una aplicación algorítmica. Por el contrario, sugiero seguir a la filósofa Kate Crawford cuando sostiene que, “comprendemos mejor el papel de la IA en el mundo a través de sus arquitecturas materiales, sus entornos contextuales y las políticas imperantes, así como de sus conexiones.” (Crawford, 2021: 12) Es decir, debemos pasar a ver a sistemas como ChatGPT en un sentido *sociotécnico* (Graham, 2021), como un conjunto de prácticas, instituciones, estructuras materiales, sistemas económicos y sociales, restricciones políticas, etc.

Pero estar atentos a la “economía política” de la IA no debería llevarnos hacia el extremo opuesto al tecnopotimismo acrítico; es decir, a una visión tecnofóbica que solo puede ver en la automatización una amenaza y una herramienta para la reificación de la injusticia y las asimetrías de poder.

Como señalan Sparrow and Hatherley, criticando justamente la tendencia de Topol a enfocarse más en la tecnología que en la sociedad, “todo ahorro de tiempo vuelto posible por una reducción en la carga administrativa de los médicos...será usada para mover más pacientes a través del sistema, antes que para permitirles a los médicos pasar más tiempo y ocuparse mejor de ellos” (2020:14). Sin embargo, este resultado depende, otra vez, de los incentivos, de las interacciones con otros puntos del sistema. No se trata, en

todo caso, de un rasgo inherente a la tecnología, sino al modo en que se operacionaliza en un contexto y con determinados objetivos en vista.

Debemos, entonces, adoptar una posición pragmática. Esto implica, en esencia, abandonar las idealizaciones asociadas a los usos potenciales de ChatGPT y enfocarnos en:

- Los usos concretos proyectados (¿es una aplicación para la asistencia virtual a pacientes? ¿para resumir historias clínicas? ¿se usará para la traducción de textos académicos?).
- Las interacciones con otros sistemas en uso (¿ayudará a consolidar la provisión de salud y la investigación o contribuirá por el contrario a erosionar otros sistemas sin un beneficio tangible? ¿es necesario usar IA o es suficiente algún otro sistema actualmente en operación mejor validado y más costo-eficiente?).
- El modo en que afectan otros criterios relevantes (¿mejora esta aplicación la relación entre pacientes y profesionales de salud? ¿implica un balance inapropiado entre la eficiencia y la privacidad, la explicabilidad o la transparencia? ¿Si las nuevas versiones de ChatGPT se entrenan usando datos con poblaciones que no tienen capacidad de pago, existen los mecanismos apropiados para compartir beneficios equitativos? ¿Si estas tecnologías están protegidas por dere-

chos exclusivos o secreto comercial, estarán disponibles, serán apropiadas y económicamente asequibles para investigadoras/es, profesionales de salud y pacientes?).

Estas preguntas (que seguramente deberán ser complementadas por otras, en un esquema de evaluación completo) se dirigen a un mismo punto: la evaluación de los usos de ChatGPT y otras tecnologías similares no puede ser hecha a priori ni de acuerdo a una concepción general de la inteligencia artificial; por el contrario, la actitud adecuada debe ser pragmática, orientada a las consecuencias concretas y tangibles de un uso potencial en un contexto específico y cualitativamente determinado.

En resumen, el verdadero desafío ético es cambiar el marco conceptual para evaluar la innovación tecnológica. Debemos asumir, en palabras del filósofo norteamericano William James, la actitud de apartarnos de “las primeras causas, principios, categorías, supuestas necesidades, y de mirar hacia las cosas últimas, frutos, consecuencias, hechos” (1948: 144). De este modo podemos evitar las tendencias gemelas a la exageración y la tecnofobia y ganar una perspectiva razonable acerca de los riesgos y desafíos de los modelos de lenguaje ampliado como ChatGPT.

Referencias

- Asociación Médica Mundial (AMM). Declaración de Helsinki de la AMM. Principios éticos para las investigaciones médicas en seres humanos [Internet]. 2013 [citado 21 de enero de 2016]. Disponible en: <https://www.wma.net/es/policies-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos/>
- Butryn. “The Shortage of Psychiatrists and Other Mental Health Providers: Causes, Current State, and Potential Solutions.” *International Journal of Academic Medicine, India*, 2017. <https://www.ijam-web.org/article.asp?issn=2455-5568;year=2017;volume=3;issue=1;spage=5;epage=9;aulast=Butryn>.
- Crawford, Kate. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021. <https://doi.org/10.2307/j.ctv1ghv45t>.
- Consejo de Organizaciones Internacionales de las Ciencias Médicas (CIOMS) en colaboración con la Organización Mundial de la Salud (OMS) (CIOMS-OMS). Pautas éticas internacionales para la investigación relacionada con la salud con seres humanos [Internet]. 2016 [citado 10 de octubre de 2016]. Disponible en: <http://iris.paho.org/xmlui/bitstream/handle/123456789/34457/9789290360902-spa.pdf?sequence=5>
- Deng, Jianyang, and Yijia Lin. “The Benefits and Challenges of ChatGPT: An Overview.” *Frontiers in Computing and Intelligent Systems* 2, no. 2 (2022): 81–83. <https://doi.org/10.54097/fcis.v2i2.4465>.
- “Expert Says There’s a 50% Chance AI Will Wipe out Humanity | Daily Mail Online.” Accessed June 13, 2023. <https://www.dailymail.co.uk/sciencetech/article-12151967/Expert-says-theres-50-chance-AI-wipe-humanity.html>.

- Graber-Stiehl, Ian. "Is the World Ready for ChatGPT Therapists?" *Nature* 617, no. 7959 (May 3, 2023): 22–24. <https://doi.org/10.1038/d41586-023-01473-4>.
- Graham, S. Scott. *The Doctor and the Algorithm: Promise, Peril, and the Future of Health AI*. New York: Oxford University Press, 2022.
- James, William. *Essays in Pragmatism*. New York: Hafner Publishing Co., 1948.
- Kostick-Quenet, Kristin M., and Sara Gerke. "AI in the Hands of Imperfect Users." *Npj Digital Medicine* 5, no. 1 (December 28, 2022): 1–6. <https://doi.org/10.1038/s41746-022-00737-z>.
- Li, Jianning, Amin Dada, Jens Kleesiek, and Jan Egger. "ChatGPT in Healthcare: A Taxonomy and Systematic Review." Preprint. Health Informatics, March 30, 2023. <https://doi.org/10.1101/2023.03.30.23287899>.
- Mastroleo I, Holzer F. New non-validated practice: an enhanced definition of innovative practice for medicine. *Law, Innovation and Technology*. 2 de julio de 2020; 12(2):318–46. <https://doi.org/10.1080/17579961.2020.1815405>
- Morozov, Evgeny. *La Locura Del Solucionismo Tecnológico*. Buenos Aires: Katz Editores, 2016. <https://www.eldiplo.org/wp-content/uploads/2018/files/6314/6463/4693/LaLocuradelSolucionismoTecnologico-Morozov-Introduccion.pdf>.
- Organización Panamericana de la Salud (OPS). *Catalizar La Investigación Ética En Emergencias. Orientación Ética, Lecciones Aprendidas de La Pandemia de COVID-19 y Agenda Pendiente*. Washington DC: OPS, 2022a. https://iris.paho.org/bitstream/handle/10665.2/56104/OPSHSSBIOCOVID-19220019_spa.pdf?sequence=5&isAllowed=y.
- Organización Panamericana de la Salud (OPS). Las Américas tienen un déficit de 600.000 profesionales de la salud, que afecta el acceso a la salud en las zonas rurales y desatendidas [Internet]. 2022b. [citado 13 de junio de 2023]. Disponible en: <https://www.paho.org/es/noticias/27-5-2022-americas-tienen-deficit-600000-profesionales-salud-que-afecta-acceso-salud-zonas>.
- Sallam, Malik. "ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns." *Healthcare* 11, no. 6 (March 19, 2023): 887. <https://doi.org/10.3390/healthcare11060887>.
- Schwartz, William B. "Medicine and the Computer: The Promise and Problems of Change." *New England Journal of Medicine* 283, no. 23 (December 3, 1970): 1257–64. <https://doi.org/10.1056/NEJM197012032832305>.
- Sedlakova, Jana, and Manuel Trachsel. "Conversational Artificial Intelligence in Psychotherapy: A New Therapeutic Tool or Agent?" *The American Journal of Bioethics* 23, no. 5 (May 4, 2023): 4–13. <https://doi.org/10.1080/15265161.2022.2048739>.
- Sparrow, Robert, and Joshua Hatherley. "High Hopes for 'Deep Medicine'? AI, Economics, and the Future of Care." *Hastings Center Report* 50, no. 1 (2020): 14–17. <https://doi.org/10.1002/hast.1079>.
- "The Superhero of Artificial Intelligence: Can This Genius Keep It in Check? | Artificial Intelligence (AI) | The Guardian." Accessed June 13, 2023. <https://www.theguardian.com/technology/2016/feb/16/demis-hassabis-artificial-intelligence-deepmind-alphago>.
- Topol, Eric J. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. First edition. New York: Basic Books, 2019.
- World Health Organization (WHO). Ethics and governance of artificial intelligence for health [Internet]. Geneva: World Health Organization. Health Ethics & Governance Team; 2021 jun [citado 13 de agosto de 2021]. Disponible en: <https://www.who.int/publications-detail-redirect/9789240029200>.
- World Health Organization (WHO). Emergency use of unproven clinical interventions outside clinical trials: ethical considerations [Internet]. Geneva: World Health Organization; 2022a [citado 26 de abril de 2022]. xii, 43 p. Disponible en: <https://apps.who.int/iris/handle/10665/352902>.
- World Health Organization (WHO). Global Strategy on Human Resources for Health: Workforce 2030: Reporting at Seventy-fifth World Health Assembly [Internet]. 2022b [citado 13 de junio de 2023]. Disponible en: <https://www.who.int/news/item/02-06-2022-global-strategy-on-human-resources-for-health--workforce-2030>.



Imagen: freepik.com

Resultados de las encuestas sobre la percepción de sesiones conjuntas entre el CEI, CI, CB o CICUAL aplicadas a los Comités de Ética en Investigación y a las Comisiones Estatales de Bioética

Areli Cerón Sánchez, Gabriela Pineda Hernández, Karla Alejandra Tovar López, Karla Gabriela Sánchez Villanueva y Flor de María Cruz Estrada*

Un elemento ético imprescindible, a nivel nacional e internacional, es la evaluación de toda propuesta de investigación que involucra seres humanos por un **Comité de Ética en Investigación** (CEI). Estos son, un grupo multidisciplinario de expertos con autonomía, plurales, consultivos, institucionales, y con independencia al grupo de investigación involucrado. Tienen objetivos y funciones definidas en los criterios que ha establecido la Comisión Nacional de Bioética (CONBIOÉTICA), a través de las Disposiciones Generales¹ y la Guía Nacional².

La evaluación de un protocolo de investigación debe realizarse desde el punto de vista ético, metodológico, **normativo**, y cuando aplique de **bioseguridad**. En el caso de México, a diferencia de otros países, esta la deben realizar el CEI, el **Comité de Investigación** (CI), y el **Comité de Bioseguridad** (CB), cuando aplique. Los tres comités deberán emitir su dictamen favorable como requisito previo para dar inicio a la investigación, adicionalmente, en los casos que involucren insumos para la salud con fines de registro sanitario³ deberán solicitar la autorización de la Comisión Federal para la Protección Contra Riesgos Sanitarios (COFEPRIS).

Por lo anterior, la regulación mexicana⁴ prevé que el CEI, el CI y el CB podrán **sesionar en conjunto, actividad fuertemente recomen-**

dada para favorecer una evaluación integral de las propuestas de investigación. Además, las **sesiones conjuntas** reducen los tiempos de evaluación y favorecen el inicio de la investigación en el menor tiempo posible. En los casos en que, en el establecimiento también se realicen proyectos de investigación con modelos animales no humanos, e integren un Comité Interno para el Cuidado y Uso de los Animales de Laboratorio (CICUAL) de conformidad con la NOM-062-ZOO-1999, también se sugiere explorar la posibilidad de adherirse a las sesiones conjuntas con los CEI, CI y CB.

Con base en lo anterior, la CONBIOÉTICA diseñó e implementó dos encuestas de sondeo con **el objetivo** de explorar la percepción de los CEI con respecto a las sesiones conjuntas de distintos comités: CEI, CI, CB y CICUAL. Asimismo, se decidió explorar el tema en paralelo con las Comisiones Estatales de Bioética (CEB), cuerpos consultivos con carácter multidisciplinario e interinstitucional, creados en las entidades federativas para extender la observación y práctica de los principios bioéticos y derechos humanos a través de capacitación, asesoría y difusión de la bioética en la atención médica e investigación, por lo que entre sus actividades está promover el cumplimiento de las disposiciones para que en los establecimientos que realicen investigación con seres humanos y se integren los CEI.

Para fines de las citadas encuestas se entiende por:

Sesión conjunta: A la sesión para evaluar un protocolo de investigación en la que participan el CEI, CI y cuando aplique el CB, con el objetivo de evaluar los aspectos éticos, metodológicos, de bioseguridad y normativos en conjunto.

¹ ACUERDO por el que se emiten las Disposiciones Generales para la Integración y Funcionamiento de los Comités de Ética en Investigación y se establecen las unidades hospitalarias que deben contar con ellos, de conformidad con los criterios establecidos por la Comisión Nacional de Bioética, publicado en el DOF (31/10/2012), así como sus modificaciones publicadas en el DOF el (11/01/2016) y el (10/12/2020).

² Guía Nacional para la Integración y Funcionamiento de los CEI, 2018, disponible en: https://www.gob.mx/cms/uploads/attachment/file/460756/7_Guia_CEI_2018_6a.pdf

³ Insumos para la salud: Incluye medicamentos, vacunas, bilógicos o dispositivos médicos.

⁴ Artículo 102 del Reglamento de la Ley General de Salud en Materia de Investigación para la Salud

*Directora de Comités de Bioética, Directora de Desarrollo Institucional, Jefa del Departamento de Apoyo a Comisiones Estatales de Bioética, Subdirectora de Enlace con Comisiones Estatales de Bioética y Jefa de Departamento de Registro de Comités de Ética en Investigación, respectivamente.

Dictamen conjunto: Al documento por escrito emitido en conjunto por los CEI, CI y cuando aplique por el CB respecto a la propuesta de investigación evaluada.

RESULTADOS

A. ENCUESTA APLICADA A LOS CEI

Se utilizó como técnica la **encuesta** y como **instrumento un formulario en Google Forms**, construido específicamente para el objetivo antes descrito. El formulario electrónico de auto-aplicación estuvo integrado por cinco preguntas, tanto de selección múltiple, como abiertas. El periodo de recepción de respuestas fue del 24 de abril al 31 de mayo del 2023. Al respecto, se recibieron **101 respuestas de los CEI**, lo que representa el 28% del total de comités registrados ante CONBIOÉTICA. La distribución de estos, por entidad federativa se muestra en la figura 1.

A continuación, el análisis gráfico de las respuestas a las cinco preguntas, se desglosa con mayor detalle. Primero se les preguntó a los comités si: 1) ¿Su CEI realiza sesiones conjuntas con el Comité de Investigación (CI), el Comité de Bioseguridad (CB) o el CICUAL de su **establecimiento**? Los resultados se describen en las figuras 2 y 3 para mayor referencia.

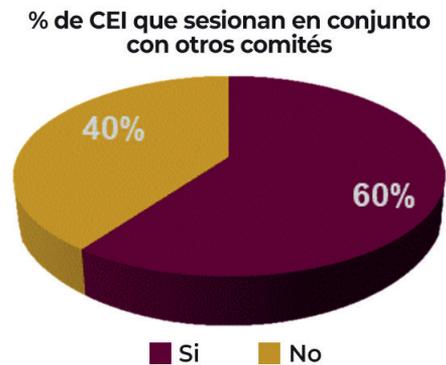


Figura 2. Porcentaje de CEI que sesionan en conjunto con otros comités (N=101).

FUENTE: Comisión Nacional de Bioética, elaboración propia.

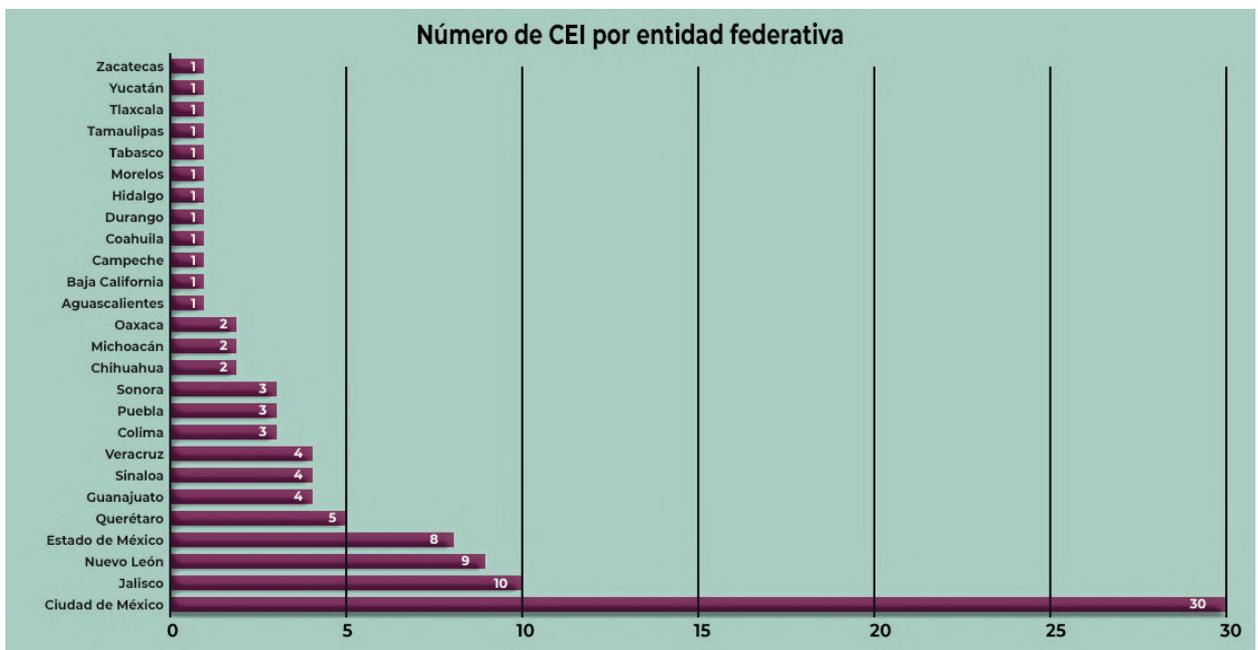


Figura 1. Distribución de los CEI que atendieron la encuesta por entidad federativa (N=101).

FUENTE: Comisión Nacional de Bioética, elaboración propia.

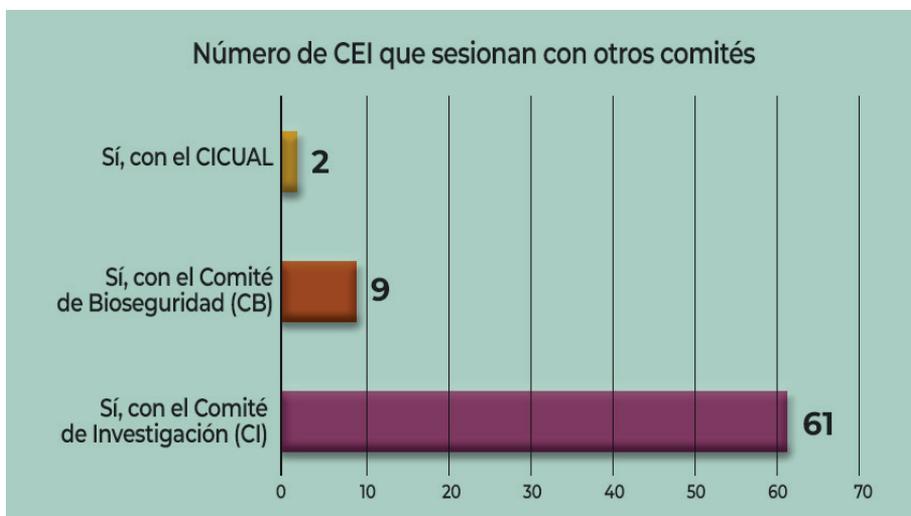


Figura 3. Número de CEI que sesionan en conjunto con otros comités (N=72). Considere que un mismo CEI puede sesionar con más de 1 comité.

FUENTE: Comisión Nacional de Bioética, elaboración propia.

Aquellos CEI que han sesionado en conjunto con alguno de los comités (CI, CB y/o CI-CUAL), con base en una escala del 1 al 10 (en la que 1 es nada útil y 10 es muy útil), com-

partieron su percepción respecto a la discusión y análisis conjunto. La cual se describe en la figura 4.

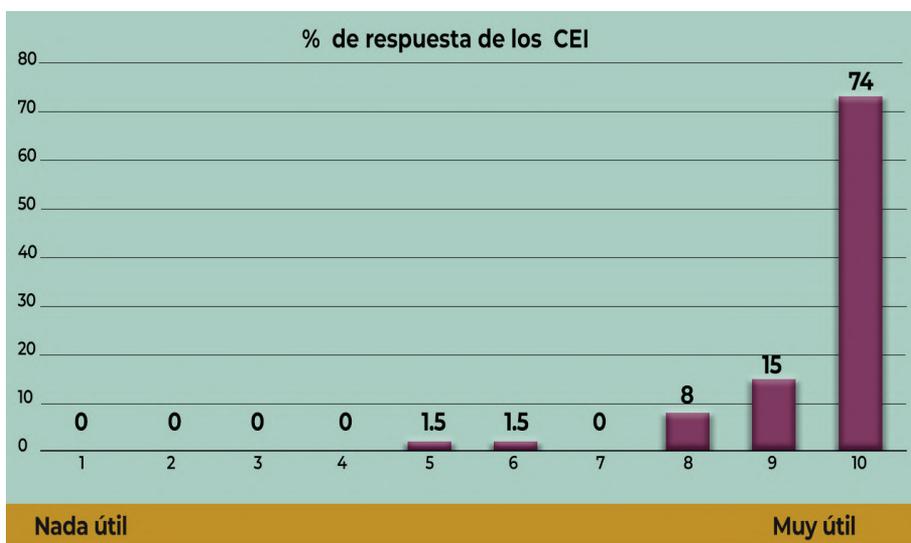


Figura 4. Percepción de la discusión y análisis en las sesiones conjuntas entre comités, en términos de porcentaje (N=61).

FUENTE: Comisión Nacional de Bioética, elaboración propia.

También, se indagó con los CEI su percepción respecto a la utilidad de las sesiones conjuntas contemplando seis preguntas, las respuestas a estas se describen en la figura 5.

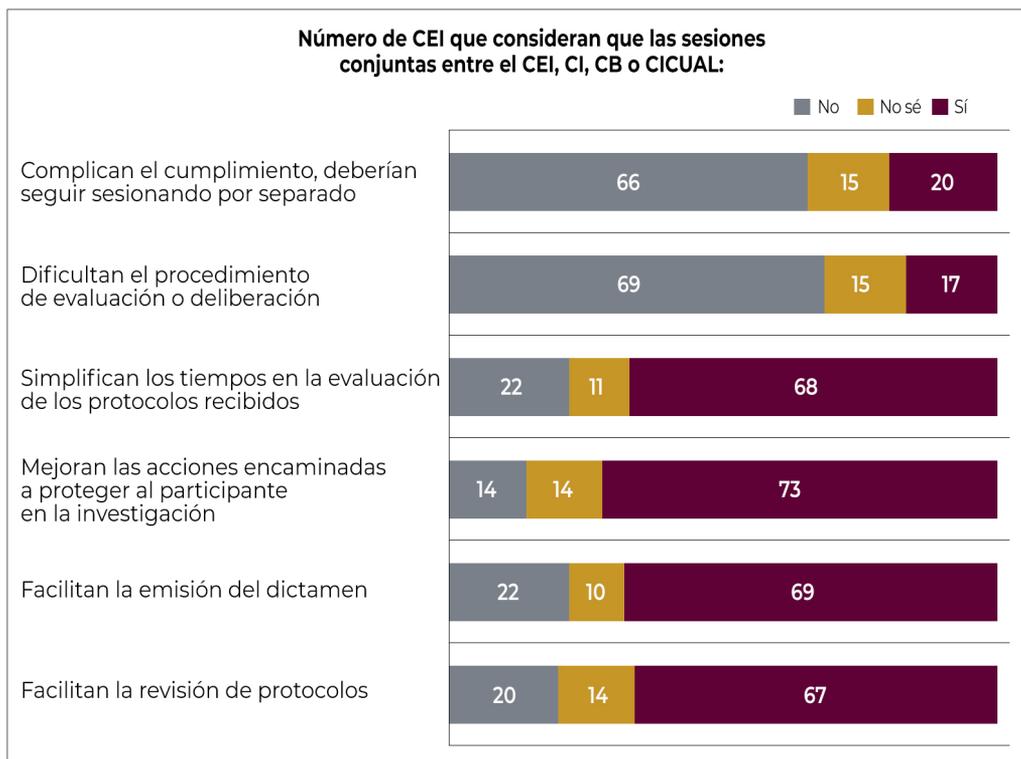


Figura 5. Percepción de utilidad de las sesiones conjuntas por los CEI (N=101).

FUENTE: Comisión Nacional de Bioética, elaboración propia.

Otro elemento importante por conocer fue: ¿Cuáles son las problemáticas que ha enfrentado o enfrentaría el CEI en la evaluación de protocolos en conjunto con el CI, CB y/o CICUAL? A continuación, se enlistan algunas de las problemáticas referidas por los CEI:

- Dificultad para reunir a todos los integrantes y/o el quórum requerido
- Divergencia de opiniones entre los comités
- Prolongación de la sesión
- Logística de la sesión
- Disponibilidad de espacio adecuado para sesionar en conjunto
- Delimitación de funciones entre los comités
- Actualización de los manuales de procedimientos y/o reglamentación interna
- Participación de personal directivo en los Comités de Investigación, en tanto que en los CEI no está permitido
- Potenciales conflictos de interés y coerción
- Comités de reciente creación
- Inexistencia de los otros comités en el establecimiento

Finalmente, se examinó la percepción de los CEI respecto a que las sesiones conjuntas sean obligatorias. En la figura 6 se representan las respuestas obtenidas.

juato, Guerrero, Hidalgo, Jalisco, Michoacán, Morelos, Nayarit, Nuevo León, Puebla, Quintana Roo, San Luis Potosí, Sonora, Tabasco, Tlaxcala, Veracruz y Zacatecas.

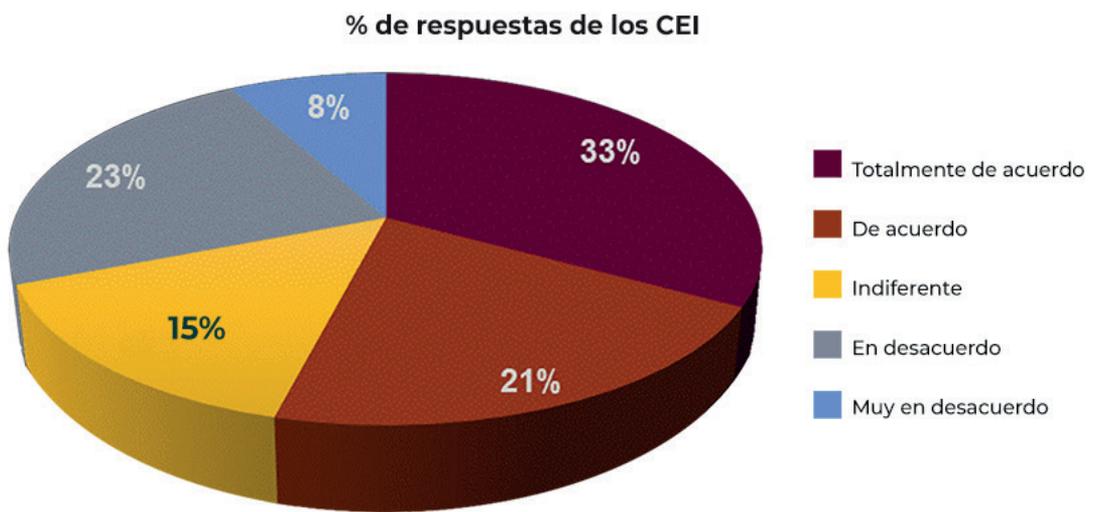


Figura 6. Percepción de los CEI respecto a que las sesiones conjuntas sean obligatorias, en términos de porcentaje (N=101).

FUENTE: Comisión Nacional de Bioética, elaboración propia.

Los resultados se dan a conocer únicamente con fines informativos, y al mismo tiempo, coadyuvar a visualizar la implementación actual de las sesiones conjuntas entre comités, con el objetivo de robustecer la evaluación ética, metodológica, de bioseguridad y normativa de los protocolos de investigación que se desarrollen en México.

B. ENCUESTA APLICADA A LAS CEB

La *Encuesta sobre sesiones conjuntas de Comités aplicada a las Comisiones Estatales de Bioética*, se implementó del 24 de abril al 02 de junio presente año. En dicha encuesta participaron 24 **entidades federativas**, lo que representa un 75% de participación respecto al total nacional. Las CEB participantes fueron Aguascalientes, Baja California, Campeche, Ciudad de México, Coahuila, Colima, Durango, Estado de México, Guana-

Es importante destacar que la representación de los integrantes de la CEB en la postura y respuestas brindadas se encontró que 36% se resolvió con mayoría simple, 20% por unanimidad de sus integrantes, 12% por mayoría calificada y 32% reportó otra situación. Con ello, 68% de las respuestas fueron discutidas por el equipo multidisciplinario que conforma a las Comisiones Estatales de Bioética, mientras que las CEB restantes reportaron casos en los que un solo representante fue el que contestó.

En relación con la percepción sobre las sesiones conjuntas, se destaca que al menos 75% de las CEB participantes (18) consideran positiva la realización de sesiones conjuntas en diferentes rubros, como: facilitar la emisión del dictamen (76%); simplificar los tiempos en la evaluación de los protocolos recibidos (76%); facilitar la revisión de protocolos (84%),

y mejorar las acciones encaminadas a proteger al participante en la investigación (88%). Además, se observa que 76% considera que las sesiones conjuntas no dificultan el procedimiento de evaluación o deliberación. Sin embargo, 68% de las CEB no tienen una postura definida sobre si las sesiones conjuntas complican el cumplimiento normativo y deberían seguir sesionando por separado. El porcentaje restante se divide equitativamente entre quienes consideran que deben sesionar por separado (16%) y quienes consideran que deben sesionar en conjunto (16%) (véase figura 7).

to de la deliberación (25%), perspectiva más amplia y plural (10%), unificar y complementar los comités y sus aportes (10%), disminución de quejas y conflictos entre los diferentes comités (5%), resultados con aplicación más amplia (5%), fortalecimiento del aprendizaje de los integrantes (5%), y aprovechar espacios (5%).

Sobre las problemáticas identificadas en la evaluación conjunta de protocolos con el CI, CB y CICUAL, se destacan aquellas de tipo logístico y administrativo como “Reunir a los

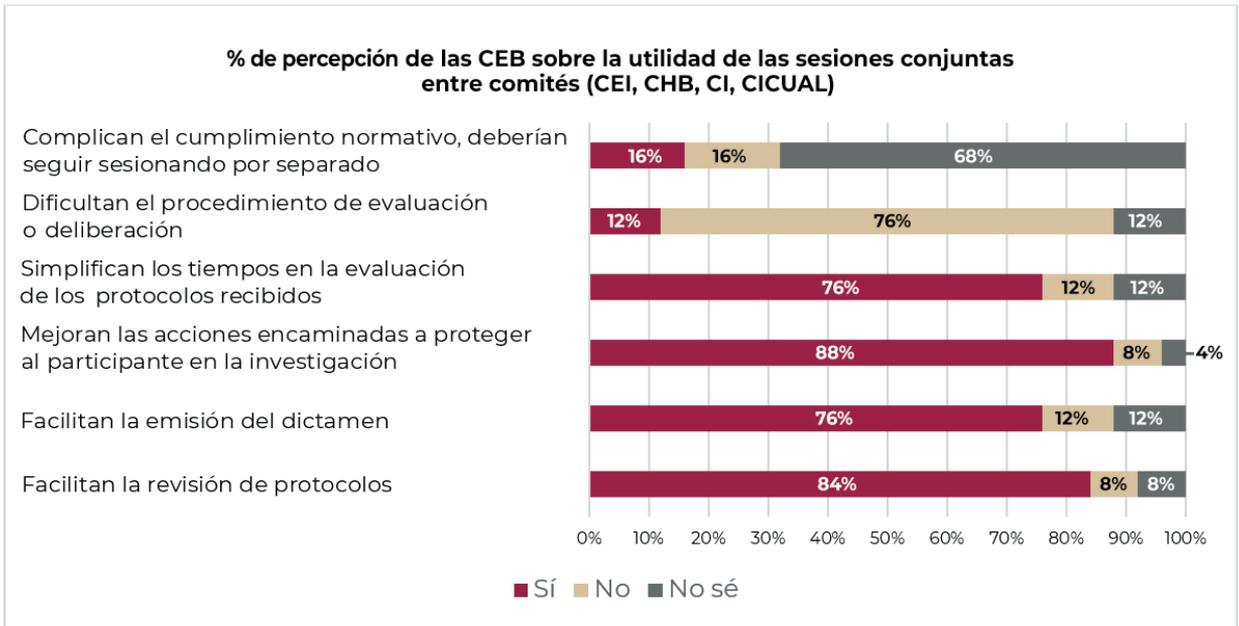


Figura 7. Percepción de las CEB sobre la utilidad de las sesiones conjuntas entre los diferentes comités.

FUENTE. Comisión Nacional de Bioética, elaboración propia.

En cuanto a los beneficios o dificultades que podrían derivarse de realizar sesiones conjuntas entre los diferentes comités, se encontró que 46% de las respuestas mencionaban beneficios, 39% dificultades y 15% restante no reportó escenarios adicionales a los enlistados en la encuesta. Los beneficios se englobaron en categorías como agilizar el dictamen y proceso (25%), enriquecimiento

miembros” con 48% de las respuestas, “Falta de criterios homologados para evaluación y operación” (24%), “Relaciones de poder entre comités” (24%), “Lineamientos y objetivos diferentes” (19%), “Conflictos al generar acuerdos” (19%), “Quórum de grupos multidisciplinarios y multiinstitucionales” (18%), “Personal compartido (duplicidad de funciones) en los diferentes comités” (12%), y “Tiem-

po de las sesiones" (7%). Otras problemáticas identificadas son de carácter sustantivo e infraestructural, como no contar con otros comités (4%) o no contar con espacios adecuados para llevar a cabo las sesiones entre los diferentes comités (4%). Además, se menciona la inasistencia de integrantes y falta de quórum en algunos de los comités, el personal compartido en CI y CEI, la rotación de personal y la falta de conocimiento de la normatividad.

Respecto a la propuesta de obligatoriedad de sesiones conjuntas entre CEI, CI, CB y CICUAL, 54% de los participantes está de acuerdo o totalmente de acuerdo. Por otro lado, 28% de los participantes está en desacuerdo con volver obligatorias las sesiones conjuntas debido a problemáticas identificadas que nuevamente incluyen aspectos logísticos y administrativos, y 8% muestra indiferencia respecto a esa modalidad de sesiones.

De igual forma, se recopilaron algunos de los comentarios y recomendaciones más desta-

cados de los participantes, como: la necesidad de regular este tipo de sesiones a través de un reglamento o disposición oficial; la consideración de que la realización de sesiones conjuntas beneficiará a los investigadores; la dificultad para lograr la asistencia de todos los miembros debido a que son participaciones honorarias; la importancia de coordinarse de la mejor manera posible; la necesidad de modificar la normativa nacional y trabajar en conjunto con la COFEPRI, y la autonomía de cada comité en su representatividad y funciones.

En suma, la encuesta revela que las sesiones conjuntas entre los comités tienen una alta aprobación por parte de las Comisiones Estatales de Bioética, sin embargo, persiste una falta de posicionamiento respecto al impacto en el cumplimiento normativo. Por lo tanto, es importante continuar la evaluación de la propuesta de las sesiones conjuntas para maximizar los beneficios y mitigar las dificultades identificadas.



Imagen: freepik.com

Encuentro Bioético

Conversación con Gabriela Ramos

En esta ocasión Encuentro Bioético sostuvo una interesante conversación con Gabriela Ramos, Subdirectora General de Ciencias Sociales y Humanas de la UNESCO, sobre Inteligencia Artificial y Ética



1. ¿Qué es la inteligencia artificial?

GR: El término Inteligencia Artificial (IA) se basa en la definición de la UNESCO como las tecnologías informáticas que, en una economía de datos, con la nube y con incremento de las capacidades de las computadoras, reproducen procesos asociados a la inteligencia humana como: el razonamiento, el aprendizaje y la adaptación, la comprensión sensorial y la interacción. Los recientes y rápidos avances de la IA han puesto de relieve las numerosas oportunidades que ofrece, así como la necesidad de una profunda reflexión ética sobre su utilización.

2. ¿Cuáles son los usos y aplicaciones comunes de la IA?

GR: Existen numerosos ejemplos de usos y aplicaciones de IA, que son cada vez más novedosos. En particular, con sus capacidades analíticas, la IA contribuye día a día a la toma de decisiones, desde los productos en internet, hasta quien obtiene un crédito,

o una entrevista de trabajo. Entre las aportaciones positivas, los modelos de IA están acelerando considerablemente el progreso científico. En 2022 contribuyeron a avanzar en la fusión de hidrógeno, a mejorar la eficacia de la manipulación de matrices y generar nuevos anticuerpos. Esto, con base en su capacidad de descubrir relaciones significativas en conjuntos inmensos de datos. Por ello, la IA también puede fortalecer la resiliencia frente a desastres naturales, diseñar edificios con eficiencia energética, mejorar el almacenamiento de energía y optimizar el despliegue de las fuentes de energía renovables.

Estas aplicaciones prometen una variedad de beneficios. Pero, lo cierto es que, sin reglas efectivas, existe el riesgo de que la IA cree nuevas desigualdades y amplíe las ya existentes. Por ejemplo, investigaciones recientes sugieren que los sistemas de IA pueden tener graves repercusiones sobre el medio ambiente. Según Luccioni, Viguier y Ligozat (2022), el entrenamiento del sistema BLOOM emitió 25 veces más carbono que un pasajero de avión en un viaje de ida de Nueva York a San Francisco. Al contrario, los nuevos modelos de aprendizaje por refuerzo como BCOOLER demuestran que los sistemas de IA pueden utilizarse para optimizar el uso de la energía.

3. ¿Cuáles son las recomendaciones de la UNESCO sobre la ética de la IA?

GR: La UNESCO tiene el mandato de asegurar que las tecnologías sean desarrolladas con una base ética, ya sea en temas del genoma humano, de la geo-ingeniería o de la inteligencia artificial. En este último rubro, en 2021, logramos la adopción por aclamación, de 193 países de **la Recomendación sobre la Ética de la Inteligencia Artificial, siendo este el primer marco universal. La Recomendación** aborda todos los riesgos y desafíos que surgen actualmente del des-

pliegue de las tecnologías de Inteligencia Artificial. Asimismo, busca maximizar los aspectos positivos. Este instrumento busca abordar el *qué de la ética* de la IA y el *cómo* de la acción política relevante, ofreciendo vías concretas para la realización del marco ético de principios y valores universales.

La Recomendación representa un marco ético integral que abarca todo el espectro de la arquitectura de derechos humanos -desde el principio de la dignidad humana hasta la cuestión de la inclusión, la equidad, la igualdad y la no discriminación- y lo hace manteniendo el enfoque en todas las etapas del ciclo de vida del sistema de IA. También lo hace a través de los principios de transparencia, rendición de cuentas, privacidad y equidad, entre otros.

En su diseño, estamos muy orgullosos de haber logrado la aprobación de principios tales como la no manipulación de los sesgos cognitivos, la determinación humana y, por consiguiente, en no aprobar la personalidad jurídica a los desarrollos de la IA. Prohíbe además el espionaje masivo y la notación social. Por otro lado, la Recomendación no pretende ser una mera declaración de principios, y por ello establece vías políticas para convertir los principios en acciones, reforzando el estado de derecho. Esto se puede lograr a través de los instrumentos de fortalecimiento de capacidades que ayudan a los países a aprovechar el poder de la IA.

Concretamente, la Recomendación incluye acciones políticas en áreas específicas: Evaluación de impacto ético, Gobernanza y rectoría éticas, Política de datos, Desarrollo y cooperación internacional, Medio ambiente y ecosistemas, Género, Cultura, Educación e investigación, Economía y trabajo, y Salud y bienestar social. Entre estas vías políticas incluimos un capítulo específico de género. Se trata del primer instrumento global con este compromiso.

La Recomendación cuenta con un marco de implementación, y la UNESCO se encuentra

aplicando un instrumento de diagnóstico, elaborado para este propósito (*Readiness Assessment Methodology*) que permite el seguimiento en 40 países. Con ello, trabajaremos hombro con hombro con los gobiernos para ayudarles a desarrollar las leyes e instituciones necesarias para normar la Inteligencia Artificial.

4.¿Cómo está cambiando la IA la forma en que se practica la medicina a nivel mundial?

GR: La IA tiene muchas ventajas en el reconocimiento de patrones, y por ello, puede mejorar sustantivamente los diagnósticos de enfermedades. Pero la IA también puede:

- Elegir, diseñar y planificar experimentos mientras mejora la medición y la observación.
- Generar hipótesis y aprender reglas científicas en áreas como la química para predecir la fabricación de medicamentos.
- Incrementar la exactitud de los diagnósticos médicos (cuando se combina con la opinión profesional).
- Predecir la replicabilidad de la investigación e incluso sugerir que los expertos y expertas revisen las propuestas de investigación.
- Identificar a los pacientes más adecuados para ensayos clínicos.

Hemos entrado oficialmente en la era de la inteligencia artificial y esto no es ajeno a la medicina. Todo esto, está lógicamente **transformando** la medicina. El mundo ahora está listo para cambiar a un ritmo no visto desde el despliegue de la imprenta hace seis siglos. La IA hace que nuestras vidas sean más fáciles, fluidas y ricas, nos ayuda a llegar a casa, informarnos, obtener créditos, conseguir un trabajo y hacer nuestros impuestos. También a diagnosticar, curar, tratar y prevenir enfermedades. Sin ir más lejos, el rápido desarrollo de la vacuna del COVID-19 fue po-

sible gracias a estas tecnologías. Ahora bien, en un área tan sensible como la salud, es imprescindible que estas tecnologías cuenten con un marco ético.

5. Respetto del uso actual de la IA, ¿será necesario implementar un marco normativo para su regulación?

GR: Éste es efectivamente el siguiente paso, en el cual estamos trabajando con los Estados miembros. La implementación de la Recomendación requiere de mecanismos concretos, y normas que se trasladan al marco legal vigente. Las tendencias recientes apuntan a una mayor actividad en la materia. En la Unión Europea, las directivas de IA comienzan a implementarse, y en todos los países hay una reflexión sobre como regular, incluyendo en los Estados Unidos.

Un análisis del Índice de IA de los registros legislativos de 127 países muestra que el número de proyectos de ley que contienen “inteligencia artificial”, que se convirtieron en ley, creció de solo 1 en 2016 a 37 en 2022. Además, un análisis de los registros parlamentarios sobre IA en 81 países muestra igualmente que las menciones a la IA en los procedimientos legislativos mundiales han aumentado casi 6,5 veces desde 2016. Pero la legislación no es suficiente. Necesitamos instituciones que aseguren el estado de derecho, y es ese el enfoque de la implementación de la Recomendación.

6. ¿Puede la IA contribuir a la equidad social?

GR: No sólo puede contribuir, sino que debe hacerlo. Este es el objetivo de los marcos éticos: evitar que se reproduzcan aún más las desigualdades que ya existen en nuestras sociedades y a nivel internacional, haciendo que la mitad del mundo no esté siempre conectado al internet, y no esté siempre representado en las bases de datos. Por ello, necesitamos una arquitectura digital que distribuya en forma más equitativa los beneficios de la creación colectiva de valor. En la actualidad, con dos países y un grupo de

empresas produciendo la totalidad de la IA, es necesario abordar las brechas.

Por otra parte, la IA bien utilizada puede apoyar a que los individuos y las comunidades reciban los servicios que necesitan tanto en salud y educación, y en tantos otros rubros. Se necesita una supervisión pública eficaz para que la digitalización y la IA generen oportunidades para la creación de valor público. Y para ello, **el uso público de la IA también debe estar asentado en una sólida base ética.** Frente al avance de esta tecnología como herramienta auxiliar para la toma de decisiones, es importante evitar un uso de los sistemas de IA que sea contrario a la democracia o a los derechos humanos.

En la UNESCO, creemos que para mantener el control sobre productos importantes y garantizar el respeto de normas éticas, **los gobiernos deben tener capacidad de desarrollar la IA sin depender del sector privado para la provisión de sistemas de carácter delicado.** También tienen que poder sostener un uso compartido de la información y la interoperabilidad de protocolos y métricas entre los diversos departamentos y ministerios. Todo esto necesita una inversión pública en las capacidades del Estado, según un enfoque basado en la idea de misión.

Y por ello en la UNESCO estamos trabajando a fin de fortalecer estas capacidades, para asistir a los Estados en la creación de la institucionalidad que necesitan para desarrollar IA.

7. ¿Cuál es el futuro de la IA?

GR: La reciente ampliación de las capacidades de la IA nos indica que el incremento puede continuar siendo exponencial. La tecnología de propósito general tendrá un impacto en todos los rubros. Pero a la vez, estos desarrollos conviven en un mundo desigual y lo refuerzan, y por ello el futuro de la IA será el que decidamos construir, con reglas claras y protección efectiva. Con una buena utilización, la IA podrían apoyar nuestra

comprensión del cambio climático (ODS 13), sistemas de transporte sustentables (ODS 11), tecnologías agropecuarias para ayudar a terminar con la desnutrición y pobreza (ODS 2), entre muchos otros beneficios.

Así, la IA se puede convertir en un aliado que el desarrollo sustentable requiere para diseñar, ejecutar, asesorar y planificar el futuro de nuestro planeta y su sustentabilidad más efectivamente. Esta combinación de IA con el desarrollo sustentable permitirá abordar las necesidades actuales sin comprometer las generaciones futuras.

Según recientes investigaciones (revista *Nature*), la IA podría convertirse en una herramienta clave para facilitar una economía circular y construir ciudades más inteligentes que utilicen sus recursos eficientemente. Un claro ejemplo del aporte de la IA es la gestión del tráfico. La aplicación de la IA en la movilidad urbana ha permitido predecir atascos y proponer rutas alternativas. Con movilidad compartida, esta tecnología predice la demanda de vehículos por zona y hora. Esto significa que las empresas pueden organizar la disponibilidad de vehículos para los ciudadanos en función de sus necesidades. Como consecuencia, esta solución no solo facilita la movilidad, sino que también minimiza su impacto ambiental.

Pero en su forma actual, y sin una buena regulación, la IA continuará reproduciendo y amplificando muchas de las brechas y los desafíos sociales a los que nos enfrentamos. Por ello tenemos que actuar. No se trata de la tecnología *per se*, sino de los marcos de gobernanza que configuran su desarrollo y uso, y los valores éticos y morales que defienden, y que son muy sensibles en el campo de la salud.

Estoy convencida de que, con un enfoque ético de la UNESCO puede ofrecer resultados justos, sostenibles e inclusivos. Pero esto no puede suceder sin gobiernos capaces que protejan el estado de derecho en línea, y desarrolladores públicos y privados que sean responsables de poner a las personas,

no a las ganancias ni a las consideraciones geopolíticas, en primer lugar. Sólo entonces la Era de la IA podrá traer el progreso que esperamos.

8. Respetto al 'CHAT GPT', ¿cuáles considera que son las implicaciones éticas en su aplicación en la educación superior?

GR: La salida del 'ChatGPT' al mercado no se hizo siguiendo las reglas éticas de la UNESCO en una manera *ex ante*. Una vez que estaba al alcance de todos, se empezaron a analizar sus impactos. Esto no debería ser así. Hay cuestiones claras de fiabilidad de los datos, de la actualidad de estos, y de los múltiples impactos que puede tener y que no se han medido. En la UNESCO realizamos un estudio el cual concluye que, con respecto a las provisiones de la Recomendación (en materia de transparencia, rendición de cuentas, resultados), el 'ChatGPT' no está alineado.

La integridad académica, por ejemplo, ha sido señalada como un área de preocupación, con mayores riesgos para el plagio, o el control de calidad. La falta de diversidad es otra área preocupante. Por último, si no nutrimos las capacidades críticas de los estudiantes, la utilización puede llegar a debilitar la excelencia o la actualidad de la producción académica.

La Recomendación incluye los siguientes elementos en materia de educación:

1. Una visión y prioridades estratégicas para todo el sistema.
2. Un principio general para la IA y las políticas educativas.
3. Una planificación interdisciplinar y una gobernanza intersectorial.
4. Políticas y normativas para un uso equitativo, inclusivo y ético de la IA.
5. Planes maestros para utilizar la IA en la gestión de la educación, la enseñanza, el aprendizaje y la evaluación.

6. Pruebas piloto, seguimiento y evaluación, y la creación de una base empírica.

7. Fomento de las innovaciones locales en IA para la educación.

La UNESCO también ha publicado orientaciones sobre la IA y la educación para los responsables que desarrollan políticas. En ella se describen las prácticas emergentes en educación y se analizan los retos que plantea el uso de la IA para alcanzar el Objetivo de Desarrollo Sostenible 4 (Educación de calidad).

Ahora bien, nuestra postura desde hace tiempo es que las evaluaciones de impacto ético de los productos de Inteligencia Artificial deben realizarse sistemáticamente “ex ante”, es decir, antes de que el producto salga al mercado. Es evidente que este consejo no se ha tenido en cuenta en los recientes lanzamientos al mercado, como el de ‘ChatGPT’ en noviembre de 2022, y ‘ChatGPT 4’ el mes pasado. Hoy en día, la Inteligencia Arti-

ficial está evaluando de manera errónea la elegibilidad de las personas para puestos de trabajo y prestaciones sociales. Identifica falsamente a las personas mediante la tecnología de reconocimiento facial. El ‘ChatGPT’ “alucina” con información falsa.

Por todo ello, en el Sector de las Ciencias Sociales y Humanas que dirijo, hemos avanzado los trabajos de la implementación de la Recomendación, con el desarrollo del Instrumento de medición de capacidades, y con la evaluación de impacto ético. Estamos aplicando en 40 países, y revisaremos las conclusiones en el Foro Global de la Inteligencia Artificial a celebrarse en Eslovenia el año que entra. Próximamente también lanzaremos el Observatorio de la Ética de la Inteligencia Artificial con el instituto Alan Turing y estamos trabajando con el sector privado para apoyar los desarrollos éticos. Con estos esfuerzos esperamos avanzar en la construcción de una IA más justa, solida, y confiable.



Imagen: freepik.com

Sugerencias de la CONBIOÉTICA en medios digitales

Ariana Leticia Landín López*

Película

Título: Blade Runner

Año de estreno: 1982

Duración: 1 hora 57 minutos

Dirección: Ridley Scott



Película de culto de 1982 que imagina un año 2019 con grandes adelantos en la Inteligencia Artificial: los “replicantes” parecen seres humanos, pero son creaciones artificiales. Tras una revuelta de estos superhumanos y ante el temor de que desarrollen ideas propias, les prohibieron regresar a la Tierra. Deckard, un Blade Runner retirado, es obligado a reincorporarse al servicio de esta especie de policías para atrapar a cuatro replicantes que osaron reingresar a la Tierra. ¿Cómo atraparlos si no se distinguen de los seres humanos? Algunos de los replicantes no son conscientes de su naturaleza. Sin embargo, otros saben lo que son y no se consideran computadoras sino entes físicos que piensan y, por lo tanto, existen. ¿Qué los impulsa a regresar a la Tierra? Algo tan fuerte como el deseo de vivir más tiempo, de alargar su existencia y para lograrlo deberán evadir a los Blade Runners y convencer a su creador de que les dé más tiempo, más vida. Sin embargo, la cuestión no será tan sencilla.

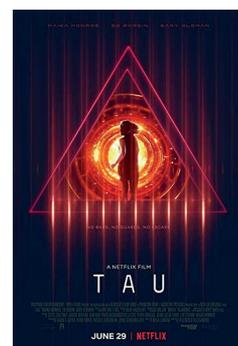
Película

Título: Tau

Año de estreno: 2018

Duración: 01 hora 37 minutos

Dirección: Federico D'Alessandro



Julia se da cuenta de que alguien está experimentando con ellos. Tras secuestrarlos, los encarcelaron y les colocaron un implante con luz en la nuca. Conscientes del peligro, intentan escapar y descubren que el espacio está resguardado por un avanzado robot con inteligencia artificial, Tau. Si el tomar decisiones a partir de un sentimiento es exclusivo de los seres humanos, en esta película se explora una posible revolución: a partir de algoritmos de inteligencia emocional, una IA puede demostrar comprensión humana. ¿Cómo se desarrollaría una tecnología con estas características? ¿Cuáles serían sus posibilidades, ventajas y riesgos? En términos de justicia, integridad y autonomía, ¿quiénes podrán tener acceso a estas tecnologías? ¿Cómo se obtienen los datos para desarrollarlas? ¿Qué sesgos tiene esa información? ¿Cómo colaborar con los participantes de las investigaciones? Tau enfatiza la importancia de realizar investigaciones con ética y responsabilidad.

*Subcoordinadora de Biblioteca, de la Comisión Nacional de Bioética.

Podcast

Nombre: Noosfera

Episodio: Noosfera 77. Dr. Robot y el machine learning en medicina | Francisco Jesús Martínez

Año de estreno: 2021

Duración: 1:07:25

Disponible: Apple Podcast, Google Podcasts, iVoox, Spotify y Spreaker



El vínculo entre la ingeniería y las ciencias de la salud, la computación y la medicina, ¿cómo funciona realmente? Francisco Jesús Martínez explica cómo se introduce el *machine learning*, en imágenes radiológicas, resonancias o encefalogramas. El aprendizaje automático o aprendizaje computacional se refiere a la creación de algoritmos para que las computadoras realicen tareas o “aprendan” a partir de ejemplos. ¿Cómo se puede utilizar el *machine learning* para mejorar la atención de la salud?

Conferencia

Nombre: Inteligencia artificial: sus desafíos éticos | Markus Gabriel

Año de estreno: 2023

Duración: 01:58:54

Disponible: YouTube



La Cátedra “Alfonso Reyes” del Tecnológico de Monterrey presentó la conferencia en la que Markus Gabriel desentraña las características de la Inteligencia Artificial (IA). ¿En qué consiste la ontología de la IA y por qué se debe abordar antes de hablar de la ética de la IA? ¿Hay una diferencia entre los sistemas de IA y la investigación en IA? Los sistemas de IA no son autónomamente inteligentes, sino que se integran y emplean en un sistema antropogénico mayor. Por lo tanto, no tienen una inteligencia intrínseca. Esto no quiere decir que la IA sea éticamente neutral, al contrario.



Imagen: freepik.com

Documental

Nombre: El cazador de cerebros

Episodio: Cuántica + IA= Futuro

Año de estreno: 2022

Duración: 29:54 minutos

Disponible: RTVE



¿Hacia dónde se dirige la computación? Visita la sede central de IBM y las instalaciones del MIT para conocer las investigaciones actuales. Los ordenadores utilizan propiedades de la física cuántica, mientras que los robots y vehículos han alcanzado un mayor nivel de autonomía y un mejor entendimiento del mundo exterior. En el ámbito de la salud, estas tecnologías ya pueden analizar grandes cantidades de datos y contribuir al diagnóstico de enfermedades. En el futuro, ¿cómo los avances digitales podrán resolver los problemas del mundo físico?

Análisis

Nombre: El uso de ChatGPT | Francesc Pedró

Director de UNESCO-IESALC

Año de estreno: 2023

Duración: 05:50 minutos

Disponible: YouTube



El director del Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe (IESALC), ofrece una reflexión sobre el uso de ChatGPT en la enseñanza. Esta herramienta impacta en los métodos tradicionales de enseñanza-aprendizaje, por lo que surge la necesidad de centrarse en el estudiante e impulsar el desarrollo de directrices y políticas claras sobre el uso de inteligencia artificial en la educación y el trabajo académico. Además, se vuelve urgente promover una alfabetización sobre la inteligencia artificial que acompañe a las enseñanzas académicas y digitales.

Sugerencias editoriales de la CONBIOÉTICA

*Karla Nallely Rosas Chelius**

Nanotecnología viva

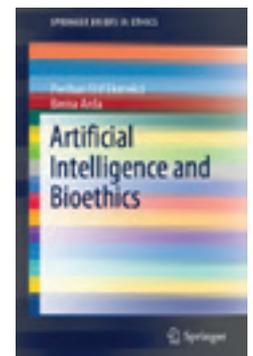
Sonia Contera
2023



La autora se basa en su experiencia como investigadora en el campo de la inteligencia artificial, describe las apasionantes formas en que la nanotecnología permite comprender, interactuar y manipular la biología. Se busca la reflexión sobre las oportunidades que emergen de los laboratorios para utilizar la tecnología con el fin de crear un futuro más justo y humano. La nanotecnología es un avance histórico que está revolucionando la medicina de formas que tendrán profundos efectos en la vida. Desde máquinas a nanoescala que pueden dirigirse a células cancerosas y administrar fármacos con mayor eficacia, hasta nanoantibióticos que combaten bacterias resistentes, pasando por la ingeniería de tejidos y órganos para trasplantes o la investigación en farmacología. Por ello se considera que el futuro traerá consigo la fusión de la nanotecnología con diversos campos de la salud y de vanguardia, esto a su vez permitirá contemplar las ventajas y los riesgos de este avance.

Artificial Intelligence and Bioethics

Perihan Elif Ekmekci & Berna Arda
2020

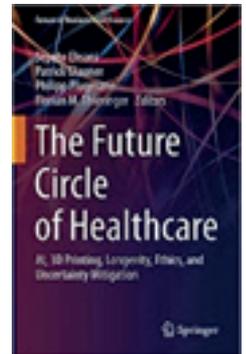


Esta obra aborda los principales desafíos al considerar los principios fundamentales de la ética médica, incluida la confidencialidad, la privacidad, la compasión, la veracidad y la fidelidad. Finalmente, los autores discuten las implicaciones éticas de la participación de agentes de inteligencia artificial en la atención al paciente al ampliar las habilidades de comunicación en un enfoque basado en casos. Asimismo, explora los principales problemas bioéticos que surgen del desarrollo y uso de la inteligencia artificial en entornos médicos. Los autores comienzan definiendo el pasado, el presente y el futuro de la inteligencia artificial en entornos médicos y luego proceden a abordar las preguntas bioéticas comunes y específicas resultantes. A su vez, analiza las investigaciones bioéticas en dos conjuntos separados: el primero se compone de debates ontológicos que se centran principalmente en la personalidad y en ser un agente ético de un artefacto. El segundo bloque aborda cuestiones bioéticas, derivadas del uso de la inteligencia artificial, se enfoca particularmente en el área del uso de inteligencia artificial en medicina y servicios de salud.

* Jefa del Departamento de Servicios de Información, de la Comisión Nacional de Bioética.

The Future Circle of Healthcare Ai, 3D Printing, Longevity, Ethics, and Uncertainty Mitigation

Sepehr Ehsani Patrick Glauner Philipp Plugmann
& Florian M Thieringet
2022



Esta publicación reúne a académicos en áreas de especialización esenciales para comprender cómo la atención médica podría cambiar y evolucionar durante la próxima década. ¿Qué lecciones se pueden extraer de los desarrollos actuales y pasados? Se abordan cuatro temas; el primero es el rápido ritmo y la ubicuidad de los avances tecnológicos en áreas como la inteligencia artificial, el aprendizaje automático, la fabricación aditiva y la electrónica portátil; el segundo pilar se refiere al envejecimiento saludable, la longevidad y el manejo de enfermedades crónicas; el tercero es el imperativo de permanecer consciente de las dimensiones éticas de las decisiones médicas, adaptando la bioética a los cambios en curso en la prestación de servicios de salud y el cuarto pilar se relaciona con cómo la incertidumbre en diferentes dominios del conocimiento médico puede mitigarse y traducirse a la práctica clínica. Los autores de los capítulos identifican los desafíos respectivos y las oportunidades prometedoras, discutiendo cómo estos podrían contribuir a visualizar el alcance futuro de la atención médica cuando se trata de proporcionar valores médicos, económicos y éticos a las sociedades humanas.

El dilema de utilizar “Chat GPT”: ¿Enemigo o aliado de la educación?

Eloy Albarrán
2023



El autor plantea una serie de consideraciones éticas y creativas que deben ser tenidas en cuenta al utilizar Chat GPT en la educación -un modelo de lenguaje que es capaz de generar texto de manera autónoma-. Hay una reflexión respecto a las posibilidades que ofrece el uso de Chat GPT en la educación, pero también sobre los riesgos que conlleva. A través de ejemplos concretos, se ilustran los dilemas éticos y creativos que surgen al utilizar esta tecnología en la educación. Por otro lado, se explora cómo la adopción de esta tecnología puede afectar la creatividad de los estudiantes, al depender en gran medida de un modelo preexistente, a medida que la tecnología avanza, cada vez son más las escuelas y universidades que adoptan herramientas de aprendizaje en línea que hacen uso de esta tecnología.

Manual práctico de inteligencia artificial en entornos sanitarios

Julo Bonis Sanz, Borja Rodríguez Vila, Juan José Beunza Nuin, Enrique Puertas Sanz y Emilia Condés Moreno
2023



Esta publicación tiene la voluntad de dar respuesta a todos los cambios esperables que el campo de la inteligencia artificial ha experimentado en estos dos años, donde se han desarrollado múltiples dispositivos y herramientas de inteligencia artificial. Por lo anterior, es urgente preparar a los profesionales de la salud y a los gestores involucrados en la adquisición, el uso, el diseño, la integración, etc. de estas herramientas. Este manual maneja un lenguaje sencillo para facilitar la comprensión de los conceptos básicos más relevantes y así facilitar la iniciación en el mundo práctico de la inteligencia artificial a personas que no disponen de ningún conocimiento previo. En el contenido se incluyen temas sobre los grandes tipos de algoritmos disponibles, qué es un algoritmo, aplicaciones concretas, el procesamiento del lenguaje natural, cómo se crean chatbots, la estructura y el funcionamiento del Internet de las Cosas, la importancia de la protección de datos y de los datos en streaming, la nube y las fuentes de datos, y, finalmente, de las claves del éxito de un programa de inteligencia artificial.

Filosofía de la medicina: discusiones y aportaciones desde México

Atocha Aliseda,
Cecilia Calderón,
Marcia Villanueva (Editoras)
2023



Esta obra es un fruto del diálogo entre profesionales de la Filosofía, Filosofía de la Ciencia, Medicina, Economía, Sociología y Antropología. Ofrece discusiones sobre aspectos filosóficos, epistemológicos y metodológicos de temas como el razonamiento clínico y las inferencias causales en las Ciencias de la Salud. También contribuye a la caracterización de la enfermedad, en relación con el sobrediagnóstico y la psiquiatría, y ahonda en temas como la deshumanización de la medicina, el sufrimiento al final de la vida y el pensamiento complejo en la práctica clínica. En suma, se trata de un libro colectivo que tiene como finalidad dar visibilidad a la filosofía de la medicina mexicana.

Conferencias magistrales 2023

Dialogando CONBIOÉTICA

31 de julio - 5 pm

The moral obligation to develop gene editing to enhance humanity

Julian Savulescu

Oxford Uehiro Centre for Practical Ethics

Comenta: Ma. de Jesús Medina Arellano
Instituto de Investigaciones Jurídicas, UNAM



3 de agosto - 12 pm

La recomendación sobre la ética de la inteligencia artificial

Dafna Feinholz-Klip

UNESCO

Comenta: Patricio Santillán Doherty
Comisión Nacional de Bioética

Liga de ubicación Auditorio Alfonso Caso:
<https://goo.gl/maps/JYhfUwptDWbzSBh18>



21 de septiembre - 12 pm

Conscience and Medical Professionalism: Positive and Negative Claims

Udo Schüklenk

Queen's University

Comenta: Rodrigo Ramos Zúñiga
Comisión Estatal de Bioética, Jalisco



Registro el día del evento, transmisión en vivo
Canal de YouTube CONBIOÉTICA



SALUD
SECRETARÍA DE SALUD



Integridad Científica



gob.mx/salud/conbioetica

- **2 de agosto**
Alfabetización en salud y atención primaria
Eduardo César Lazcano Ponce / *Instituto Nacional de Salud Pública*
- **9 de agosto**
Final de la vida en unidades pediátricas
Luis Enrique Juárez Villegas / *Hospital Infantil de México, Federico Gómez*
- **16 de agosto**
Principios bioéticos en la práctica pericial médica
Zoraida García Castillo / *Escuela Nacional de Ciencias Forenses, UNAM*
- **23 de agosto**
Turismo, ética y medio ambiente
Gino Jafet Quintero Venegas / *Instituto de Investigaciones Sociales, UNAM*
- **30 de agosto**
Retos de la protección de los derechos animales en México
Samuel León Martínez / *Programa Universitario de Bioética, UNAM*
- **6 de septiembre**
La dignidad como eje en la atención de los procesos de enfermería
Martha Eugenia Martínez Pinedo / *Consejo Estatal de Bioética, Zacatecas*
- **13 de septiembre**
Integridad científica y ética de la investigación: diferencias y sinergias
Irene Córdova Jiménez / *Universidad de Guadalajara*
- **20 de septiembre**
Recursos para la educación en bioética
Ana Carolina Sepúlveda Vildósola / *Facultad de Medicina, UNAM*
- **27 de septiembre**
Filosofía del dolor y cuidados paliativos
David Fajardo Chica / *Instituto de Investigaciones Filosóficas, UNAM*
- **4 de octubre**
Causas biológicas, humanas y ambientales de las pandemias
Ana María Carrillo Farga / *Facultad de Medicina, UNAM*
- **11 de octubre**
Principios éticos y habilidades en la atención de la enfermedad mental
María Soledad Rodríguez Verdugo / *Hospital San José Hermosillo*
- **18 de octubre**
Bioética y salud colectiva
David Benítez Valladares / *Comisión Estatal de Bioética, CDMX*
- **25 de octubre**
Aspectos éticos de la APS en la salud pública
Adolfo Javier Romero Garibay / *Universidad de la Salud*

Se dará constancia de capacitación por una hora, por videoconferencia
Transmisión por canal de YouTube. Liga de evaluación en la descripción

