

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