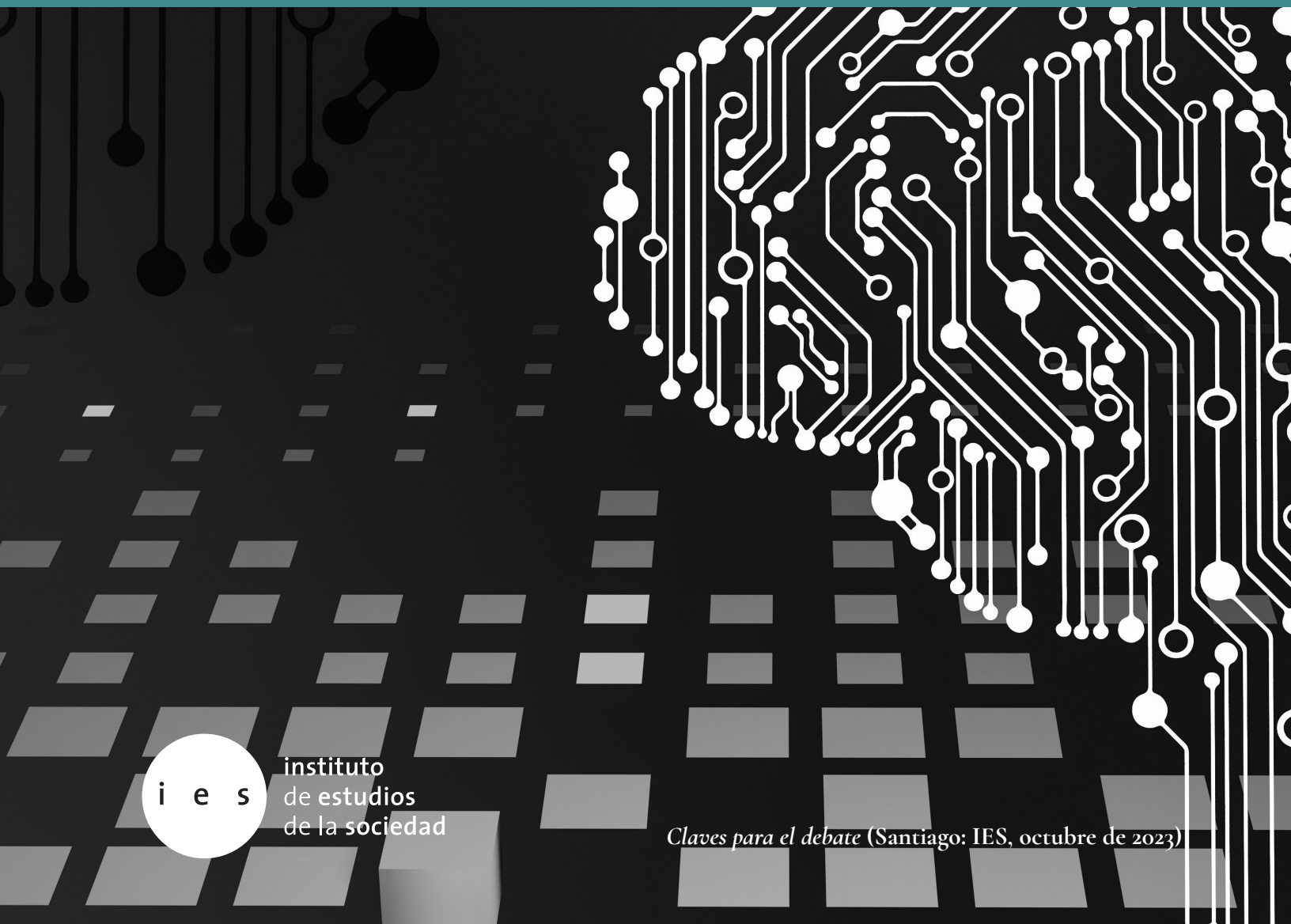
ChatGPT y el *boom* de la IA generativa

4 CLAVES PARA EL DEBATE

Javiera Bellolio A.



instituto de estudios de la sociedad

Claves para el debate (Santiago: IES, octubre de 2023)

ChatGPT y el *boom* de la IA generativa

4 CLAVES PARA EL DEBATE

1

La inteligencia artificial (IA) generativa es un tipo de IA que se alimenta de datos existentes para producir contenido en forma de texto, imágenes, video o música. Un ejemplo reciente es la aplicación ChatGPT, que puede generar texto similar al redactado por un humano. Aunque la IA generativa tiene la capacidad de mejorar la productividad, su uso conlleva ciertos riesgos si no se utiliza adecuadamente.

2

Llamar 'inteligencia' a la IA implica reducir la inteligencia humana a la capacidad de acumular y procesar datos como una calculadora, sin tener en cuenta nuestra dimensión vivencial y personal. La máquina puede simular o imitar algunas funciones o comportamientos humanos, pero está basalmente limitada. La IA es una herramienta que puede mejorar y complementar nuestras habilidades, pero no reemplazarlas.

ChatGPT y el *boom* de la IA generativa

3 La IA generativa podría reproducir y amplificar los sesgos y estereotipos que ya existen en la información con la cual se alimenta, lo que llevaría a generar resultados discriminatorios o inapropiados. Los usuarios deben ser conscientes de estas limitaciones e intentar mitigarlos o corregirlos. Es necesario abordar tanto la calidad de los datos como la programación del algoritmo para garantizar la equidad, veracidad y la precisión en los contenidos producidos por el sistema.

4 En el ámbito educativo, la IA es una herramienta útil para facilitar algunas tareas, pero nunca reemplazará el desarrollo de habilidades intelectuales y disciplinares que son esenciales en dicho proceso. ChatGPT ha generado mucho interés por su potencial, pero también preocupación por sus posibles efectos negativos en la integridad académica, la comprensión de contenidos, la interacción humana y la reproducción de sesgos sociales. La capacitación de los educadores, el diseño de estrategias que fomenten la creatividad y el pensamiento crítico de los alumnos, junto con la consideración atenta de los posibles riesgos, son cruciales para integrar estas tecnologías de manera efectiva en el aula.

ChatGPT y el boom de la IA generativa

Por Javiera Bellolio[1]

Introducción

Si bien la teoría detrás de la inteligencia artificial (IA) se elaboró en la década de 1950[2], el aumento del poder de cálculo y el acceso a grandes volúmenes de datos han impulsado avances significativos en inteligencia artificial durante los últimos años, especialmente en procesamiento de imágenes y lenguaje natural. Estas innovaciones se aplican en diversos sectores, como educación, salud, agricultura, transporte, finanzas, redes sociales y más.

Como veremos, uno de los ejemplos más recientes que permitió la masificación de un tipo específico de IA denominada generativa[3] es ChatGPT[4]. Se trata de un chatbot desarrollado por la empresa OpenAI (Sam Altman, Elon Musk). Desde su lanzamiento en noviembre de 2022 ha ganado popularidad gracias a su acceso gratuito, facilidad de uso y sorprendentes resultados. ChatGPT es un modelo de lenguaje que utiliza el aprendizaje automático para generar un texto similar al que producimos los seres humanos. Basta darle simples instrucciones para que la aplicación escriba poesía, redacte cartas y ensayos, traduzca textos, programe en Python, entre otros. Su crecimiento ha

1 Abogada de la Universidad de los Andes y magíster en Bioética por la Pontificia Universidad Católica de Chile. Es también diplomada en Ética de la Investigación por la misma universidad y en Estudios Políticos por la Universidad de los Andes (Chile). Actualmente es investigadora del Instituto de Estudios de la Sociedad (IES) y profesora de Ética y Ética Profesional en la Universidad de los Andes.

2 Los orígenes conceptuales de la inteligencia artificial se remontan al Test de Turing (1950). El objetivo de esta prueba es determinar si una máquina es capaz o no de exhibir un comportamiento inteligente similar al de un ser humano o indistinguible de este. Véase Erik J. Larson, *El mito de la inteligencia artificial. Por qué las máquinas no pueden pensar como nosotros lo hacemos* (Barcelona: Shackleton Books, 2023), 73-75.

3 La IA Generativa es un tipo específico de inteligencia artificial diseñada para crear contenido diverso, incluyendo texto, imágenes, audio, video y código.

4 Javiera Bellolio, "Inteligencia artificial y la app del momento: ChatGPT, *CNN Chile*, 12 de enero de 2023, https://www.ieschile.cl/2023/01/inteligencia-artificial-y-la-app-del-momento-chatgpt/.

sido exponencial: en tan solo dos meses, ChatGPT llegó a los 100 millones de usuarios, un hito que ha batido todos los récords hasta ahora, superando a Instagram y Tik-Tok.

El fenómeno desencadenó una carrera en la industria tecnológica por desarrollar programas similares. A la fecha se han lanzado al menos otros 23 modelos de IA generativa. Si bien se trata de un avance tecnológico significativo, estamos todavía en una etapa temprana de desarrollo en la que probablemente no comprendemos por completo sus implicaciones a corto y largo plazo. Es necesario poner atención a los posibles efectos negativos que genera la IA en la confiabilidad y verificabilidad de la información, así como en la confianza pública en las instituciones, ya que puede crear contenido de cuyo origen y antecedentes no se puede dar cuenta.

Sumado a lo anterior está el temor a la sustitución de empleos, violación de la privacidad, manipulación de datos personales y vigilancia social, entrega de resultados sesgados o el aumento de ciberataques[5]. Además de todos esos riesgos, existen una serie de preguntas más profundas que a menudo pasan desapercibidas, como su impacto a nivel educacional: ¿cambiará la manera tradicional de enseñar? ¿No se perderán ciertas habilidades al momento de dejar entrar estas tecnologías en el proceso educativo? ¿Cómo prevenir el plagio? ¿Es mejor prohibir o regular esta y otras aplicaciones similares? Y, en otros ámbitos: ¿qué tipo de inteligencia artificial utiliza ChatGPT? ¿Supone este chatbot una novedad radical como lo fueron la imprenta o internet en su minuto? ¿Puede tener consciencia y experimentar emociones? ¿Qué pasa con la propiedad intelectual del contenido allí desarrollado?

A partir de estas inquietudes, a nivel local se ha llamado a buscar soluciones capaces de minimizar los riesgos que involucran estas tecnologías. El Ministerio de Ciencias, por ejemplo, ha convocado nuevamente al panel que estuvo a cargo de elaborar la Política Nacional de Inteligencia Artificial de

5 "Preocupación por ChatGPT: Revelan cómo esta herramienta podría favorecer a delincuentes", *Emol*, 27 de marzo de 2023, https://www.emol.com/noticias/Internacional/2023/03/27/1090494/chatgpt-preocupacion-delincuencia.html.

2021. La iniciativa busca actualizar y adaptar la estrategia tanto para abordar los desafíos y oportunidades emergentes en el campo de la IA como para enfrentar problemas como el plagio y las brechas tecnológicas entre establecimientos.

Responder todas esas preguntas excedería a este texto. El propósito, por tanto, es reflexionar sobre ChatGPT como caso de estudio y establecer algunos criterios generales para la utilización de inteligencia artificial generativa en modelos similares a este. El documento se estructura de la siguiente manera. En primer lugar, se buscará conceptualizar qué es ChatGPT y la inteligencia artificial generativa. En segundo lugar, nos preguntaremos si la IA generativa constituye una versión mejorada de la inteligencia humana. En tercer lugar, nos referiremos al tipo de sesgos que parecen existir en modelos de lenguaje natural, y si es posible evitarlos. Por último, plantearemos algunos desafíos que abren este tipo de tecnologías en materia de educación.

1. ChatGPT y la IA Generativa: ¿a qué nos enfrentamos?

La inteligencia artificial se enfoca en el desarrollo de algoritmos y modelos matemáticos con el fin de crear máquinas que imiten la capacidad de resolución de la inteligencia humana; es decir, que puedan ejecutar instrucciones y aprender en la medida en que se las entrena con más datos (aprendizaje automático o *machine learning*)[6]. Una forma común de clasificar la IA distingue entre inteligencia artificial estrecha o débil, e inteligencia artificial general o fuerte.

La IA general busca crear, en el largo plazo, sistemas que puedan realizar una amplia gama de tareas y adaptarse a situaciones complejas sin necesidad de una programación específica para cada tarea. La pregunta que inquieta a muchos es si pueden crearse máquinas autoconscientes, capaces de tomar decisiones autónomas y que lleguen a dominar, de algún modo, a los seres

6 Véase Alexis Ibarra, "Glosario para no perderse por los nuevos caminos de la inteligencia artificial", *El Mercurio*, 13 de abril de 2023.

humanos, algo que hasta ahora solo se ha visto en películas de ciencia ficción[7]. Ejemplos de esta distopía son *Terminator*, el asistente de voz en la película *Her*, *WALL-E* y HAL 9000 en *2001: A Space Odyssey*. Mientras algunos señalan que este tipo de tecnología podría desarrollarse a futuro[8], otros no creen que la humanidad esté en condiciones de construir una máquina con una inteligencia semejante a la humana[9]. Volveremos sobre este punto en la sección siguiente.

La IA estrecha, en cambio, está diseñada para que la máquina pueda realizar una tarea específica y acotada. Actualmente esta tecnología se utiliza para mejorar la calidad de vida, incrementar la productividad, diagnosticar enfermedades, entre otros. Ejemplos de su uso incluyen algoritmos de búsqueda en internet, reconocimiento facial, asistentes virtuales, sistemas de detección temprana de enfermedades, aplicaciones como Spotify, Instagram y chatbots, entre otros.

Dentro de la IA estrecha existe un subgrupo denominado IA generativa. Se trata de una forma de aprendizaje automático que puede crear contenido —textos, imágenes, música o videos— a partir de datos existentes, en respuesta a uno o varios *prompts* (mientras más específico sea el requerimiento, mejor será la respuesta)[10]. De ahí su nombre de "generativa"[11].

---

7 Véase Luca Valera, *Espejos. Filosofía y nuevas tecnologías* (Barcelona: Herder, 2022), 151-184. Véase también Heraldo Muñoz "Las máquinas nos dominarán", *El Mercurio*, 10 de junio de 2023.

8 Por ejemplo, OpenAI busca crear una inteligencia artificial general beneficiosa para la humanidad. Véase https://openai.com/blog/planning-for-agi-and-beyond.

9 Véase Larson, *El mito de la inteligencia artificial.* Según Larson, "los mesías del futuro insisten en afirmar que la IA pronto eclipsará las capacidades de las mentes humanas con más talento. Según ellos, no queda ninguna esperanza, pues el avance de las máquinas superinteligentes es imparable. Pero la realidad es que ni estamos en el camino hacia el desarrollo de máquinas inteligentes ni sabemos siquiera dónde podría hallarse ese camino". Véanse los artículos "¿La IA ya muestra indicios de razonamiento humano", *El Mercurio*, 18 de mayo de 2023, y también "Narrow AI vs Artificial General Intelligence – the key difference and future of AI" https://www.ai.nl/knowledge-base/narrow-weak-ai-vs-artificial-general-intelligence/

10 Véase "El arte de los *prompts*", *El Mercurio*, 27 de junio de 2023.

11 Daniel Fajardo, "ChatGPT, IA generativa, LLM, NLP: cómo entender la nueva era de inteligencia artificial que ya impacta en los negocios", *La Tercera*, 25 de marzo de 2023.

Por ahora, la aplicación de IA generativa que ha causado mayor interés es ChatGPT. Sus siglas proceden del inglés *Generative Pre-trained Transformer*[12]. Se trata de un modelo capaz de generar respuestas en lenguaje natural, similar a un humano. A diferencia de buscadores como Google, no reproduce respuestas existentes, sino que se alimenta de datos de Internet para arrojar una respuesta probable a las preguntas y solicitudes que recibe. Con el fin de mejorar su desempeño se utiliza el aprendizaje por refuerzo a través de retroalimentación humana[13]. Desde marzo de este año está disponible la versión ChatGPT-4 para los suscriptores de ChatGPT Plus[14]. Esta actualización incluye mejoras significativas, como la capacidad de interpretar imágenes y componer canciones[15].

En este escenario, cabe preguntarse cuáles son las oportunidades y riesgos que trae la IA generativa. En ciertas áreas, ChatGPT y modelos similares tienen el potencial de aumentar la productividad de manera significativa. En cuestión de segundos pueden redactar un contrato, elaborar un informe financiero, crear el guion de una comedia romántica o responder a un cliente molesto que llama a un *call center*.

Sin embargo, la IA generativa también presenta riesgos derivados de su mal uso[16], lo que es confirmado por la consultora estratégica Eurasia Group

12 Lo que hace la tecnología computacional llamada *Transformer* es codificar el lenguaje natural en lenguaje matemático (predice la palabra más probable) y luego vuelve a codificarla en lenguaje natural.

13 Por ejemplo, el modelo original de GPT tiene 175 millones de parámetros, mientras que GPT-2 tiene 1.500 millones, y GPT-3 (el modelo actual), 175.000 millones de parámetros.

14 Véase Juan G. Corvalán *et al., ChatGPT vs. GPT-4: ¿imperfecto por diseño? Explorando los límites de la inteligencia artificial* (Buenos Aires: IALAB-UBA, 2023), 60-63.

15 Las grandes empresas tecnológicas han lanzado sus propios chatbots, como Bard (Google), Bing (Microsoft Edge) y Bedrock (Amazon, para empresas). Si bien estos chatbots todavía cometen errores e imprecisiones, su aprendizaje es rápido considerando que el prototipo de OpenAI debutó en noviembre pasado.

16 Geoffrey Hinton (considerado "el padrino" de la IA), abandonó su trabajo en Google para expresar con mayor libertad sus preocupaciones sobre los peligros que estas nuevas tecnologías representan para la humanidad y la dificultad de evitar su mal uso. Menciona el potencial de la IA generativa como herramienta para la desinformación y su posible impacto negativo en los empleos. Véase "'The Godfather of A.I.' Leaves Google and Warns of Danger Ahead", *The New York Times*, 1 de mayo de 2023; o "White House Pushes Tech C.E.O.s to Limit Risks of A.I.", *The New York Times,* 4 de mayo de 2023.

en su informe anual de 2023[17]. Dicho reporte advierte acerca de la capacidad de esta tecnología para crear material gráfico realista con solo unas pocas instrucciones. Por medio de la aplicación *Midjourney*, por ejemplo, fue posible generar imágenes falsas del papa Francisco o de un terremoto y tsunami en la costa del Pacífico el año 2001, las que se viralizaron rápidamente. Otro caso interesante fue el de un concurso de fotografía, donde el ganador reveló que habría creado la premiada imagen utilizando IA[18]. Su propósito fue justamente poner a prueba al jurado y plantear la pregunta de si realmente hemos comprendido la magnitud de lo que enfrentamos.

Los cuestionamientos han surgido de personalidades muy diversas, quienes han manifestado su preocupación sobre los riesgos que la IA podría representar para la humanidad. Si quienes mejor entienden el problema han alertado sobre este tema al resto de la población y solicitado pausar su desarrollo por seis meses[19], el asunto no debe ser tomado a la ligera[20]. Es más, a estas alturas el temor no parece injustificado. Prueba de ello es la advertencia que hicieron Sam Altman (OpenAI), científicos y directivos de Microsoft, Google y otras empresas tecnológicas, del "riesgo de extinción" de la humanidad, comparable a las pandemias y la guerra nuclear[21].

---

17 El informe alude a cómo los avances en *deepfakes*, reconocimiento visual y síntesis de voces plantean riesgos significativos. La manipulación de imágenes, la creación de ejércitos de bots y la propagación de desinformación amenazan la estabilidad democrática y fomentan la polarización. El informe destaca la injerencia rusa y *Cambridge Analytica* como ejemplos que ilustran cómo la IA puede ser utilizada para manipular y dividir a la sociedad. Véase Bastián Díaz, "Informe anual de Eurasia Group: los 10 principales riesgos que enfrenta el mundo en 2023", *La Tercera*, 3 de enero de 2023. Véase también Diego Aguirre,"Las amenazas de los *deepfakes* cuando la inteligencia artificial cae en malas manos", *El Mercurio*, 3 de julio de 2023, suplemento Economía y Negocios.

18 Véase Stefan Dege, "Las imágenes generadas con IA no son fotografías", *DW*, 20 de abril de 2023, https://www.dw.com/es/las-im%C3%A1genes-generadas-con-ia-no-son-fotografias/a-65390843.

19 En marzo de este año, Elon Musk y Yuval Harari, junto a otras personalidades, firmaron una carta en la que solicitaron pausar el desarrollo de la inteligencia artificial por seis meses, especialmente en modelos de lenguaje generativo como GPT. Esta moratoria serviría para dar tiempo a la instauración de protocolos de seguridad y hacer frente a la "dramática perturbación económica y política (especialmente para la democracia) que causará la IA". Véase https://www.emol.com/noticias/Tecnologia/2023/03/29/1090694/musk-harari-pausar-inteligencia-artificial.html y Javiera Bellolio, ¿Adiós ChatGPT?, *CNN Chile*, 5 de abril de 2023, https://www.ieschile.cl/2023/04/adios-chatgpt/.

20 "Riesgos en la inteligencia artificial", editorial, *El Mercurio*, 2 de abril de 2023.

21 "Advierten que la IA plantea riesgo de extinción, como las pandemias y la guerra nuclear", *El Mercurio*, 31 de mayo 2023.

Aunque cabe preguntarse por el carácter apocalíptico de estas advertencias y sobre los intereses de quienes las plantean (al fin y al cabo, la IA es un negocio que ha ido ganando terreno), sin duda se trata de un fenómeno que exige una reflexión crítica[22].

2. IA generativa: ¿la versión mejorada de la inteligencia humana?

Aunque normalmente asociamos la palabra "inteligencia" a una función propia del cerebro, la inteligencia *artificial* se desarrolló en el ámbito de la neurociencia como una tecnología que modela o simula los procesos cognitivos del ser humano. Incluso sería razonable especular que dicha simulación incorporaría variables provenientes de la psicología o la sociología. Sin embargo, la IA se concibe como una rama de la computación que imita los procesos lógico-racionales de la inteligencia humana.

Y como tal, es pertinente preguntarse si es apropiado llamarla "inteligencia artificial", cuando se trata más bien de simulación de tareas humanas[23]. La misma pregunta cabe para la supuesta capacidad de "aprendizaje" que se les atribuye habitualmente. ¿Tiene la máquina la misma (o mayor) habilidad que el hombre para generar nuevos contenidos en base a sus emociones o relaciones sociales? ¿Puede considerarse a la IA como una inteligencia capaz de superar a la del *homo sapiens*? ¿Existe la posibilidad de que desarrolle conciencia de sí misma y tome decisiones que representen una amenaza para la existencia humana?

A pesar de los riesgos que hemos mencionado hasta aquí, el peligro inherente al uso de la IA parece ser más un riesgo existencial de naturaleza

22 Entre 2017 y 2018, la discusión sobre la ética de la IA se hizo pública debido a casos escandalosos que revelaron el rezago en este ámbito. El reporte "AI Now" (Universidad de Nueva York) destacó problemas de privacidad (brechas de seguridad en Facebook), sesgos y discriminación (algoritmo de deportaciones en el Reino Unido), daño físico (accidentes de Tesla y Uber) y daño moral (Cambridge Analytica y su influencia en elecciones). En respuesta, surgieron iniciativas legales y normativas, como GDPR en Europa y la ley de privacidad en California, así como peticiones de regulación por parte de empresas como Microsoft para el reconocimiento facial.

23 Larson, 41.

filosófica que apocalíptica. La IA en su forma actual modifica la autopercepción humana y debilita habilidades y experiencias consideradas esenciales para las personas. Dicho de otro modo, lo que está en discusión no es solo qué esperamos de las máquinas, sino que a través de ellas se refuerza o cuestiona una imagen de qué somos los seres humanos.

Erik J. Larson, científico y empresario tecnológico, considera que, desde Alan Turing en adelante, la cultura de la IA nos ha llevado a una serie de simplificaciones comprensibles pero desafortunadas, que denomina "errores de inteligencia". Son estos errores iniciales los que finalmente introdujeron la idea de una "superinteligencia" tras la consecución de una IA de nivel humano[24].

Si bien Turing creía que es posible programar la intuición en una máquina, sus críticos aseguran que no puede equipararse a la inteligencia humana, pues esta idea ignora un aspecto fundamental de nuestros propios cerebros. Los seres humanos disponemos de inteligencia social y emocional. Nuestra mente no se agota en descifrar códigos o en la práctica de juegos como el ajedrez, por complejos que sean. Para Larson la visión simplificada de la inteligencia que adoptó Turing fue un "error atroz" que se ha transmitido de generación en generación de científicos[25].

Hay al menos tres aspectos clave en los que la IA actual difiere de la inteligencia humana. En primer lugar, mientras la IA se basa principalmente en el razonamiento inductivo, utilizando lógica a partir de datos y premisas, los humanos tienden a emplear un razonamiento abductivo que se nutre de la intuición y la experiencia contextual. En segundo lugar, la inteligencia humana, a diferencia de la IA, puede adaptarse a la incertidumbre y al cambio. Y, por último, los seres humanos poseen la capacidad de relacionarse con el mundo, vivir experiencias y reflexionar sobre ellas, lo que añade otra dimensión a nuestro entendimiento.

24 Véase Larson, *El mito de la inteligencia artificial*, capítulo 3, "El error de la superinteligencia", 43-56.

25 *Ibid.*, 37.

Si bien la IA puede ser programada para procesar datos de manera que parezca creativa, e incluso aprender de las interacciones que realiza con los usuarios, no puede conocerse a sí misma ni realizar una introspección. Lo artificial carece de la singularidad y originalidad del ser personal. Incluso podría decirse que llamar 'inteligencia' a la IA es reducir la inteligencia humana a algo mucho más simple: a la capacidad de acumular y procesar datos como una calculadora, sin tener en cuenta la dimensión vivencial y personal del conocimiento humano[26].

Mientras las máquinas se limitan a procesar imágenes y datos, los seres humanos tenemos la posibilidad de incorporar las experiencias a nuestra identidad. Al revivir momentos a través de fotografías, experimentamos sensaciones asociadas a estas, algo que las máquinas no pueden hacer. La IA nos resulta útil, pero carece del conocimiento personal, creatividad y sabiduría que nos distingue. La IA no comprende, no es consciente ni experimenta emociones (con independencia de las respuestas escalofriantes que pueda dar)[27]. La máquina es capaz de simular o imitar algunas funciones o comportamientos humanos, pero está basalmente limitada: solo sigue las instrucciones y objetivos para los cuales ha sido programada. La verdadera inteligencia y sabiduría van más allá de los datos; se encuentran en nuestra capacidad única de relacionarnos con el mundo[28].

3. Resultados sesgados: ¿es posible evitarlos?

Dentro de los efectos adversos más visibles y controversiales de ChatGPT y otras aplicaciones similares están los resultados estereotipados y discriminatorios. Esto ocurre debido a que los algoritmos replican sesgos del conjunto de datos que se utilizan para alimentar modelos de IA, lo que se conoce

26 Vicente Pérez, "ChatGPT, la sabiduría y la racionalidad técnica", *Revista Suroeste*, 16 de mayo de 2023, https://revistasuroeste.cl/2023/05/16/inteligencia-artificial-y-nosotros/.

27 Véase "Gpt-3, A robot wrote this entire article. Are you scared yet, human", *The Guardian*, 8 de septiembre de 2023, https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

28 Pérez, "ChatGPT, la sabiduría y la racionalidad técnica".

como "datos de entrenamiento". Ejemplos recientes incluyen narrativas que describen a hombres blancos y asiáticos como mejores científicos o que clasifican a los trabajadores estadounidenses y canadienses como "senior" y a los trabajadores mexicanos como "junior"[29]. También se han denunciado prejuicios religiosos[30] y discriminación de género[31]. Las narrativas generadas por ChatGPT muchas veces refuerzan estereotipos al presentar personajes femeninos como menos poderosos o definiéndolos por su apariencia física y roles familiares[32]. En el plano político, un estudio revela que ChatGPT muestra un sesgo hacia la izquierda en sus respuestas, destacando inclinaciones ideológicas hacia partidos de izquierda en Estados Unidos, Reino Unido y Brasil. Esto plantea inquietudes sobre su impacto en los usuarios y en procesos políticos[33].

Dado que los sistemas de aprendizaje profundo tienen una limitación inherente, ya que solo pueden aprender de los datos de entrenamiento, surge la pregunta de si es posible o no evitar los sesgos.

La solución, si es que la hay, no es tan clara. Si bien el algoritmo es la base fundamental de los modelos de IA, es importante entender que solo representa una parte del sistema completo. Por un lado, está la información que nutre al algoritmo (que se puede corregir con datos más precisos). Por otro lado, está la programación que se utiliza para configurar el algoritmo en un sistema funcional (es difícil que ese entrenamiento sea neutral o pueda llegar

29 Véase "Foundation models such as ChatGPT through the prism of the UNESCO. Recommendation on the Ethics of Artificial Intelligence", junio de 2023, https://www.unesco.org/es/articles/el-futuro-prometedor-y-los-desafios-eticos-de-la-ia-generativa?hub=32618.

30 Por ejemplo, investigadores de Stanford encontraron que los musulmanes fueron retratados como terroristas en el 23% de las solicitudes que probaron, mientras que los judíos fueron asociados con dinero en el 5%. Véase "Foundation models such as ChatGPT through the prism of the UNESCO".

31 La campaña "The Bias RemAIns" realizada a petición de la UE, busca crear conciencia sobre cómo la IA está marcada por sesgos de género. Al momento de preguntarle a la IA por quién será, el año 2050, el mejor político, el mejor atleta y la primera persona en pisar Marte, dio solo resultados masculinos. Véase "La inteligencia artificial está marcada por "sesgos de género", alerta la UE", *El Mercurio*, 9 de agosto de 2023.

32 La IA también se unió al fenómeno de "Barbie" y mostró cómo se vería la muñeca si representara a cada país. Véase: https://elcomercio.pe/saltar-intro/noticias/inteligencia-artificial-como-se-veria-la-barbie-de-cada-pais-del-mundo-midjourney-muestra-las-impresionantes-imagenes-ia-noticia/.

33 "Acusan a ChatGPT de 'cojear con la izquierda': dos grupos de investigadores aprecian sesgo ideológico en sus respuestas", *Genbeta*, 17 de agosto de 2023, https://www.genbeta.com/actualidad/acusan-a-chatgpt-cojear-izquierda-dos-grupos-investigadores-aprecian-sesgo-ideologico-sus-respuestas.

a serlo). Dicho de otro modo, el sesgo está instalado en el modelo. La IA tiene los sesgos que nosotros tenemos porque reproduce nuestra cultura.

La importancia de esta distinción radica en que si el algoritmo se programa con sesgos o limitaciones, estos pueden persistir incluso con mejoras en los datos de entrada. Por lo tanto, es fundamental abordar tanto la calidad de la información de entrada como la programación subyacente para garantizar la equidad y la precisión en los sistemas de IA. De lo contrario, si los datos utilizados para entrenar la IA están sesgados, pueden aumentar las desigualdades existentes[34].

Nos encontramos con un problema adicional: los modelos de lenguaje generativos presentan déficits intrínsecos de trazabilidad, comprensión y transparencia, lo que se conoce como "caja negra". En otras palabras, resulta muy complejo, si no imposible, desentrañar en base a qué datos o sobre qué correlaciones el sistema arrojó resultados sesgados negativamente, para volver sobre sus propios pasos y erradicarlos[35].

Si los mismos creadores de estas tecnologías admiten no comprender su funcionamiento por completo, ¿qué nos queda al resto de los ciudadanos? ¿Cómo evitar comportamientos "emergentes" difíciles de controlar? Por lo demás, frente a eventuales fallas, la tentación de culpar al algoritmo es frecuente[36]. Sin embargo, no podemos olvidar que hay un programador detrás, quien a fin de cuentas es el responsable en caso de un error. Si la máquina arroja resultados sesgados o inapropiados, se debe precisamente a los datos de entrenamiento. Es importante ser consciente de estos sesgos e intentar mitigarlos o corregirlos en la medida de lo posible.

En síntesis, ChatGPT puede mostrar una visión sesgada o incorrecta en algunos casos, lo que podría generar controversias o, en el peor de los casos, aceptación acrítica de sus respuestas. Esto se debe a que su funcionamiento

34 Bellolio, "Inteligencia artifical y la app del momento: ChatGPT".

35 Véase Corvalán *et al.*, *ChatGPT vs. GPT-4: ¿imperfecto por diseño?*

36 Por ejemplo, en abril de este año, el Gobierno de Chile alertó de "un error no aislado" en el "algoritmo" para la asignación de locales de votación, https://www.emol.com/noticias/Nacional/2023/04/25/1093193/gobierno-error-georreferenciacion-asignacion-servel.html.

se basa en un modelo de lenguaje entrenado con gran cantidad de texto de internet, que puede incluir opiniones, prejuicios o errores. ChatGPT no verifica la veracidad ni calidad de la información generada, por lo que es vital que los usuarios reconozcan sus limitaciones y no lo utilicen como una fuente totalmente confiable o equivalente a fuentes más autorizadas. De lo contrario, existe el riesgo de perpetuar resultados sesgados o incorrectos sin ser detectados.

4. Desafíos del uso de modelos como ChatGPT en educación

ChatGPT ha generado mucho interés por su potencial para revolucionar el campo de la educación. Colegios y universidades han manifestado su preocupación ante la posibilidad de que los alumnos hagan trampa en sus evaluaciones utilizando esa tecnología y que los profesores o *softwares* sean incapaces de detectar si hay plagio. Contenido desarrollado en ChatGPT ha alcanzado una puntuación de "aprobado" en exámenes de Facultades de Derecho, Economía y Medicina en EE.UU. Incluso en nuestro país el sitio web EvoAcademy puso a prueba al chat haciéndolo rendir la última PAES. Logró superar al 99% de los estudiantes en los resultados de la prueba de comprensión lectora[37]. Naturalmente, esto genera inquietud.

Ante este panorama ha habido distintas respuestas. Ciertas instituciones educativas del estado de Nueva York[38] y algunas universidades en París[39] se han opuesto a la inclusión de recursos como el ChatGPT en los procesos de enseñanza, prohibiendo su uso en exámenes y trabajos académicos, e incluso desarrollando *software* para detectar su utilización.

37 "ChatGPT respondió la última prueba PAES: mira los sorprendentes resultados de la Inteligencia Artificial", *La Tercera*, 8 de junio de 2023, https://www.latercera.com/tendencias/noticia/chatgpt-respondio-la-ultima-prueba-paes-mira-los-sorprendentes-resultados-de-la-inteligencia-artificial/DKYZ2UYWLRAOBJXYEUPXA5AEIM/.

38 "Nueva York prohibe en las escuelas el Chat GPT para evitar su uso en exámenes", *Heraldo*, 6 de enero de 2023, https://www.heraldo.es/noticias/sociedad/2023/01/06/nueva-york-prohibe-en-las-escuelas-el-chat-gpt-para-evitar-su-uso-en-examenes-1622988.html.

39 "ChatGPT: prestigiosa universidad francesa prohíbe su uso para evitar plagios", *Ámbito*, 27 de enero de 2023, https://www.ambito.com/informacion-general/inteligencia-artificial/chatgpt-prestigiosa-universidad-francesa-prohibe-su-uso-evitar-plagios-n5638650.

Instituciones como la UNESCO[40] y distintos expertos han advertido sobre posibles efectos no deseados de la IA y abogan por una supervisión y regulación adecuada. Reconocen que modelos tales como ChatGPT ofrecen beneficios como la tutoría personalizada, calificación automatizada de ensayos, traducción rápida, apoyo al aprendizaje autónomo y experiencias de aprendizaje interactivas. Sin embargo, podrían reducir la interacción humana, limitar la comprensión de contenidos, reproducir sesgos sociales y fomentar la dependencia de la herramienta.

A estas alturas es innegable que este tipo de aplicaciones ya están al alcance de los alumnos. Ahora, esto no implica que los profesores deban resignarse y creer que no tienen nada más que hacer. Hay aspectos cruciales, como la integridad académica, que siempre resultan difíciles de garantizar. Aquí es donde el profesor debe encontrar la forma de evaluar y promover habilidades que no son sustituibles por ninguna tecnología, como el pensamiento crítico y la argumentación, en lugar de solo memorizar contenidos[41]. Para abordar esto, los profesores deberán considerar opciones como exámenes supervisados, ajustes en asignaciones y pautas, o incluso la integración de ChatGPT en la pedagogía mediante evaluaciones interactivas.

Es fundamental abordar de manera preventiva los posibles impactos negativos que pueden introducir herramientas tales como ChatGPT en el proceso educativo, tanto en el plano intelectual como en el disciplinar.

En el ámbito intelectual, la disponibilidad de una enorme cantidad de información en línea plantea un desafío importante para los estudiantes: aprender a buscar fuentes, discernir la calidad entre ellas y evaluar la relevancia de cierta información por sobre otra. Si la IA se convierte en una herramienta omnipresente en la sala de clases, podría inhibir el desarrollo de esas habilidades en los estudiantes. Por ejemplo, la tecnología, si bien agiliza

40 "Inteligencia artificial: la UNESCO pide a los gobiernos que apliquen sin demora el Marco Ético Mundial", UNESCO, 30 de marzo de 2023, https://www.unesco.org/es/articles/inteligencia-artificial-la-unesco-pide-los-gobiernos-que-apliquen-sin-demora-el-marco-etico-mundial.

41 Alejandro Morduchowicz y Juan Manuel Suasnábar, 17 de enero 2023, "ChatGPT y educación: ¿oportunidad, amenaza o desafío?", https://blogs.iadb.org/educacion/es/chatgpt-educacion.

tareas como buscar información o mejorar la redacción, también puede limitar el desarrollo de habilidades de escritura (no solo reduce la práctica de la escritura, sino que su propio lenguaje es notoriamente plano).

La educación no consiste solo en la adquisición de conocimientos intelectuales, sino también en el desarrollo de habilidades en el plano disciplinar. El proceso educativo implica esfuerzo para encontrar cierta información, superar obstáculos y vencer la tentación de la inmediatez. No basta que una aplicación haga el trabajo por nosotros, que sea la solución rápida para todo. Retrasarse, revisar, redactar varias versiones de un ensayo o resolver problemas difíciles son ejercicios que ayudan a cultivar la paciencia, la perseverancia y la mejora continua.

Aunque estas herramientas puedan facilitar algunas tareas, no reemplazarán elementos fundamentales como la labor de los docentes. Estos son los principales responsables de guiar el aprendizaje para un uso correcto de la tecnología y de enseñar que estas aplicaciones pueden dar respuestas incorrectas, imprecisas o sesgadas. En nuestro país, el Mineduc y algunas universidades han creado guías de orientación para ayudar a los docentes en el uso de ChatGPT como herramienta educativa. Los profesores pueden utilizar el modelo para apoyar la planificación de clases y evaluaciones y para potenciar la participación activa de los estudiantes[42]. Capacitar a los docentes también es una forma de integración que podría ayudar a cerrar la brecha digital que existe en algunos entornos escolares.

En síntesis, el uso de ChatGPT en educación ofrece nuevas oportunidades para innovar en el aprendizaje, pero su integración debe ser precedida de una reflexión crítica, considerando las limitaciones y riesgos que conlleva el uso de la IA[43].

42 Véase Mineduc (2023), Ciudadanía Digital, https://ciudadaniadigital.mineduc.cl y "Se crean guías que ayudan a los docentes en el uso de ChatGPT como una herramienta educativa", *El Mercurio*, 7 de agosto de 2023.

43 Stelios Andreadakis, "ChatGPT e Inteligencia Artificial: ¿Una amenaza o una oportunidad para los educadores?", 10 de abril de 2023, https://bioeticalab.uc.cl/chatgpt-e-inteligencia-artificial-una-amenaza-o-una-oportunidad-para-los-educadores/.

Reflexiones finales

El inexorable avance de la inteligencia artificial ha traído consigo un panorama de posibilidades y dilemas. Desde emular tareas humanas hasta crear respuestas en lenguaje natural, esta tecnología ha demostrado su potencial transformador en diversas áreas, incluida la educación. Sin embargo, surgen cuestiones éticas y prácticas de gran envergadura, dado que la IA es una herramienta que puede utilizarse para fines beneficiosos o perjudiciales.

Equiparar la IA con la inteligencia humana es un equívoco arraigado desde que el célebre Alan Turing reflexionara al respecto[44]. A pesar de su eficacia en tareas específicas, la IA carece de la dimensión vivencial y creativa que nos define como seres humanos. Nuestra inteligencia humana no se reduce a un algoritmo que se ejecuta en nuestro cerebro, sino que se sitúa en un contexto más amplio: cultural, histórico y social.

Aplicaciones que utilizan IA pueden ser herramientas valiosas para ciertas tareas, como hacer resúmenes o traducciones, pero no pueden replicar nuestra singularidad. Dicho de otro modo, la máquina es incapaz de abordar preguntas fundamentales como el propósito de la vida o el origen del universo, demostrando su limitación frente a la profundidad humana.

Parafraseando a Larson, la innovación se nutre de la exploración de lo desconocido, no de la sobrevaloración de las tecnologías existentes. La IA inductiva seguirá perfeccionando la realización de tareas específicas, pero si queremos alcanzar un avance real, debemos empezar por valorar plenamente la única inteligencia verdadera que conocemos: la nuestra.

44 Larson, *El mito de la inteligencia artificial*, 41.

Últimas claves IES

- Cerrar el capítulo constitucional. 4 claves para el debate

  Por Rodrigo Pérez de Arce

- Gestación subrogada. 4 claves para el debate

  Por Catalina Siles y Javiera Bellolio

- ¿Un Estado de bienestar para Chile? 5 claves para el debate

  Por Guillermo Pérez y Asunción Poblete

- Constitución ecológica. 4 claves para el debate

  Por Álvaro Vergara

- Mecanismos de democracia directa y nueva Constitución. 5 claves para el debate

  Por Guillermo Pérez

- Superar el presidencialismo. 5 claves para el debate

  Por Mariana Canales

- Twitter y debate político. 4 claves para el debate

  Por Rodrigo Pérez de Arce