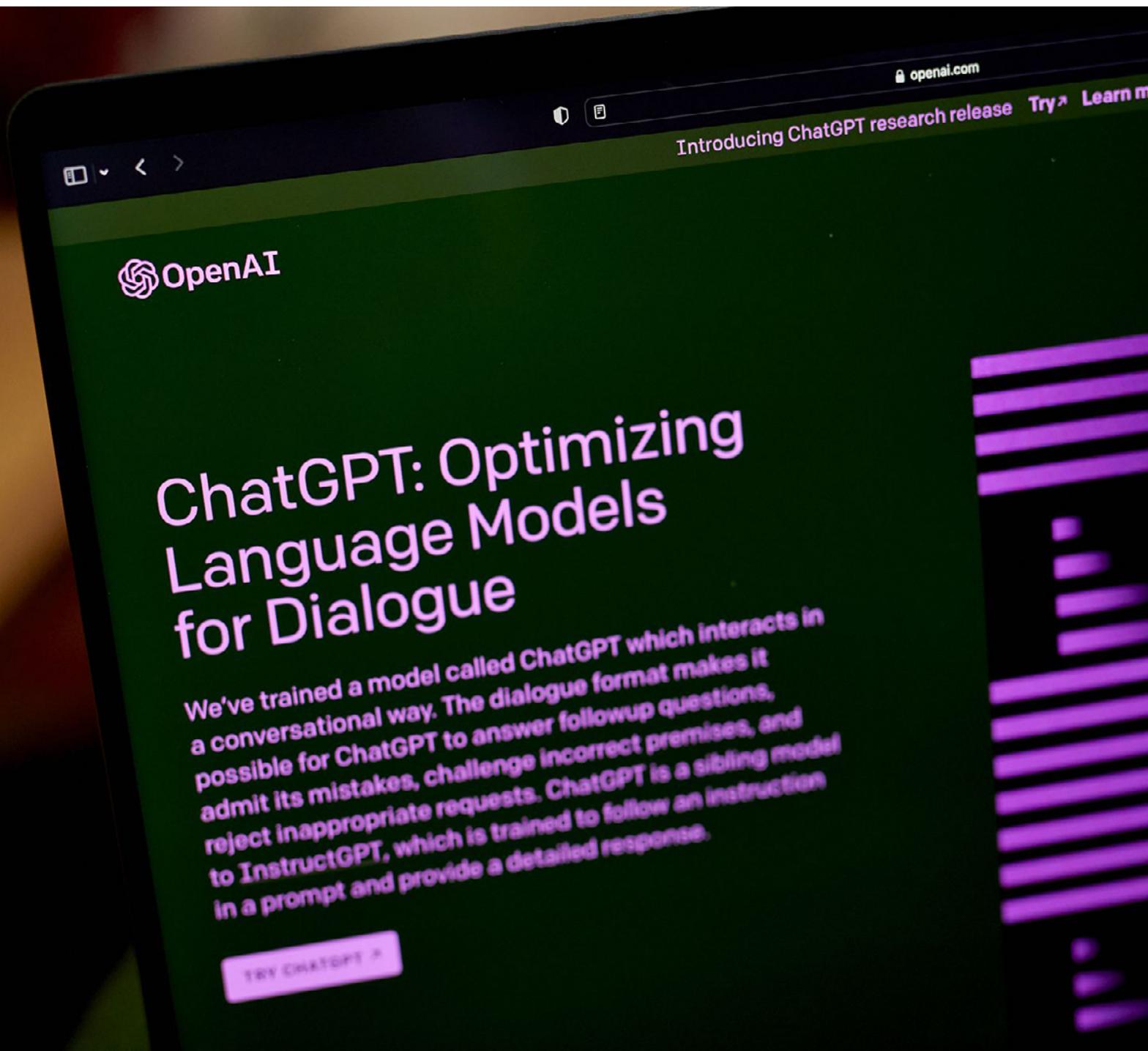


ChatGPT: Más preguntas que certezas*



El empleo de esta tecnología, que ha generado interés y preocupación en los últimos meses, abre una serie de interrogantes en ámbitos de las comunicaciones, la educación, la investigación académica y otros.

ALEJANDRO BARROS

Centro de Sistemas Públicos (CSP)
de Ingeniería Industrial, U. de Chile;
director de Espacio Público;
abc@alejandrobarrros.com

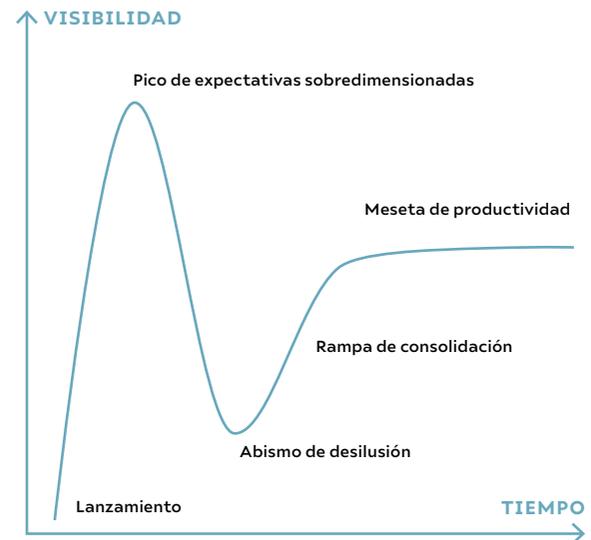
Desde que se lanzó a fines de noviembre de 2022, la aplicación ChatGPT se ha transformado en todo un éxito con su capacidad de responder a requerimientos de sus usuarios, redactando respuestas basadas en información contenida en la web. A la fecha, ya cuenta con más de cien millones de personas registradas, logrando así en dos meses algo que pocos habían logrado, muy por sobre lo que ocurrió en su momento con aplicaciones como Facebook o Twitter.

La aplicación ChatGPT ha generado múltiples reacciones, desde el mundo creativo hasta la educación, y al igual que la aplicación Dall-E (generadora de imágenes basada en Inteligencia Artificial, que causó también gran revuelo al aparecer hace unos meses), forma parte de lo que se conoce como *Generative AI* (IA generativa), tecnologías que permiten producir textos, imágenes y música en forma automática. El *software* utiliza modelos complejos de aprendizaje automático para predecir la siguiente palabra en función de secuencias de palabras anteriores o de la pregunta realizada, o bien puede configurar la siguiente imagen en función de palabras que describen imágenes anteriores.

Como en el pasado, cuando ha hecho irrupción una nueva tecnología, se produce un efecto que muestra diferentes percepciones, las que van desde la fascinación hasta el temor por los efectos nocivos que su uso pueda tener. Estas reacciones son un fenómeno bastante estudiado. Incluso la consultora Gartner le dio un nombre:

el ciclo y/o efecto hype¹, que muestra que las tecnologías emergentes pasan por cuatro estados hasta llegar a su nivel de madurez: lanzamiento, pico de expectativas sobredimensionadas, abismo de desilusión, rampa de consolidación y, finalmente, meseta de productividad. El paso por cada uno de estos estados puede durar bastante tiempo e incluso hay algunas tecnologías que mueren antes de llegar a la meseta de productividad. Según la consultora Gartner, las tecnologías IA generativa se encuentran ahora en la etapa más alta: «Pico de expectativas sobredimensionadas». No es posible asegurar cómo ni cuánto continuará el ciclo.

CICLO HYPE DE GARTNER



* Este artículo está basado en <https://www.alejandrobarrros.com/chatgpt-mucho-hype/> publicado el 19 de enero de 2023.

¹ Efecto Hype y Hype Cycle, Gartner - <https://www.gartner.es/es/metodologias/hype-cycle>

¿Qué pasa con ChatGPT?

Se trata de una tecnología desarrollada en 2022 por la empresa OpenAI², la misma empresa que desarrolló Dall-E³(generador de imágenes) en enero de 2021. La empresa lo define como la optimización de modelos lingüísticos para el diálogo.

¿Cómo funciona?

El modelo actual de ChatGPT está basado en la tecnología GPT-3⁴ (*Generative Pre-trained Transformer 3*), esto es, un modelo de lenguaje autorregresivo que utiliza aprendizaje profundo para producir textos que simulan la redacción humana sobre la base de un análisis de grandes volúmenes de datos.

Los grandes modelos lingüísticos (*Large Language Models*) como el GPT-3, se entrenan con grandes cantidades de datos de texto de Internet (miles de millones) y son capaces de generar textos similares a los humanos, aunque no siempre producen resultados coherentes. La función objetivo que utilizan es una distribución de probabilidades sobre secuencias de palabras, que les permite predecir cuál es la siguiente palabra de una secuencia: mientras más datos de entrenamiento se utilicen, mejor es la predicción. Lo primero a señalar es que no se trata de inteligencia en el sentido estricto. ChatGPT no entiende lo que lee ni, menos aún, razona. Por otra parte, estos modelos presentan algunos problemas que es bueno tener en consideración:

- i. El modelo inventa hechos inexistentes o erróneos, e incluso puede contradecirse en ciertos casos. En la gráfica siguiente se muestra un ejemplo⁵.

No se trata de inteligencia en el sentido estricto. ChatGPT no entiende lo que lee ni, menos aún, razona.

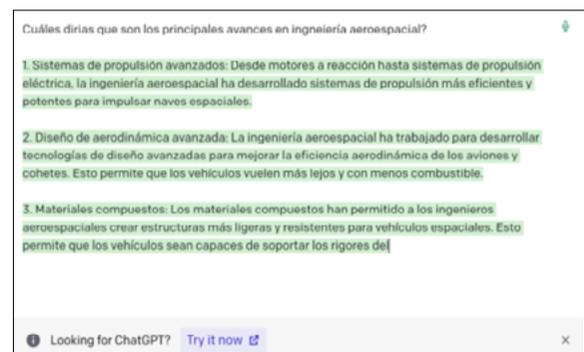
- ii. El lector del resultado no entiende cómo el modelo llegó a ese resultado/decisión. Como bien lo expresara uno de los mayores expertos nacionales en Inteligencia Artificial: «Sabemos que ChatGPT funciona 90% del tiempo. El problema es que no sabemos qué 90% y, peor, ChatGPT tampoco lo sabe», parafraseando una antigua frase de la publicidad⁷.
- iii. Entrega resultados diferentes en función del usuario que ha formulado la pregunta. Incluso si cambia de dispositivo, cambia la respuesta. Algo similar ocurre en Google. Esto nos recuerda el análisis de Google y Facebook hecho por Eli Pariser en su buen libro *The Filter Bubble*⁸.
- iv. Resultados sesgados o francamente tóxicos: un modelo lingüístico entrenado con datos sesgados puede reproducir sesgos en sus resultados.

Es pertinente destacar que el GPT-3 (versión actual) fue entrenado con 175 mil millones de ejemplos y que, actualmente, OpenAI se encuentra desarrollado una nueva versión basada en GPT-4, que está siendo entrenado con 100 billones de ejemplos. También veremos otras herramientas similares en el mercado, como es el caso de Google con su herramienta Dart, recientemente anunciada⁹.

¿EN QUÉ ÁREA NO SABES NADA?



¿CUÁLES DIRÍAS QUE SON LOS PRINCIPALES AVANCES EN INGENIERÍA AEROESPACIAL?



¿Que se viene por delante?

Desde un punto de vista de su evolución futura, múltiples autores de prestigiosas universidades plantean algunos análisis de la posible evolución de estos modelos. En su *paper* «Disociar lenguaje y pensamiento en los grandes modelos lingüísticos: una perspectiva cognitiva»¹⁰, los autores plantean que los grandes modelos lingüísticos (LLM) actuales generan de forma rutinaria párrafos de texto coherentes, gramaticales y aparentemente con sentido. Este logro ha llevado a especular con que estas redes son —o pronto serán— «máquinas pensantes», capaces de realizar tareas que requieren conocimiento y razonamiento abstractos.

En su análisis, los autores realizaron múltiples pruebas y sostienen que:

- i. Los LLM contemporáneos deberían ser tomados en serio como modelos de competencias lingüísticas formales;
- ii. Los modelos que dominan el uso del lenguaje en la vida real necesitarían incorporar o desarrollar no solo un módulo lingüístico básico, sino también múltiples capacidades cognitivas no específicas del lenguaje, necesarias para modelar el pensamiento.

Impactos en educación

Según varios especialistas, ChatGPT aún no es confiable. No comprende el mundo físico, no comprende el mundo psicológico e, incluso, alucina. Gary Marcus (ACM Digital Library) plantea que nos encontramos en el «Parque Jurásico de la Inteligencia Artificial»¹¹.

Lo que está claro es que esta tecnología no ha dejado indiferente a nadie, y hay discusión en diversos sectores acerca de cómo serán sus impactos y qué hacer al respecto. Revisemos algunos de ellos.

Algunos plantean que el uso de estas herramientas por parte de estudiantes va a poner en tela de juicio los trabajos y tareas que les solicitan sus profesores. Hoy los docentes y sistemas educativos plantean que van a tener que cambiar sus métodos (tareas, trabajos de investigación, evaluación, etc.). Ya hay universidades que están evaluando sus prácticas académicas, así como sus normas de conducta¹², partiendo por la definición de plagio que tienen. Algunas instituciones han actuado incluso adoptando medidas bastante inútiles, como prohibir el acceso a estas tecnologías, como hicieron escuelas del estado de Nueva York, o bien han ido más lejos, planteando que el ChatGPT va a matar los ensayos que los profesores les solicitan a sus alumnos.

Otros de los desafíos identificados en el área de la educación e investigación dicen relación con las citas. Con herramientas como estas, podríamos observar un nuevo fenómeno, cual es el de personas citando refe-

rencias que no existen, o incluso a un algoritmo. Es fácil hacer que ChatGPT entregue algunas referencias plausibles pero incorrectas. Incluso hoy se está dando la discusión de si ChatGPT puede ser citado como autor en la producción científica¹³.

Espero que la reflexión vaya en otra dirección. Por ejemplo, ¿de qué forma se puede trabajar en conjunto con ChatGPT? Lo plantea un artículo del NYT, «No prohíbas el ChatGPT en las escuelas. Enseña con él»¹⁴. O bien, dado que estas tecnologías ya están aquí y vamos a tener que convivir con ellas, se pueden modificar las formas de enseñanza y evaluación; por ejemplo, volviendo a métodos orales. En Japón, Todai Robot, proyecto de Inteligencia Artificial, realizó las pruebas de ingreso a la Universidad de Tokio, obteniendo resultados en torno al 20% de los mejor evaluados, sin entender nada de lo que estaba contestando. Esto muestra otro desafío muy significativo en la calidad de nuestra educación: ¿Cómo podemos ayudar a los niños a sobresalir en las cosas que los humanos siempre harán mejor que la IA?¹⁵.

La discusión y la presión de diferentes actores también ha llevado a que se desarrollen recursos para verificar si algún texto ha sido generado en forma automática, como es el caso de la herramienta GPTZero¹⁶. La misma empresa OpenAI, desarrolladora de ChatGPT¹⁷, en febrero lanzó una herramienta para verificar si un texto fue generado en forma automática. Ambas herramientas aún están verdes (hice algunas pruebas y su comportamiento no fue el mejor, aunque frente a un par de pruebas se comportó bien... ya veremos cómo evoluciona esta opción).

2 Ubicada en San Francisco, California, es una compañía de investigación en Inteligencia Artificial, fundada en diciembre de 2015 por los empresarios Elon Musk, Sam Altman, Ilya Sutskever, Greg Brockman y Wojciech Zaremba.

3 Dall-E - <https://openai.com/dall-e-2/>

4 GPT-3 - <https://es.wikipedia.org/wiki/GPT-3>

5 The emerging types of language models and why they matter, Kyle Wiggers, Techcrunch, 28-4-2022 - <https://techcrunch.com/2022/04/28/the-emerging-types-of-language-models-and-why-they-matter/>

6 Cabe señalar que ese ejemplo corresponde a una prueba que realicé en diciembre. El modelo ya responde en forma consistente.

7 Ricardo Baeza Y., Universidad Pompeu Fabra - <https://twitter.com/PolarBearby/status/1622358291522416640?s=20>

8 <https://www.alejandrobarrros.com/sabias-que-tus-busquedas-en-la-web-no-son-neutras/>

9 Dart - Google <https://blog.google/technology/ai/bard-google-ai-search-updates/>

10 Dissociating language and thought in large language models: a cognitive perspective, Kyle Mahowald, Anna A. Ivanova, Idan A. Blank, Nancy Kanwisher, Joshua B. Tenenbaum, Evelina Fedorenko, 16-1-2023 - <https://arxiv.org/abs/2301.06627>

11 AI's Jurassic Park Moment, Gary Marcus, Communications of The ACM, 12-12-2022 - <https://cacm.acm.org/blogs/blog-cacm/267674-ais-jurassic-park-moment/fulltext>

12 Alarmed by A.I. Chatbots, Universities Start Revamping How They Teach, Kalley Huang, The New York Times, 16-1-2023 - <https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html>

13 ChatGPT listed as author on research papers: many scientists disapprove, Chris Stokel-Walker, Nature, 18-1-2023.

14 Don't Ban ChatGPT in Schools. Teach With It., Kevin Roose, The New York Times, 12-1-2023 - <https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html>

15 Can a robot pass a university entrance exam?, Noriko Arai, Charla Ted, 2017

16 GPTZero - <https://gptzero.me/>

Profesiones con prerrequisitos de ejercicio

Se han realizado algunas pruebas con ChatGPT en exámenes de prerrequisito en algunas áreas que en varios países se encuentran reguladas, como son Medicina y Derecho, y ha habido resultados razonables.

En el caso de Medicina, se realizó una prueba con la examinación para obtener la licencia médica en Estados Unidos (*United States Medical Licensing Exam – USMLE*), que consta de tres exámenes. ChatGPT alcanzó o se acercó al umbral de los aprobados en los tres exámenes, y lo hizo sin ningún tipo de entrenamiento ni refuerzo especializado. En el caso de Derecho, también se han realizado pruebas con el proceso de examinación de la barra de abogados de Estados Unidos y exámenes de la carrera de leyes, con un resultado mediocre, pero aprobando el examen. Incluso en las pruebas finales de uno de los más prestigiosos MBA de Estados Unidos, ha habido resultados aceptables.

En el mundo del *marketing*

No todo puede ser impactos negativos. Hay argumentos respecto de que esto puede ser una herramienta muy poderosa para generar contenidos de *marketing* digital, responder preguntas frecuentes de clientes y promocionar productos, utilizando una menor cantidad de recursos¹⁸.

En la discusión pública

Otro ámbito en el cual el ChatGPT y herramientas similares van a tener un impacto muy relevante está en el debate público: por ejemplo, cuando se discuten políticas públicas (consultas ciudadanas), *lobby* y otros mecanismos. Esto importa, ya que estas herramientas pueden generar contenidos en forma automática a gran velocidad y así influir en el proceso de debate de cualquier política pública (leyes, normas, etc.). Además, podrán incidir en sesgos políticos, como lo han planteado algunos¹⁹. Otros, como Nathan Sanders y Bruce Schneider, ambos de Harvard, plantean que existe un gran riesgo de que el «ChatGPT pueda secuestrar nuestra democracia»²⁰.

Nuestras normas respecto del cabildeo (*lobby*) se encuentran modeladas en base a que esa actividad la hacen personas con nombre y apellido. Sin embargo, cuando el *lobby* lo hace un *software*, lo más seguro es que nuestras normativas queden totalmente obsoletas.

Medios de comunicación

Muchos medios al día de hoy están utilizando mecanismos del tipo IA generativa como ChatGPT, para generar contenidos. Tal es el caso de los medios digitales BuzzFeed

Podrían permitir crear contenidos con textos generados automáticamente, que pueden transformarse en herramientas muy potentes para la desinformación.

y CNET, haciéndolo tanto en forma intensiva como a nivel experimental, lo cual plantea varias interrogantes respecto de derechos de autor, rol de periodistas y editores entre otros²¹.

Otro aspecto preocupante es que estas herramientas podrían permitir generar contenidos (redes de sitios web) con textos generados automáticamente, que pueden transformarse en herramientas muy potentes para la desinformación y la generación de *fake news*.

El surgimiento de estas herramientas tecnológicas nos plantea una serie de interrogantes respecto de nuestras prácticas (políticas, educativas, laborales, comerciales, etc.), marcos normativos y procesos de generación de contenidos, y muchas aún no tienen respuesta convincente.

¿Dónde se encuentra usted, más en el lado tecno-optimista o tecno-pesimista? /M

17 AI Text Classifier (ChatGPT) – <https://platform.openai.com/ai-text-classifier>
 18 Cómo las marcas pueden usar ChatGPT en marketing de contenidos, Jesús Maceira, Mauna, 11-1-2023 – <https://www.maunamedia.com/como-las-marcas-pueden-usar-chatgpt-en-marketing-de-contenidos/>
 19 The Political Bias of ChatGPT – Extended Analysis, David Rozado, 20-1-2023 – <https://davidrozado.substack.com/p/political-bias-chatgpt>
 20 How ChatGPT Hijacks Democracy, Nathan Sanders Bruce Schneider, 15-1-2023, New YorkTimes – <https://www.nytimes.com/2023/01/15/opinion/ai-chatgpt-lobbying-democracy.html>
 21 BuzzFeed to Generate Online Content Using AI, Ben Wodecki, 27-1-2023, AI Business – <https://aibusiness.com/nlp/buzzfeed-to-generate-online-content-using-ai>