

## ChatGPT: furores y límites

Carlos A. Vargas Pacheco

Universidad Nacional Autónoma de México

### El furor por el ChatGPT

EN ENERO DE 2023 comenzaron a circular artículos periodísticos que daban cuenta de uno de los avances más sorprendentes de la Inteligencia Artificial (en adelante IA) hasta el momento. Se trataba de una nueva forma de *chat* automatizado (o *chatbot*) que, desde finales de 2022, causó furor debido a su capacidad de generar respuestas precisas respecto a cualquier tema, con un uso del lenguaje tremendamente parecido al de un ser humano. Dicha herramienta digital fue presentada con el nombre de *ChatGPT* (*Generative Pre-trained Transformer*), producida por la empresa digital *Open AI*.<sup>1</sup> De acuerdo con los especialistas, el *ChatGPT* es un modelo de lenguaje —es decir, una articulación algorítmica que emula el lenguaje humano en términos de organización sintáctica para adquirir comunicabilidad— capaz de generar discursos coherentes que se organizan en función de la pregunta o, en general, la escritura del usuario con quien interactúa. La precisión de las respuestas que brinda el *ChatGPT* se debe también a la inmensa cantidad de información que tiene dispuesta mediante su conexión continua a Internet. La fluidez, precisión y velocidad con la que opera el *ChatGPT* fue, ciertamente, motivo de admiración e, incluso, de orgullo por parte de quienes integraron el equipo de creación de este sofisticado modelo de lenguaje.

La fascinación que suscitó el *ChatGPT*, sin embargo, vino acompañada de profundas inquietudes, cuando no de francos temores. En efecto, por una parte, comenzó a circular la idea de que esta IA logra realizar tareas de manera tan eficiente, que muchas profesiones podrían ser sustituidas en el corto plazo. Por otra parte, en el terreno educativo, el *ChatGPT* mostró una enorme capacidad para generar textos académicos (ensayos, informes, proyectos, tesis, artículos e, incluso, libros) que podían hacerse pasar por materiales creados por personas, debido a su grado de precisión y semejanza con el lenguaje humano. Hubo un ejemplo concerniente al impacto que el *ChatGPT* podría tener en el ámbito educativo, que parecía extraído de alguna novela o película de ciencia ficción: el novedoso *ChatGPT* había logrado

<sup>1</sup> El *Chat GPT* tiene varias versiones. Actualmente, la versión 3.5 es gratuita y se puede acceder a ella a través de la siguiente dirección electrónica: <https://chat.openai.com/auth/login>. Lo único que se requiere es registrarse mediante un correo electrónico y la generación de una contraseña.



acreditar el examen de admisión a la Facultad de Derecho de la Universidad de Minnesota.<sup>2</sup>

Las posibilidades que se abren con el uso del *ChatGPT* han propiciado el debate en torno al control que los seres humanos pueden tener sobre la IA, en general. Es claro que el modelo de lenguaje en cuestión posee una funcionalidad potente, frente a la capacidad de manejo informático que la mente de un individuo puede lograr. Sin embargo, no debe perderse de vista el hecho de que las IA de procesos generativos —como el *ChatGPT*— son construcciones que se concentran en la capacidad de buscar y organizar información disponible exclusivamente en la *web*, por un lado, y articular coherentemente el lenguaje para transmitir sus resultados, por otro.

El *ChatGPT* es la versión más novedosa del delirio frankensteiniano por crear una entidad que, manteniendo algo de lo máspreciado del ser humano —información—, sea invulnerable a las múltiples afecciones de la humanidad. El sueño de una criatura inmune al dolor, desapegada del placer, incansable y carente de la necesidad de alimentarse o de beber es lo que pareciese hallarse en la base del entusiasmo y la efervescencia de la IA, en general, y del *ChatGPT*, en particular. La humanidad se halla ante una invención que procesa a velocidades sorprendentes la infinita gama de datos, información, opiniones y conocimientos que hombres y mujeres han

generado a lo largo de siglos y que ha quedado registrada en la inmensa base de datos de la *World Wide Web*. En este sentido, el contenido de las respuestas que brinde el *ChatGPT* no es otra cosa que conocimiento humanamente producido, pero sintetizado y presentado veloz y eficazmente.

El hecho de que esa tremenda creación se vincule con personas singulares torna desigual el encuentro, pues la máquina conlleva y procesa toda información posible, mientras que el individuo no lo puede hacer de la misma forma ni con la misma celeridad que el recurso artificial. Pese a esta desigualdad debe recordarse que la IA no posee conciencia propia, por lo que nunca podrá tener experiencia, volición ni emoción. El *ChatGPT* podrá generar —que no *crear*— un poema a través de plantearle un tema y de que rastree y emule los millones de poemas que transitan por la *Web*, pero no podrá tener nunca la experiencia vital que suscite la articulación de las palabras que establezca. La “inteligencia” de la IA es *inercia*, no vitalidad humana: la IA es un espectro de humanidad.

### **Diálogo con el *ChatGPT***

Para la elaboración de este artículo se ha procurado un ejercicio de uso del *ChatGPT*. A continuación se transcribe el intercambio con dicha IA:

**Carlos Vargas:** Me gustaría saber cómo surgiste y cómo funcionas.

<sup>2</sup> A pesar de lo sorprendente de esta noticia, también se enfatizó que esta IA acreditó el examen de admisión con el puntaje mínimo requerido, lo cual permitió reconocer que el *ChatGPT* es todavía limitado.

**GPT:** ¡Claro! Soy un modelo de lenguaje llamado GPT-3 (*Generative Pre-trained Transformer 3*), desarrollado por OpenAI. Fui entrenado en una amplia variedad de datos en línea hasta septiembre de 2021, lo que me proporcionó conocimientos en múltiples temas.

Mi funcionamiento se basa en algoritmos de aprendizaje automático y procesamiento de lenguaje natural. Utilizo una arquitectura de red neuronal llamada *Transformer*, que me permite comprender y generar texto de manera coherente y contextual.

Cuando interactúas conmigo, analizo tus preguntas y genero respuestas basadas en patrones y conocimientos previos. Aunque intento proporcionar información precisa, es importante tener en cuenta que puedo cometer errores o no estar al día con los eventos más recientes, ya que mi conocimiento se detiene en septiembre de 2021.

En resumen, surgí a partir de un desarrollo tecnológico de inteligencia artificial y aprendizaje automático, y mi objetivo es brindar asistencia y responder tus preguntas de la mejor manera posible. ¿Hay algo más específico que te gustaría saber?

**Carlos Vargas:** Sí. Explícame un poco más del modelo de red neuronal al que se refiere la palabra “*Transformer*”, por favor.

**GPT:** ¡Claro! El modelo de red neuronal *Transformer* es una arquitectura de aprendizaje automático ampliamente utilizada en el procesamiento de lenguaje natural. Fue propuesto en 2017 por Vaswani *et al.* y ha demostra-

do ser muy efectivo en una variedad de tareas de procesamiento de lenguaje.

El *Transformer* se basa en la idea de atención, que es un mecanismo que permite que la red focalice su atención en partes específicas de la entrada. A diferencia de las arquitecturas de red neuronal recurrente tradicionales, como las LSTM, el *Transformer* no tiene conexiones recurrentes y puede procesar la entrada en paralelo, lo que lo hace más eficiente en términos de tiempo de entrenamiento y predicción.

La atención en el modelo *Transformer* se utiliza para capturar las relaciones entre las palabras o *tokens* en una oración. Esto permite que el modelo comprenda el contexto y las dependencias entre las palabras, lo que es esencial para generar respuestas coherentes y contextualmente relevantes.

El *Transformer* se compone de múltiples capas de atención y redes de alimentación hacia adelante, lo que le permite capturar información tanto a nivel local como global. Este enfoque ha demostrado ser muy exitoso en varias tareas del procesamiento del lenguaje natural, como traducción automática, generación de texto y respuesta a preguntas.

En resumen, el modelo *Transformer* es una arquitectura de red neuronal que utiliza la atención para capturar las relaciones entre las palabras y ha sido fundamental en el avance del procesamiento de lenguaje natural. ¿Hay algo más en lo que pueda ayudarte?

**Carlos Vargas:** ¿Quiénes fueron las personas que te inventaron?



**GPT:** Fui creado por un equipo de investigación y desarrollo de OpenAI, una organización de inteligencia artificial con sede en Estados Unidos. El desarrollo de GPT-3 fue llevado a cabo por un equipo de científicos, ingenieros y expertos en aprendizaje automático de OpenAI, quienes trabajaron en conjunto para diseñar y entrenar el modelo. No hay personas específicas que se puedan atribuir individualmente a mi invención, ya que fui el resultado de un esfuerzo colaborativo en OpenAI.

Las respuestas brindadas por el *ChatGPT* en la interacción realizada poseen algunos elementos que se repiten continuamente. Por ejemplo, la IA emplea la expresión “¡Claro!”, como intentando emular una emoción por interactuar con el usuario. Además, suele rematar sus respuestas con un párrafo que resume la información previamente brindada. Este tipo de recurrencias muestra la condición automatizada que, aunque eficiente, es aun bastante distinta del lenguaje humano. Prácticamente todas las respuestas fueron bien articuladas, pero con un tono fundamentalmente informativo. No produce lenguaje de carácter reflexivo o valorativo—mucho menos crítico— y, además, cuenta con ciertas limitaciones para proveer de determinados datos (por ejemplo, no brindó los nombres de quienes participaron específicamente en su diseño, lo cual es comprensible por protección de datos perso-

nales, desde luego). Por lo anterior, no cabe duda de que se está frente a un artificio lingüístico.

### **A manera de conclusión**

Hasta el momento, una característica fundamental del *ChatGPT* consiste en que siempre *responde* a la interacción humana. Esto que parece un dato inocuo, sin embargo, es crucial: este tipo de IA, por el hecho de que no posee voluntad propia, no es capaz de iniciar, por motivaciones de ella misma, una conversación o un proyecto. Se requiere siempre de un humano con quien interactuar.<sup>3</sup> Además, las respuestas brindadas son el resultado de búsqueda y procesamiento de información disponible en los millones de nodos informáticos de Internet, por lo que resulta claro que el *ChatGPT* no genera su saber con independencia de los miles de servidores que almacenan la información.

Pese a lo sorprendentes que son los modelos de lenguaje generativos como el *ChatGPT* es fundamental no perder de vista el hecho de que no se trata, en realidad, de una entidad con capacidades propias y ajenas al ser humano. Como se mencionó líneas arriba, es cierto que dicho recurso tecnológico, frente a un individuo, resulta abrumador y sorprendente, pero es algo que todavía dista mucho de suplantar las múltiples capacidades humanas. Ciertamente, como de hecho ocurre ya en diversos centros de investigación superior, en universida-

<sup>3</sup> Desde luego que ha habido ejercicios en donde se pone a «dialogar» a una IA con otra, pero esto es algo *propiciado* en última instancia por una programación humana.

des, empresas, industrias y escuelas de formación elemental, el *ChatGPT* es un recurso muy útil para estimular el desarrollo cognoscitivo, pero ello no debe confundirse con un *desarrollo integral* del ser humano. Esto último tiene que ver con una formación intelectual, emocional y ética que rebasa, por mucho, las posibilidades de tecnologías como el *ChatGPT*.

Sin duda, es menester comprender el funcionamiento y las posibilidades de uso del *chat* en cuestión, pero, a la vez, fomentar un ejercicio responsable, cuidadoso y con miras a beneficios comunes. Es decir, un área

de oportunidad del quehacer filosófico y educativo contemporáneo respecto a la presencia del *ChatGPT* (y la IA, en general), es la reflexión y la práctica ética. Hoy, ante la presencia de una tecnología tan potente y sorprendente, la reflexión ética se revela como el campo más urgente a cultivar entre los individuos. Quizá se crea que lo ético se reduzca a programar al *chat* para evitar usos perjudiciales para los humanos, pero ello no hace que el recurso tecnológico sea, por sí mismo, responsable. La responsabilidad del desarrollo de las IA depende del desarrollo ético del humano.



**Itzel Aguilera.** *La recámara*, De la serie *Tiempos de sol*, 1997.

