

El cerebro detrás del algoritmo

Revista Riberas

Por Diego A. Evin* | Ilustración: Staff Revista RIBERAS**

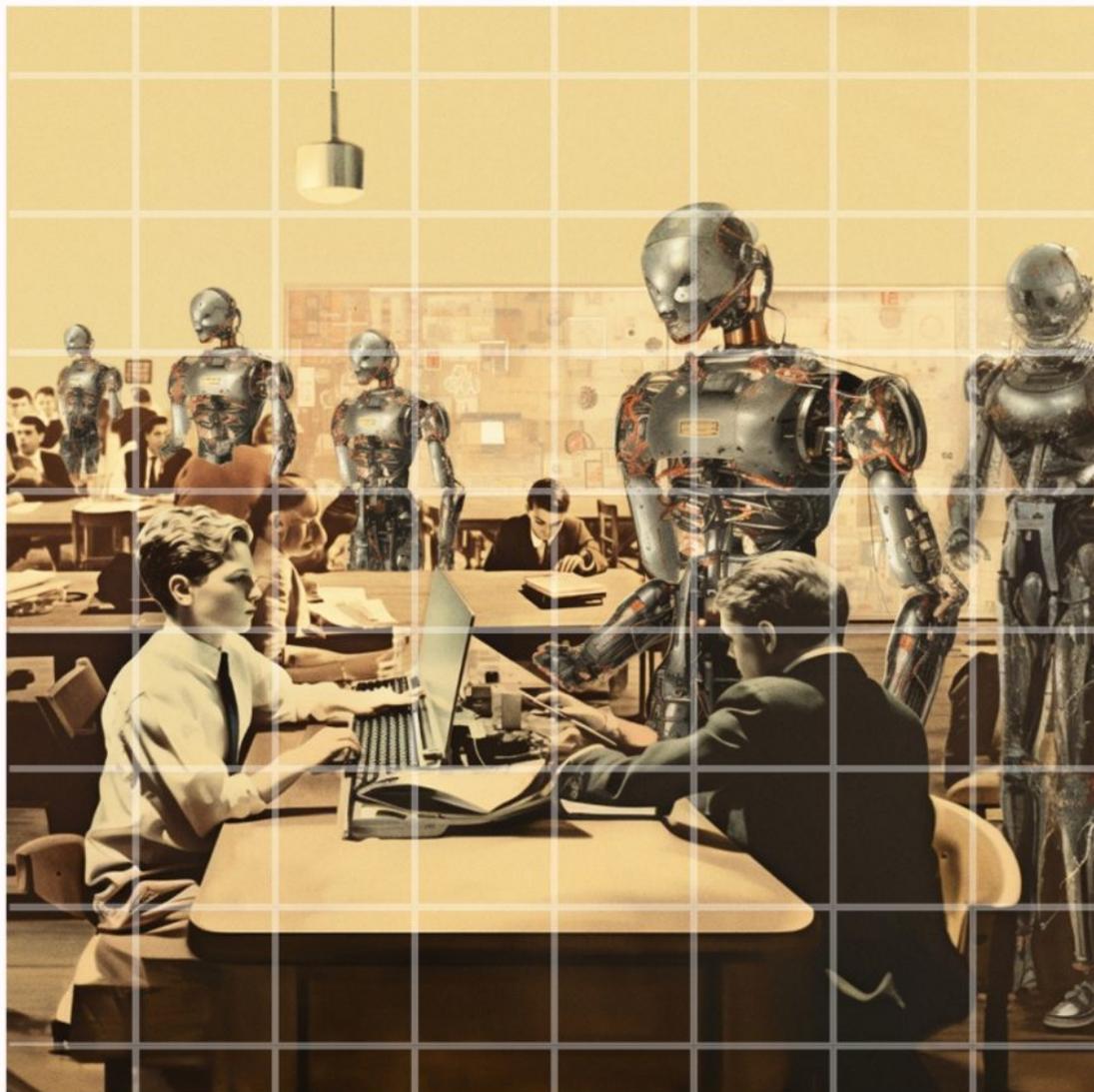
ChatGPT es un modelo de lenguaje basado en Inteligencia Artificial. Desde que llegó y se globalizó su circulación social, entró a los ámbitos educativos y abrió preguntas como cuál es el uso que le dan los estudiantes, qué resistencias provoca, y por qué estos nuevos contextos educativos abren desafíos frente al proceso enseñanza-aprendizaje y construcción de conocimiento.

La irrupción de ChatGPT trajo al ámbito educativo a un nuevo actor, tan popular como controvertido. Entre su adopción natural y como una herramienta más por parte de los estudiantes, y las diferentes reacciones de los docentes, este cambio invita conocer de qué se trata esa tecnología y pensar su rol e implicancias en el ámbito educativo.

Los avances recientes en Inteligencia Artificial nos sorprendieron incluso a quienes trabajamos en el área. No tanto por la novedad técnica que suponen, sino por la velocidad y extensión con la que se popularizaron y adoptaron. Para la comunidad docente, significó además, la irrupción de un actor nuevo en el escenario de la educación.

Una de las primeras experiencias que impactaron al interactuar con *ChatGPT* fue la de hacerle las preguntas de un examen y corregirlas con el mismo criterio que utilizaba en nuestra asignatura. Un estudiante con esas respuestas, sin dudas lo aprobaría. Por el tipo de uso del lenguaje, no podría asegurar que no lo haya escrito una persona. Además, su funcionamiento mostraba cierta variabilidad. Tras probarlo repetidamente con las mismas preguntas, los resultados eran ligeramente diferentes, pero igual de certeros. Tiempo después, fue noticia y parte de la campaña publicitaria de *ChatGPT*, que sus resultados habían superado varios exámenes en facultades de Derecho y Negocios [1].

El impacto del rendimiento de su funcionamiento está no sólo en la precisión de las respuestas, sobre todo de aquellas teóricas, sino en la elaboración lingüística que es la que precisamente, nos resulta razonablemente natural. Asimismo, es un tipo de respuesta adaptable ya que si pedimos al sistema que lo redacte como si fuera un niño de 10 años, eso se verá reflejado en el resultado. Por lo tanto, una primera reacción es la de pensar en cómo impactarían estas dimensiones en la forma de evaluar el proceso de aprendizaje.



De la era Wikipedia a la de ChatGPT

Junto con la aparición de Wikipedia, en sus orígenes, llegaron los temores sobre el modelo de enciclopedia libre y sus alcances, cómo la utilizarían los usuarios, cuáles y cómo controlar normas de fiabilidad, veracidad, actualización, entre otras. Cuando se socializa y comienza a circular *ChatGPT* se la pensó como una alternativa a Wikipedia que le agregaría excelentes capacidades de síntesis. Y en cuanto a buscadores tradicionales como Google, se la pensó potencialmente en relación a la posibilidad de hacer consultas en forma de diálogo, sin modificar la formulación lingüística que usamos habitualmente entre humanos para obtener respuestas mucho más precisas, donde su aplicabilidad potencial es más profunda.

Analizando la adopción y apropiación de esta tecnología en varios estudiantes y becarios, encontramos que lo usaban como la figura de un tutor personal al que le pueden hacer las preguntas que muchas veces en el aula prefieren reservarse. Es decir, funciona como una herramienta amplia en varios aspectos: en el caso de estudiantes de TICS acompaña el inicio de programación en un nuevo lenguaje o permite buscar soluciones a errores en el código,

también se utiliza como fuente de consulta para explorar un tema nuevo, en el proceso de lluvia de ideas, o para sugerir la redacción o traducción de un texto. Muchas personas, incluso, generaron con ella una dependencia cotidiana y hasta natural, que obviamente tiene implicancias en el proceso educativo. Todo este panorama, nos demanda a la comunidad docente un conocimiento mínimo acerca de qué es esta tecnología para que experimentemos con ella, discutamos su valor y pensemos su rol en el proceso educativo.

Sin embargo, no todo lo que brilla es admirable. Como habrán experimentado las lectoras/es que interactuaron con esta herramienta, sabrán que los resultados no son perfectos. Muestran muchas falencias como la de “alucinar” respuestas, e incluso algunas de ellas pueden ser asociadas a esos momentos tan conocidos como el de cuando los estudiantes “guitarrear”.

Existen también limitaciones en las argumentaciones tras esperar una elaboración más profunda sobre un tema específico, sesgos notorios en las respuestas, y otras restricciones que seguramente se irán haciendo más patentes al extender su uso. Curiosamente, los resultados reportados son mejores en áreas como Economía que en Derecho, Medicina y Matemáticas [2]. Y muchos de esos comportamientos y limitaciones se entienden mejor si se conoce cómo están contruidos estos modelos.

Queremos conversar sobre cualquier tema: ¿Qué hay detrás de *ChatGPT*?

En primer lugar, la tecnología que está por detrás de *ChatGPT* se conoce como Grandes Modelos de Lenguaje (o LLM por su acrónimo en inglés). Podemos pensar a un modelo de lenguaje como el cerebro detrás del algoritmo que predice y sugiere cuál será la próxima palabra que vamos a escribir al redactar un mensaje de texto. Predice la siguiente palabra en una secuencia, considerando un contexto, una historia de palabras previas. En los modelos de lenguaje desarrollados hace algunos años, que usaban un tipeo predictivo, reconocimiento óptico de caracteres o el reconocimiento automático del habla, las probabilidades de las palabras siguientes se aprendían simplemente observando y contando sus co-ocurrencias en grandes colecciones o corpus de texto. Hoy se lo hace usando unas redes neuronales especiales llamadas *Transformers* que, en vez de recurrir al conteo, emplean un mecanismo denominado *Atención*[3], que establece vínculos contextuales y de afinidad entre secuencias de palabras, que también se aprenden en un proceso de entrenamiento sobre corpus de texto.

A ese modelo de lenguaje, capaz de predecir la palabra siguiente, se lo puede usar de modo inductivo para generar secuencias de palabras: a partir de una secuencia de palabras previas se predice la siguiente, se la incorpora como la más reciente a la lista de palabras previas y se repite el procedimiento. Es justamente así como funcionan los modelos como *ChatGPT* y la razón por las que se los denominan generativos. De hecho, GPT es un acrónimo en inglés que significa *Transformers Preentrenados Generativos*. Se puede pensar la parte de la pregunta

que le hacemos como la secuencia de palabras previas, y la respuesta esperada como la parte que se completa con ese mecanismo generativo. Es decir, el primer punto que debería comprenderse es que esa caja negra que responde nuestras inquietudes no tiene un mecanismo de razonamiento lógico de fondo, sino puramente estadístico. Esto hace que muchos discutan el concepto de cuán “inteligentes” son realmente estos sistemas que han llegado a llamar “loros estocásticos” [4].

¿Y cómo hacen estos sistemas para elaborar una respuesta cuando hay datos contradictorios? Así como en el ámbito jurídico cuando hay un tema controvertido se dice “la mitad de la biblioteca opina de una forma y la mitad de otra”, y la decisión depende de cuál se elija, en los LLMs las respuestas dependen principalmente de los datos con los que se haya alimentado al modelo. Retomando la frase previa: si se le enseña con una mitad de la biblioteca, la biblioteca completa o muchas bibliotecas. De esa manera, los sistemas como *ChatGPT* manifiestan un sesgo de visiones, opiniones y evidencias. Si bien no hay información muy detallada al respecto, *ChatGPT* está entrenado con enormes cantidades de datos libres como la Wikipedia, base de datos de trabajos científicos, compilaciones de libros, compilaciones de páginas web en general, así como datos propios y privados. En su caso se buscó crear un sistema generalista, capaz de conversar sobre cualquier tema, que pueda responder cualquier pregunta. Seleccionando otros conjuntos de texto de entrenamiento, la misma tecnología se puede usar para desarrollar desde cero o para adaptar los modelos generales y producir un interlocutor especialista, por ejemplo, del área médica o legal y cuya precisión en esos campos sea mucho mayor a las de un sistema generalista.

Ese proceso de aprendizaje o preentrenamiento es la parte más importante del sistema y requiere miles de placas de procesamiento gráfico para cálculos, meses de entrenamiento y tiene un costo de entre 5 y 10 millones de dólares. A esa fase *ChatGPT* agrega otras, diseñadas para mejorar sus salidas. Entre ellas una de ajuste fino que reentrena los modelos preentrenados usando otro subconjunto de datos. En este caso conformado por ejemplos similares a las entradas y salidas deseadas de una interacción esperada para el sistema. La elaboración de ese conjunto de datos (denominados consultas o prompts y sus respuestas), es generalmente manual y muy laboriosa. La fase siguiente se denomina de aprendizaje por refuerzo, a partir de la retroalimentación humana y se la puede ver como otro paso de ajuste fino en el que se adapta el modelo usando un criterio de preferencia, obtenido a partir de usuarios que generan un tercer conjunto de datos de entrenamiento, en este caso comparando y ordenando salidas alternativas del sistema ante una misma entrada. La aplicación de esa secuencia de pasos termina determinando la personalidad que podemos atribuirle al sistema. De esta forma, además de buscar precisión y utilidad en las respuestas, en *ChatGPT* se busca que el mensaje generado sea políticamente correcto, que evite entrar en conflictos o tocar temas que pueden resultar sensibles por cuestiones legales. Sin embargo, en las cantidades

masivas de datos con las que se entrena existen múltiples sesgos que es muy difícil filtrar y muchas veces se trasuntan en el texto generado. Ejemplos como esos se dieron en varios casos de textos racistas, sexistas u ofensivos que se hicieron públicos principalmente en las primeras etapas en que el sistema se hizo público.

La forma descrita de funcionamiento de estos sistemas hace que las salidas sean dependientes de la forma que tengan las entradas o consultas. Eso ha dado lugar a la necesidad de aprender un poco los secretos a la hora de saber cómo preguntar, y a un área de estudio denominada ingeniería de prompts.



Es como hablar con un docente pero sin la presión

La gran cantidad de recursos tanto de datos como de capacidad de cómputo y el costo que supone, hace que por ahora generar modelos tan grandes como los de *ChatGPT* estén vedados a las grandes compañías, y por ende, que los sesgos de los sistemas que usemos estén determinados por ellas. Esto dio lugar a debates interesantes respecto al rol de estos modelos interviniendo en los procesos de creación de significado en las sociedades, así como la delegación de habilidades cognitivas y sus implicancias. Todas esas controversias abren interesantes oportunidades para la educación. Si bien muchos de los usos de estas tecnologías están asociados principalmente a una de las partes del proceso educativo como lo es la adquisición de conocimientos, el aprendizaje de conceptos, de hechos, teorías y habilidades específicas, ése no es el fin último de educar. Como ilustra la popular frase atribuida a Plutarco “La educación no es llenar un balde, sino encender un fuego”. La educación además supone otros objetivos como la formación de ciudadanía responsable y comprometida, que

fomente valores como respeto, tolerancia, igualdad y justicia; el desarrollo personal y la mejora de oportunidades, el desarrollo del pensamiento crítico y creativo, y la promoción del aprendizaje continuo. En cada una de estas dimensiones es posible aprovechar las ventajas, y trabajar en las amenazas y desafíos de esta tecnología para mejorar la educación con una visión integral.

Desde que se hizo pública esta herramienta resonaron en el ámbito educativo diversas reacciones. Se pueden encontrar varios estudios que toman las implicancias de esta tecnología [2, 5-7].

Entre las amenazas y riesgos se plantean:

- Plagio y posibilidad de aprobar sin aprender
- Sobreestimación y dependencia excesiva, que limite el pensamiento crítico y la independencia para resolver problemas, o soslaye otras estrategias de aprendizaje
- Pérdida de creatividad y originalidad, en caso que comiencen a tomar al sistema como la principal fuente de ideas o inspiración
- Desinformación a partir de los fenómenos de alucinación y falta de revisión de sus resultados
- Amplificación de los sesgos o prejuicios de estos sistemas
- Vulneración de la privacidad y uso sin consentimiento de datos, así como la propiedad y control de los contenidos generados
- Aumento de las desigualdades entre estudiantes que tienen los recursos materiales para hacer uso de estas herramientas para el aprendizaje y los que no lo tengan
- Sustitución o degradación del rol de educadores humanos
- Reducción de la interacción humana

Con respecto a las ventajas de la tecnología se destacan:

- Inmediatez al acceso de información
- Tutoría personalizada para estudiantes
- Ayuda para aprender idiomas y lenguajes de programación
- Ayuda a redactar ensayos, resumir y hace preguntas sobre trabajos científicos
- Auxilio a los docentes para estructurar la currícula, generar material para sus cursos, sugerencias para formular preguntas en exámenes

Finalmente, sobre uno de los aspectos que más preocupación despierta, la posibilidad de plagio, se hacen algunas recomendaciones a la comunidad

docente:

- Incorporar imágenes en las preguntas de los exámenes que hacen más difícil generar respuestas correctas a ChatGPT.
- Hacer preguntas enfocadas más en el análisis, aplicación y creación (niveles más altos en la taxonomía de Bloom) que recordar respuestas específicas
- Hace preguntas que involucren experiencias de lo que se hizo o discusión en clase, información que los sistemas no cuentan para generar respuestas
- En el caso de monografías, revisar que existan y sean pertinentes las referencias bibliográficas
- Establecer guías anti-plagio y educar a los estudiantes sobre la integridad académica.

Como comunidad docente y educativa tenemos por delante el desafío de explorar esas oportunidades, aprender y experimentarlas con este tipo de herramientas que ya están y existen entre nosotros. Es necesario en este sentido, re debatir el rol docente y cuáles son los otros espacios de aprendizaje que abren estas tecnologías sobre los usos que habitualmente hacen de ellas los estudiantes. De esa forma, podremos enriquecer la experiencia del aprendizaje, construyendo y compartiendo una visión no solo utilitarista sino formando un espíritu crítico entre estudiantes y docentes.

*Diego A. Evin es Bioingeniero, por la Universidad Nacional de Entre Ríos, Doctor en Ciencias de la Computación de la UBA. Realizó su formación de posgrado en el Laboratorio de Investigaciones Sensoriales (INIGEM CONICET-UBA). Es investigador asistente de CONICET en CINTRA CONICET-UTN FRC, y profesor adjunto de la materia Inteligencia Artificial en la Facultad de Ingeniería de la Universidad Nacional de Entre Ríos.

**Las ilustraciones de portada y de interiores fueron creadas con apps de IA.

Referencias

- [1] ChatGPT Passes Exams from Law and Business Schools. Accedido online: <https://www.reuters.com/legal/transactional/chatgpt-passes-law-school-exams-despite-mediocre-performance-2023-01-25/> (accedido el 27 de mayo de 2023).
- [2] Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410.
- [3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

[4] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).

[5] Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 1-15.

[6] Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN 4337484.

[7] Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*.

Tags: [ChatGPT](#)[Docencia](#)[Educación Superior](#)[TIC's](#)[Universidad y tecnología](#)[usos sociales](#)