

Asomándonos a la ventana contextual de la Inteligencia Artificial: decálogo de ayuda para la identificación del uso de ChatGPT en textos académicos



Por Gonzalo Jover*, David Carabantes* y José L. González Geraldo**

* Universidad Complutense de Madrid

** Universidad de Castilla La Mancha

Palabras clave: ChatGPT, OpenAI, Inteligencia Artificial

Publicado el 2 de junio de 2023

Que la Inteligencia Artificial (IA) ha contaminado cada rincón de nuestra sociedad a un ritmo tan espeluznante como atractivo y peligroso es más que evidente a estas alturas. Independientemente de su mayor o menor adaptación o rechazo, los que nos encargamos de aportar granitos de arena en educación superior no solo podemos, sino que debemos, pronunciarnos. Después de todo, coincidiendo con Lennox, no tenemos por qué ser físicos nucleares o climatólogos para poder discutir el impacto de la energía nuclear o del cambio climático (2020, p. 17).

Sin embargo, este razonamiento no nos proporciona una patente de corso bajo la que actuar y dictaminar a diestro y siniestro, sin más. Sin importar el campo al que nos dediquemos, el respeto que debemos hacia la ciencia nos fuerza a indagar sobre las bases que sostienen los actuales modelos de lenguaje generativos y, en particular, cuáles son sus potencialidades y sus limitaciones. Como heraldo catalizador de toda esta debacle o dicha, nos centraremos en el ya archiconocido ChatGPT.

Nuestra intención aquí no es la de explicar cómo funciona, algo que ni los propios desarrolladores de OpenAI pueden -ni quieren- concretar de manera absolutamente precisa, de ahí que sea identificado como un modelo de tipo "caja negra" (Zhai, 2023), pero sí querríamos atajar y compartir uno de los aspectos que, al menos de momento, sin duda condiciona las posibilidades que este rey del pastiche nos ofrece. Nos referimos al límite de lo que se conoce como "ventana contextual", un concepto clave a la hora de entender por qué ChatGPT puede resultar por momentos tan genial como patético (Marcus, 2022).

Importe o no el tamaño de los datos en los que se basó el aprendizaje no supervisado de GPT, un aspecto que tras el anuncio de PaLM 2 parece ser cada vez menos relevante (Merino, 2023), y dejando de lado que su adaptación supervisada a *chatbot* es precisamente eso, una concreción conversacional propia de una inteligencia estrecha (ANI), partiremos de una premisa básica: sí, el saber ocupa lugar. Además, este espacio está habitado por lo que se conoce como *tokens*, palabras o partes de palabras que, simplificando en exceso, sirven de enlace entre nuestro idioma y el de los actuales ordenadores, entre nuestras ideas y sus ceros y unos. Un espacio que, por supuesto, está acotado.

Así, el número de tokens totales que pueden manejar los distintos modelos que pugnan por llevarse el gato al agua (ChatGPT, Bard y Bing, principalmente) es uno de los principales umbrales -y problemas- al que hemos de prestar atención como académicos. Como ejemplo, encontramos los 4.096 tokens en los que se basa ChatGPT 3.5 o los 8.192 de ChatGPT 4. Este límite marca, para entendernos, el rango de contexto al que puede prestar atención en cada consulta (*prompt*) cada modelo. Dicho de otra forma, mientras que ChatGPT 3.5 empieza a "olvidar" información cuando nos acercamos a las 2.500 palabras de conversación, ChatGPT 4 aguanta nuestra barrila con algo más de paciencia, desorientándose conforme alcanzamos las 6.000 palabras.



Fuente de la imagen: [Pixabay](https://pixabay.com/)

Si reflexionamos sobre este aspecto, pronto comprenderemos por qué los textos elaborados a través de ChatGPT, más pronto que tarde, tienen serias dificultades para mantener la coherencia, desarrollar discursos bien hilados o, simplemente, recordar los conceptos que ya mencionaron. Desde la academia, esta suerte de “memoria pez” (Stern, 2023) se presenta ante nosotros como un talón de Aquiles que hemos de aprovechar desde el punto de vista pedagógico, pues nos proporciona una serie de claras pistas que nos ayudarán a identificar, simplemente leyendo cualquier trabajo académico que caiga en nuestras manos y sin acudir a otros programas que chequean la probabilidad del uso del algoritmo, si estamos ante un caso de “plagio de alta tecnología” (Chomsky, Roberts y Watumull, 2023). Por ello, por supuesto contando con la ayuda del propio protagonista de todas estas líneas, hemos recopilado el siguiente “Decálogo de ayuda para la identificación del uso de ChatGPT en textos académicos”:

- (1) *Repetición de ideas*: ChatGPT, debido a su limitada ventana contextual, puede repetir conceptos o ideas que ya se han presentado antes en el texto. Si notas que un trabajo tiene una redundancia significativa, podría ser una señal de que se ha utilizado ChatGPT.
- (2) *Inconsistencia a lo largo del texto*: ChatGPT puede perder la coherencia a medida que genera textos más largos. Si un trabajo parece comenzar de manera precisa, pero se vuelve progresivamente menos coherente o relevante, esto puede ser una señal de uso de ChatGPT.
- (3) *Falta de referencias cruzadas*: ChatGPT tiene dificultades para recordar y referirse a elementos del texto que están fuera de su ventana contextual. Si un trabajo carece de referencias cruzadas adecuadas o no logra mantener un hilo argumental sólido a lo largo del texto, puede ser una señal de uso de ChatGPT.
- (4) *Errores de contexto*: ChatGPT puede producir errores contextuales, como referencias incorrectas o inconsistentes, debido a su incapacidad para mantener un contexto a largo plazo.
- (5) *Uso excesivo de jerga*: ChatGPT puede sobreutilizar la jerga técnica o académica para parecer más autoritativo. Si un trabajo está lleno de jerga pero carece de una argumentación sólida, puede ser un indicio de que se ha utilizado ChatGPT.
- (6) *Falta de profundidad crítica*: Aunque ChatGPT puede generar contenido basado en patrones en los datos de entrenamiento, carece de la capacidad para la reflexión crítica y la profundidad de análisis que caracteriza al pensamiento humano.
- (7) *Respuestas generales o vagas a preguntas específicas*: ChatGPT puede proporcionar respuestas que, aunque gramaticalmente correctas, no abordan de manera precisa y específica la pregunta o el tema en cuestión.
- (8) *Inconsistencias en el estilo de escritura*: Un cambio abrupto en el estilo de escritura puede ser una señal de que se ha utilizado ChatGPT. Por ejemplo, si partes del texto parecen escritas por diferentes autores, esto podría indicar que se han utilizado varias ejecuciones de ChatGPT.

(9) *Referencias bibliográficas irrelevantes o inapropiadas*: ChatGPT puede generar citas que parecen relevantes, pero que al examinarlas no tienen relación directa con el argumento presentado. A ello se une otro de sus grandes problemas, las alucinaciones que derivan en referencias inventadas.

(10) *Conclusiones débiles o no relacionadas*: Debido a la limitada ventana contextual, ChatGPT puede tener dificultades para generar conclusiones coherentes y sólidas que se vinculen adecuadamente a los argumentos presentados en el texto.

Es evidente que todo lo dicho no solo puede y debe aplicarse a los trabajos estudiantiles sino también, por supuesto, a los que nosotros mismos elaboramos; el uso de la IA en actividades académicas está siendo abordado desde todas las partes del tablero y sería pueril pensar que su abuso queda lejos de nuestras puertas. El potencial y el peligro que estos modelos ofrecen como asistentes a tareas docentes y de investigación es tan grande como el que, por ejemplo, puede entrecerarse al relacionarlo con las tareas de gestión de nuestras revistas. En este sentido, esperamos poder compartir en breve los resultados de una investigación en la que comprobamos hasta dónde se puede confiar en estos modelos de lenguaje como posibles revisores de revistas especializadas en el campo educativo.

Finalmente, los más duchos en la materia habrán notado la omisión del modelo GPT-4 32k, que dispone de un motor gemelo del 8k capaz de mantener en su ventana contextual hasta 32.000 tokens, unas 25.000 palabras. Un gran avance del que seguramente pronto podremos disfrutar como usuarios (de momento solamente accesible con invitación y a través del uso API), pero que se queda prácticamente en mantillas si lo comparamos con el reciente anuncio de otro modelo patrocinado por Anthropic: Claude (<https://www.anthropic.com/product>), capaz de prestar atención, de una sola vez, nada más y nada menos que a 100.000 tokens, unas 75.000 palabras aproximadamente.

No obstante, como ya han comprobado los que tienen acceso a estos avanzados modelos, quizá no todo sea únicamente cuestión de ventana contextual, sino de otras cuestiones que, de nuevo de momento, nos permiten seguir preocupantemente ilusionados por lo que hay de “inteligente” sobre el sustrato de lo artificial. Sea como fuere, es indudable que en la universidad estamos ante un cambio que empequeñece aquellas promesas -quizá también falsas- que desde Bolonia nos prometieron a comienzos de siglo. Hagamos lo posible por estar a la altura, hagamos lo imposible por seguir presumiendo de humanidad.

Referencias:

Chomsky, N., Roberts, I. y Watumull, J. (2023). The False Promise of ChatGPT. *New York Times*. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

Lennox, J. C. (2020). *2084. Artificial Intelligence and the future of humanity*. Zondervan.

Marcus, G. (2022). How come GPT can seem so brilliant one minute and so breathtakingly dumb the next? <https://garymarcus.substack.com/p/how-come-gpt-can-seem-so-brilliant>

Merino, M. (2023). No estamos aquí para masturbarnos por el número de parámetros. El CEO de OpenAI quita importancia a lo que más se alababa de GPT-4. <https://www.msn.com/es-es/noticias/tecnologia/no-estamos-aqui%C3%AD-para-masturbarnos-por-el-n%C3%BAmero-de-par%C3%A1metros-el-ceo-de-openai-quita-importancia-a-lo-que-m%C3%A1s-se-alababa-de-gpt-4/ar-AA1b0tem>

Stern, J. (2023). *GPT-4 Has the Memory of a Gold fish*. <https://www.theatlantic.com/technology/archive/2023/03/gpt-4-has-memory-context-window/673426/>

Zhai, X. (2023). ChatGPT for Next Generation Science Learning. *Crossroads*, 29(3), 42-46. <https://doi.org/10.1145/3589649>

Cómo citar esta entrada:

Jover, Gonzalo; Carabantes, David y González Geraldo, José L. (2023). Asomándonos a la ventana contextual de la Inteligencia Artificial: decálogo de ayuda para la identificación del uso de ChatGPT en textos académicos. *Aula Magna 2.0* [Blog]. <https://cuedespyd.hypotheses.org/13299>