

Breve introducción a ChatGPT para Dejusticia¹

¿Qué es exactamente *ChatGPT*? Es una herramienta de inteligencia artificial creada por [OpenAI](#) capaz de generar respuestas con base a preguntas u órdenes dadas por los usuarios. Este “chat” utiliza un tipo de Inteligencia artificial llamado “IA generativa” que crea contenido a solicitud del usuario². Su uso ha generado debates de todo tipo. Este corto texto busca abordar tres cuestiones relacionadas con el funcionamiento, de dónde provienen las respuestas generadas por este tipo de herramienta y cómo genera las mismas. Luego pasa a analizar el grado de confianza que podemos tener en las respuestas dadas por ChatGPT.

Cómo funciona

Una forma simple para entender cómo funciona esta tecnología es pensar que realiza dos procesos (casi) simultáneos. El primer proceso es el de búsqueda de información relacionada con los términos que introduce el usuario. El segundo proceso es el de generación de contenido que responda a la pregunta del usuario basada en esa búsqueda. Desde esta óptica, podríamos decir que estos sistemas generan respuestas luego de “buscar” en una base de datos. Como la respuesta se produce por asociación de palabras, esta puede provenir directamente de lo que encuentre en su base de datos o ser una construcción autónoma que parezca proveniente de los datos con los que fue entrenado el sistema, pero en realidad sea completamente nueva.

De dónde proviene la información de ChatGPT

En términos sencillos, el “conocimiento” de ChatGPT proviene de información pre-programada y de los datos con los que se entrenó el sistema. ¿Qué información conforma la base de datos? En resumen, casi todo lo que había en internet hasta septiembre de 2021. Esto incluye textos académicos, libros, todas las entradas de Wikipedia, información disponible en las redes sociales, blogs, etc.³. Como dijo Lucía Camacho en un [programa](#) organizado por el medio Razón Pública: “la base de datos de ChatGPT es un pantallazo de internet a septiembre de 2021”.

Generación de respuestas

ChatGPT no es un motor de búsqueda como Google o Bing. Su principal función es generar una conversación que asemeje el lenguaje humano. Para nutrirla, acude a la base de datos con las que fue entrenado el sistema y produce texto predictivo basándose en patrones estadísticos del lenguaje. El sistema ha sido entrenado con millares de textos en distintos idiomas y es capaz

¹ Documento elaborado por Daniel Ospina Celis y María Alejandra García. Difundido de manera interna en Dejusticia.

² Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations. *arXiv preprint arXiv:2301.04246*; pág. 15. <https://arxiv.org/abs/2301.04246>

³ Ibid, pág. 17.

de predecir las formas lingüísticas comunes. Así, cuando se le hacen preguntas, es capaz de interactuar y ofrecer respuestas acordes a lo que el usuario pidió⁴. Por eso, incluso si la pregunta es idéntica, las respuestas van a ser diferentes⁵.

Confianza en las respuestas dadas por ChatGPT

Confiar en que las respuestas dadas por ChatGPT son una fuente de información veraz sería incorrecto. La base de datos con la cual fue entrenado el sistema (internet) puede contener información falsa, errónea o sesgada⁶. Y como el sistema no indica cuáles son sus fuentes como lo haría un buscador estilo Google, podemos caer en error. La herramienta también es capaz de “inventar” información pues su función principal es mantener una conversación, no responder ajustándose a la verdad. [Este](#) hilo de Twitter muestra varios casos en donde el sistema inventó información en el contexto colombiano. A este tipo de errores se les llama “alucinaciones” y la empresa que creó ChatGPT recomienda no basarse en sus resultados para tomar decisiones importantes⁷.

Un ejemplo de una alucinación sucedió cuando un Juez en Cartagena utilizó ChatGPT para [motivar una tutela](#). Al preguntar sobre las sentencias relacionadas con el tema, la herramienta inventó varias e incluso explicó cuál era el contenido de estas sentencias ficticias. Los creadores de ChatGPT son claros al advertir que la herramienta puede producir errores respondiendo preguntas, por ejemplo al resolver operaciones matemáticas⁸.

Esto no quiere decir que este tipo de herramientas no sean sumamente útiles. Lo son, pero no por eso debemos confiar ciegamente en sus respuestas.

Recursos sugeridos

Preparamos una lista de recursos sugeridos (sobre todo videos o textos cortos) para aprender más sobre esta herramienta, sus potencialidades, sus dilemas y el impacto que podrá tener en nuestro trabajo y en los derechos humanos.

- **Sobre el funcionamiento de ChatGPT:**

⁴ OpenAI (2023). GPT-4 Technical Report. Pág 42. <https://cdn.openai.com/papers/gpt-4.pdf>.

⁵ Goldstein et al, pág 18.

⁶ Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). Understanding the capabilities, limitations, and societal impact of large language models. *arXiv preprint arXiv:2102.02503*.

⁷ OpenAI, págs 10-11. <https://cdn.openai.com/papers/gpt-4.pdf>

⁸ Von Hipel, P. (Enero 04 de 2023). ChatGPT Is Not Ready to Teach Geometry (Yet). Education Next. <https://www.educationnext.org/chatgpt-is-not-ready-to-teach-geometry-yet/#:~:text=It%20states%20that%20in%20a,of%20the%20other%20two%20sides>.

1. Acadaimy (04 de enero de 2023). ChatGPT Explained: Coding, Examples and How to Use. <https://www.youtube.com/watch?v=NH5h26cy4M>.
2. ByteByteGo (24 de abril de 2023). How ChatGPT Works Technically | ChatGPT Architecture. <https://youtu.be/bSvTVREwSNw>.
3. Arvin Ash (abril de 2023). So How Does ChatGPT really work? Behind the screen!. <https://youtu.be/WAiqNav2cRE>.
4. CNBC (Abril de 2023). Why OpenAI's ChatGPT it's such a big deal?. <https://youtu.be/pOmpqdlVCoo>.
5. Universidad de Chile (13 de abril de 2023). Podcast - ¿Qué es Chat GPT? Académicos explican el funcionamiento de esta Inteligencia Artificial. <https://youtu.be/rnCGru5eEUw>.

- **Sobre los Modelos de Lenguaje Natural:**

1. Simplilearn, 17 de marzo de 2021. Natural Language Processing In 5 Minutes | What Is NLP And How Does It Work? | Simplilearn. <https://www.youtube.com/watch?v=CMrHM8a3hqw>.
2. Ciencia del dato (Noviembre de 2022). El procesado de lenguaje natural: ¿nos entienden los ordenadores? <https://open.spotify.com/episode/4PLpswTG6BLjUkK0gE7seL?si=7fc8b395635b4712>.
3. Podcast. (25 de enero de 2023). 14. Fernando Balbachan - Modelos de lenguaje, ChatGPT, inteligencia artificial, lingüística. <https://open.spotify.com/episode/1fWApYwhQ4ZTBzZkMcqRwN?si=b87d7ec51fe64454>.
4. IBM. (2022). What is NLP (Natural Language Processing)?. <https://youtu.be/fLvJ8VdHLA0>.
5. Dot CV (2021). La Siguiete Gran Revolución: NLP (Procesamiento del Lenguaje Natural). <https://youtu.be/cTQiN9dewIg>.

- **Sobre las implicaciones a futuro de esta tecnología:**

1. ChatGPT: La carrera de la inteligencia artificial y la desigualdad. El hilo. Podcast. <https://open.spotify.com/episode/0KQjdoTxw0pJ5K2boRRzfN?si=200f8b14a3114d2b>.
 2. Human Rights Watch (web). Pandora's Box: Generative AI Companies, ChatGPT, and Human Rights. <https://www.hrw.org/news/2023/05/03/pandoras-box-generative-ai-companies-chatgpt-and-human-rights>.
 3. Paul Yan. (27 de abril de 2023). OpenAI Admits 80% Jobs Will Be Affected By AI | Research Paper Analysis. https://youtu.be/Jb2Z_uJbSiI. Pueden encontrar el paper en que se basó este vídeo aquí: <https://arxiv.org/pdf/2303.10130.pdf>.
 4. EO. (Marzo de 2023). Survival Strategies in the Era of AI Taught by Stanford | Stanford AIRE Director. <https://www.youtube.com/watch?v=ZA9K0JMrbWg>.
 5. ColdFusion. (Marzo de 2023). ChatGPT Has A Serious Problem. <https://www.youtube.com/watch?v=ce616lsed9Y>.
 6. Jyotirmoy Banerjee (2023). International Journal of Human Rights Law Review. ChatGPT and Human Rights: Navigating the Technological Frontier. <https://humanrightlawreview.in/chatgpt-and-human-rights-navigating-the-technological-frontier/>.
- **Sobre cómo utilizar esta herramienta para la investigación:**
1. Science Grad School Coach. (Febrero 2023). ChatGPT for Scientific Research: How to use AI as a Partner in Your Research. <https://youtu.be/KhRY-ktKB-4>.
 2. Profkip. (15 de diciembre de 2022). Using ChatGPT for writing in research. https://www.youtube.com/watch?v=tEdM9e_ycFU.
 3. Andy Stapleton. (Abril de 2023). EPIC ChatGPT Prompts for Research. <https://www.youtube.com/watch?v=w-GiUY-DcJY>.
 4. Scientific Writing with Karen L. McKee (Enero de 2023). Can ChatGPT Write Your Next Scientific Paper?. <https://www.youtube.com/watch?v=wnGpt030IG4>.
 5. Jorge Cruz. (Febrero de 2023). Usa ChatGPT para formular tu PREGUNTA de investigación y revisión de LITERATURA (lo último, sale mal). <https://www.youtube.com/watch?v=DvkqIGHUabk>.