

Un tsunami en expansión

Dos años de ChatGPT e IA Generativa



Con el apoyo de:

Dirección y coautoría:

Juan G. Corvalán y María Victoria Carro

Investigación y colaboración:

Marcos Adrogué, María Luciana Cardoso, Romina Décima Canovas, Gian Di Chiara D Agostino, Augusto Federico Córdoba, Abril Fernández Marquez, Melania Gadea, Margarita González Pereiro, Aylén Irrera Martino, Francisca Gauna Selasco, Agostina Jara Rey, Lola Ramos Pereyra, Felicitas Rodríguez Contreras, Priscila Silva, Michelle Truchot, Dolores Val Eyras y Ana Luna Villegas

Diseño gráfico

María Victoria Mafud

Not by AI. Hecho 100% por humanos

Índice

- I.** Introducción
Pantallazo de la revolución generativa.
- II.** El momento Big Bang
Las redes neuronales Transformers. Breve recorrido de los modelos GPTs.
- III.** La llegada de Copilot: Microsoft vs. Google
La batalla de los buscadores. Edge, Bing y Copilot.
La competencia por los modelos generativos.
- IV.** Lanzamiento de la API: escalando la integración de ChatGPT
Posibilidades de la API. Fine-tuning, transfer learning y RAG.
- V.** GPT-4: de la promesa open a un presente profit
ChatGPT Plus. GPT-4 y sus métricas de performance. GPT-4 Turbo. Mezcla de expertos.
- VI.** ¿La nueva apuesta de Pascal? Riesgos catastróficos de la IA
Carta abierta de expertos para pausar la IA. Riesgos catastróficos. Despido de Sam Altman.
- VII.** Opciones para personalizar y editar tu ChatGPT sin programar
Plugins. Custom instructions (instrucciones personalizadas).
GPTs
- VIII.** Un nuevo ecosistema para mejorar la productividad
ChatGPT Enterprise. ChatGPT Teams. Canvas.
- IX.** Multimodalidad y nuevas formas de interacción
Primeros avances multimodales. Sora. GPT-4o.
ChatGPT Search.
- X.** Agentes de IA y el camino hacia la AGI (inteligencia artificial general)
Demandas de Elon Musk a OpenAI. Niveles para alcanzar la superinteligencia de OpenAI. OpenAI o1-preview o “modelo frutilla”. Agentes de IA.
- XI.** Línea de tiempo

I

Introducción

00101



I. Introducción

El 30 de noviembre de 2022 se produjo uno de los hitos más importantes de la historia en el desarrollo de la IA: la empresa OpenAI liberó ChatGPT versión 3.5, un modelo de lenguaje multipropósito y multitarea para que cualquier persona con un usuario y una contraseña pudiera interactuar con él.

Desde su lanzamiento, ChatGPT rompió récords e inició el cambio más significativo en la industria tecnológica en este siglo. En dos meses, logró una hazaña sin precedentes: se convirtió en la plataforma digital con el crecimiento en cantidad de usuarios más rápido en toda la historia de Internet al alcanzar, en ese brevíssimo periodo, la asombrosa cifra de 100 millones de usuarios¹. También potenció la industria y estableció una forma de interactuar con nuestros dispositivos: el formato “genio de Aladino”, basado en preguntar y dar instrucciones a sistemas versátiles, capaces de realizar cualquier tarea del lenguaje natural. En dos años, podemos “hablar” con cualquier formato de información como pdfs y videos, crear clones digitales indistinguibles de humanos, generar podcasts a partir de documentos y manipular texto como si fuéramos Robert Downey Jr. con su Jarvis en Iron Man.

Este tsunami de innovación está cambiando el paradigma sobre el que se construyó la cuarta revolución industrial (las computadoras e internet). Por un lado, gran parte del software está siendo atravesado por los modelos de lenguaje, que se utilizan, incluso, como asistentes para generar el propio código informático. Por el otro, se vive una democratización inédita en el acceso a la inteligencia artificial, específicamente, a la inteligencia artificial generativa (GenAI). Sin necesidad de conocimientos de programación y de forma gratuita o a un costo muy accesible, los usuarios ahora pueden predecir y automatizar o semi-automatizar una cantidad creciente de tareas. Una realidad que, hace apenas unos años, habría parecido inimaginable debido a su complejidad y los elevados costos asociados.

En segundo plano, vivimos dos grandes disruptpciones que cambian la ecuación acerca de cómo construimos el conocimiento. Históricamente, entendemos los datos, obtenemos y producimos información para luego avanzar hacia conocimientos contextualizados que nos permitan realizar con éxitos tareas y tomar decisiones. La IA generativa, una vez puesta a disposición, licua todo el tiempo y espacio que se requería para avanzar en ese camino. Fragmentos de conocimiento humano sintetizado se ofrecen

¹Lo que a Instagram le llevó 30 meses y a TikTok 9, ChatGPT lo logró en tan solo 2 meses. Ver Garfinkle, Alexandra, *ChatGPT on track to surpass 100 million users faster than TikTok or Instagram: UBS, Yahoo! Finance*, disponible en <https://finance.yahoo.com/news/chatgpt-on-track-to-surpass-100-million-users-faster-than-tiktok-or-instagram-ubs-214423357.html> . Consultado el 27 de noviembre de 2024.

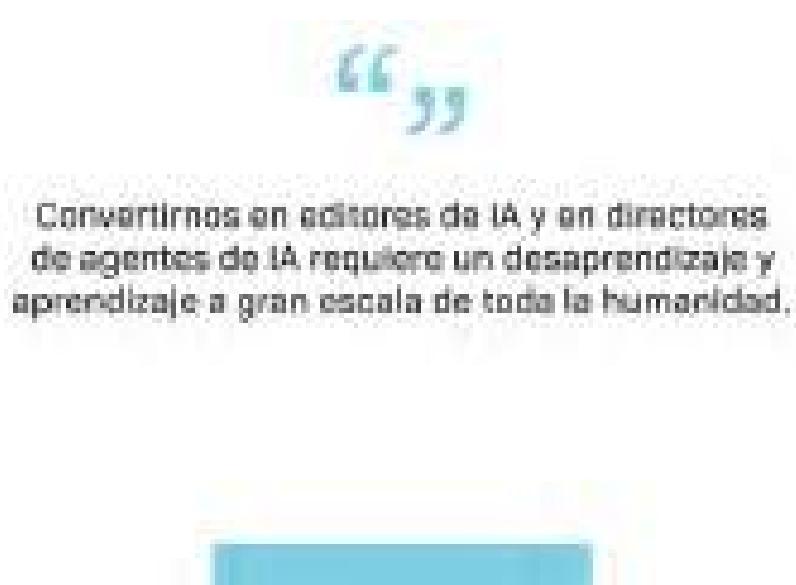
a las personas en un formato de “commodity” probabilístico, basado en predicciones masivas en tiempo real.

Por ejemplo, se pueden cargar 50 hojas de vida para que una IA como Notebook LM las convierta en un podcast. Pero no se trata simplemente de que dos personas nos cuenten en un relato de voces sintéticas la trayectoria profesional de cincuenta personas. La IA generativa predice y genera al final de la conversación del podcast, una sugerencia acerca de quiénes serían los tres mejores perfiles que deberíamos considerar. Todo eso en pocos minutos.

Por último, este cambio de paradigma nos pone en un escenario muy diferente a cómo se plantean los desafíos hasta hoy. Convertirnos en editores de IA y en directores de agentes de IA requiere un desaprendizaje y aprendizaje a gran escala de toda la humanidad. Nuestro cerebro no está preparado para hacer coworking masivo con sistemas probabilísticos, que reproducen y crean conocimiento sintético en función de la calidad de nuestras peticiones e instrucciones.

Si miramos en retrospectiva, 2 años hacia atrás, la versión de ChatGPT de noviembre de 2022, parece una tienda Blockbuster si lo comparamos con las funcionalidades y el ecosistema que rodea a los seis modelos que hoy se ofrecen en la interfaz de OpenIA. Ni que hablar de otras empresas y otros modelos de lenguaje como Claude de Anthrophic. Por ejemplo, esta última empresa puso a disposición de los usuarios una versión beta de una IA generativa que puede controlar tu computadora y puede realizar tareas como si un humano hiciese los clics. Una suerte de RPA -*Robot process automation*- anabolizado.

La aceleración y magnitud de la transformación es excepcional y por eso escribimos este documento. Aunque parece que hubiéramos transitado veinte o treinta años de innovación, sólo transcurrieron dos.



Convertirnos en editores de IA y en directores de agentes de IA requiere un desaprendizaje y aprendizaje a gran escala de toda la humanidad.

II

El momento Big Bang

00101

II. El momento Big Bang de la IA Generativa

La publicación de Vaswani et al. (2017), *Attention is All You Need*², fue en gran parte el disparador de todo este fenómeno. Allí se sentaron las bases teóricas para lograr que la IA diera un gran salto en materia de aprendizaje profundo al introducir la arquitectura de redes neuronales *Transformer*. De allí en más, fueron surgiendo modelos más grandes, entrenados con cantidades astronómicas de datos. Y mientras la industria se encontraba estancada en la IA estrecha entrenando cada sistema para tareas acotadas, concretas y específicas, los modelos de lenguaje se posicionaron como la excepción. Así, evolucionaron hacia sistemas generales multitarea, superando incluso el rendimiento humano en muchos casos, liderando los avances y eclipsando el protagonismo en todo el campo de la IA.

El libro *ChatGPT vs. GPT-4: imperfecto por diseño*³ publicado en marzo de 2023 sintetizó los resultados de nuestra investigación y seguimiento, que comenzó con las primeras generaciones de GPT (GPT-1 y GPT-2), y luego continuó con el análisis de las versiones posteriores. Todo este trabajo culminó en uno de los primeros conjuntos de pruebas independiente, de los más grandes y completos hasta ese momento, con más de 700 ejemplos que realizamos durante diciembre, enero y febrero.

A partir de allí, y especialmente a lo largo de este último año, la historia de ChatGPT ha estado marcada por una serie de numerosos eventos notables, desde escándalos empresariales y debates legislativos, hasta profecías sobre la llegada de máquinas superinteligentes al estilo *terminator*.

El combo que viene junto a este tsunami de disrupción, se basa en lanzamientos semanales o diarios de modelos por parte de distintas empresas y organizaciones, avances colaborativos, sistemas de IA generativa de código abierto, plataformas de IA para usos específicos en casi todos los rubros de todas las industrias, incorporación de funcionalidades en los modelos existentes y miles de investigaciones que evidencian mejoras, límites y desafíos. A continuación, intentaremos desglosar los acontecimientos más significativos de forma cronológica.

² Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. Long Beach, CA, USA. Disponible en <https://arxiv.org/abs/1706.03762>. Consultado el 18 de noviembre de 2024.

³ Corvalán, JG, Estevez, E., Le Fevre Cervini, EM, Schapira, D. y Simari, G. (2023). *ChatGPT vs GPT-4: ¿Imperfecto por diseño? Explorando los límites de la inteligencia artificial conversacional* . Buenos Aires: Thomson Reuters La Ley. Disponible en: <https://ialab.com.ar/webia/wp-content/uploads/2023/03/Libro-ChatGPT-vs-GPT-4.-UBA-Thomson-Reuters-La-Ley.pdf> (consultado el 21/11/2024)

III

La llegada de Copilot: Microsoft vs. Google



III. La llegada de Copilot: Microsoft vs. Google

La primera gran incógnita que generó ChatGPT fue cómo revolucionaría las búsquedas web. Para qué recurrir a Google, leyendo mucho y filtrando información, si el nuevo chatbot de OpenAI puede elaborar una respuesta completa con solo darle una instrucción. Inmediatamente, comenzaron las especulaciones sobre si la nueva herramienta constituía una verdadera amenaza para la empresa liderada por Sundar Pichai, que durante años había dominado el negocio, prácticamente sin competidores⁴.

Sin embargo, pronto se hicieron evidentes sus limitaciones. Al no tener conexión a internet, la información que brindaba el chatbot estaba desactualizada. Por ejemplo, a la pregunta, *¿quién ganó el último mundial de fútbol?*, no podía responder de manera correcta y se limitaba a decir que su base de entrenamiento llegaba hasta septiembre de 2021. Además, no revelaba sus fuentes y si se le preguntaba por ellas, no era capaz de identificarlas.

También aparecieron las “alucinaciones”, fabulaciones o salidas sin correspondencia con las bases de datos con las que se entrenó al modelo⁵. Con frecuencia ChatGPT inventaba información falsa y la presentaba con apariencia de verdadera, de forma coherente o plausible. Ello producía confusión en los usuarios, especialmente si no tenían conocimientos anteriores sobre el tema en cuestión⁶. Aunque estas limitaciones persisten⁷, ya existe todo un esquema de abordaje para mitigarlas.

En este contexto, Microsoft, principal accionista de OpenAI vió la oportunidad de incorporar el famoso modelo de lenguaje a su nuevo buscador Bing que se lanzó el 7 de febrero de 2023. El objetivo era abordar la grieta entre las expectativas de los usuarios al

⁴ Grant Nico y Metz Cade, “A new Chatbot is a Code Red for Google’s Search Business” New York Times, diciembre 2022, disponible en: <https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html> (consultado el 12 de marzo de 2024).

Paresh Dave, “ChatGPT inauguró una nueva era en las búsquedas en internet. Microsoft podría arruinarla”, The Wired, marzo 2023, disponible en: <https://es.wired.com/articulos/chatgpt-inauguro-una-nueva-era-en-busquedas-en-internet-microsoft-podria-arruinarla> (consultado el 21 de marzo de 2024).

⁵ Yehuda, Y., Malkiel, I., Barkan, O., Weill, J., Ronen, R., & Koenigstein, N. (2024). *InterrogateLLM: Zero-Resource Hallucination Detection in LLM-Generated Answers*. Disponible en <https://arxiv.org/abs/2403.02889v3> (consultado el 21/11/24)

⁶ Karen Weise y Cade Metz, “When A.I. Chatbots Hallucinate” The New York Times, Mayo 2023, disponible en <https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html> (consultado el 9 de abril del 2024)

⁷ Xu Ziwei, Jain Sanjay, Kankanhalli Mohan, “Hallucination is Inevitable: An Innate Limitation of Large Language Models” arXiv:2401.11817, enero 2024, disponible en: <https://arxiv.org/pdf/2401.11817> (consultado el 25 de julio de 2024).

realizar una búsqueda y las respuestas de los motores convencionales. Según los representantes de la compañía, a pesar de que se realizan 10.000 millones de búsquedas cada día, aproximadamente la mitad de ellas no obtienen respuestas satisfactorias⁸.

Aquí Microsoft se propuso crear una experiencia unificada implementando mejoras o novedades en tres dimensiones: Edge, Bing y Copilot.

En primer lugar, Edge es un navegador que permite a los usuarios acceder a internet mientras que Bing es un motor de búsqueda para encontrar contenido en línea y que a su vez está disponible desde otros navegadores como por ejemplo, Firefox o Safari. Los motores son la principal forma en que las personas encuentran información y por lo tanto, establecen los criterios para decidir qué sitios web aparecerán entre los primeros resultados y qué otros quedarán relegados. De ahí que la competencia entre tecnológicas por controlar e innovar en la puerta de entrada al mundo del contenido en línea sea tan feroz.

Por su parte, Copilot es un asistente de IA basado en la tecnología de OpenAI que puede interactuar con los usuarios en forma de chat conversacional. Al integrarse a Bing⁹, se propone mejorar la experiencia de búsqueda en línea. Cuando el usuario introduce un prompt, el buscador revisa los resultados de toda la web y los resume para ofrecer respuestas individualizadas y contextualizadas a preguntas complejas, añadiendo explicaciones. Por ejemplo, se pueden obtener instrucciones detalladas sobre cómo sustituir huevos por otro ingrediente justo cuando se está cocinando, sin tener que desplazarse por varias recetas¹⁰.

De esta forma, el copiloto del usuario en la web reúne lo mejor de los dos mundos. Por un lado, supera ciertas limitaciones de ChatGPT-3.5 porque tiene acceso a datos en línea actualizados. Por el otro, amplía y mejora a Bing porque permite introducir

⁸ Yusuf Mehdi, “*La reinvención de las búsquedas con un nuevo Microsoft Bing, potenciado por IA, y Edge, un copiloto para la web*”, Microsoft Prensa, Febrero 2023, disponible en <https://news.microsoft.com/es-es/2023/02/08/la-reinvencion-de-las-busquedas-con-un-nuevo-microsoft-bing-potenciado-por-ia-y-edge-un-copiloto-para-la-web/> (consultado el 26 de marzo de 2024).

⁹ Lo cierto es que Copilot no está solo integrado en Bing, también puede encontrarse de forma independiente en copilot.microsoft.com o en sus aplicaciones móviles, disponibles en App Store para iOS y en Google Play para Android. También está en el chatbot de Microsoft Edge accesible desde una barra lateral. Pastor Javier, “*Bing Chat ya no se llama Bing Chat. Ahora se llama Copilot, y se confirma como la gran apuesta de futuro de Microsoft*”, Xataka, Noviembre 2023, disponible en <https://www.xataka.com/servicios/bing-chat-no-se-llama-bing-chat-ahora-se-llama-copilot-tiene-version-empresarial-copilot-pro> (consultado el 14 de abril de 2024).

¹⁰ Yusuf Mehdi, “*La reinvención de las búsquedas con un nuevo Microsoft Bing, potenciado por IA, y Edge, un copiloto para la web*”, Microsoft Prensa, Febrero 2023, disponible en <https://news.microsoft.com/es-es/2023/02/08/la-reinvencion-de-las-busquedas-con-un-nuevo-microsoft-bing-potenciado-por-ia-y-edge-un-copiloto-para-la-web/> (consultado el 26 de marzo de 2024).

búsquedas de hasta 16.000 caracteres¹¹. Así, al ingresar una consulta, la interfaz se divide en dos: primero los resultados tradicionales a los que estamos acostumbrados; segundo, una respuesta personalizada, basada en cuatro o cinco fuentes seleccionadas con sus links para comprobar la información¹².

Pero esto no es todo. Microsoft ha incluido esta herramienta en todas las aplicaciones de su paquete Office. Por ejemplo, con darle una instrucción, Copilot en PowerPoint puede convertirlas en una presentación de forma automática que luego el usuario podrá modificar y mejorar¹³. También es posible interactuar con un documento, generar imágenes y si tu plan es incluirlo más allá de los momentos de trabajo, puede ser tu asistente de ocio. La herramienta ofrece la capacidad de planear tus vacaciones o ser tu entrenador de gimnasia. Para esto, hasta te sugiere opciones de *prompts* prediseñadas como “sabrosas mezclas de condimentos para pollo” o “consejos para prevenir lesiones durante el entrenamiento”.

En fin, a pesar de estas nuevas funcionalidades en los productos de Microsoft, lo cierto es que la posición dominante de Google continúa consolidada y los datos demuestran que los usuarios prefieren ampliamente esta alternativa. En 2019, antes de la explosión de la IA generativa, el porcentaje del mercado de buscadores que representaba Bing se mantuvo entre 2,31% y 2,63%, mientras que Google alcanzó el 92,96%. En noviembre de 2024, Bing concentra el 4% del mercado, mientras que Google mantiene el porcentaje con un leve descenso hacia el 90%¹⁴.

Pero la batalla no solo se da en las búsquedas web, sino que también alcanzó los grandes modelos de lenguaje, ámbito en el que Microsoft sí obtuvo revancha.

¹¹ Patton Seth, *What's New in Copilot | August 2024*, Microsoft, 30 de Agosto 2024, disponible en: <https://techcommunity.microsoft.com/blog/microsoft365copilotblog/whats-new-in-copilot-august-2024/4226565> (consultado el 21 de noviembre de 2024).

¹² Microsoft Prensa, “*Llega Microsoft Copilot, el compañero de Inteligencia Artificial para el día a día*”, Microsoft Centro de Noticias, Septiembre 2023, disponible en <https://news.microsoft.com/es-es/2023/09/21/llega-microsoft-copilot-el-companero-de-inteligencia-artificial-para-el-dia-a-dia/> (consultado el 14 de abril de 2024).

¹³ Colette Stallbaumer, “*Introducing Microsoft 365 Copilot—A whole new way to work*”, Microsoft Blog, Marzo 2023, disponible en https://www.microsoft.com/en-us/microsoft-365/blog/2023/03/16/introducing-microsoft-365-copilot-a-whole-new-way-to-work/?ranMID=24542&ranEAID=nOD/rLJHOac&ranSiteID=nOD_rLJHOac-ukeW0ZTGggqEheTLk2gouA&epi=nOD_rLJHOac-ukeW0ZTGggqEheTLk2gouA&irgwc=1&OCID=AIDcmm549zy227_aff_7791_1243925&ranMID=46133&ranEAID=uX9G0lYjaAY&ranSiteID=uX9G0lYjaAY-ljcSufbaf80UR5_TrZ8FSw&epi=uX9G0lYjaAY-ljcSufbaf80UR5_TrZ8FSw&tduid=%28ir_qaeayt3tcgkfdyq22dxnhxqhim2xdkpapmdxgnwz00%29%287791%29%281243925%29%28uX9G0lYjaAY-ljcSufbaf80UR5_TrZ8FSw%29%28%29&irclickid=_qaeayt3tcgkfdyq22dxnhxqhim2xdkpapmdxgnwz00 (consultado el 14 de abril de 2024).

¹⁴ Los datos sobre las cuotas de mercado de los diferentes buscadores han sido extraídos de StatCounter, disponible en: <https://gs.statcounter.com/search-engine-market-share> (consultado el 21 de noviembre de 2024).

Tras el lanzamiento de ChatGPT, las organizaciones debieron comenzar a abrir sus sistemas para demostrar que estaban a la altura. Recordemos que hasta este momento el rendimiento de los modelos se medía en relación a la cantidad de parámetros que los desarrolladores declaraban que sus redes neuronales tenían, siempre bajo el lema: cuanto más grande, mejor. Mientras tanto, puertas adentro del laboratorio se hacían experimentos que revelaban métricas que eran imposibles de auditar. Algunas excepciones consistieron en accesos restringidos para usuarios en lista de espera¹⁵.

Frente al furor de ChatGPT Google salió a la cancha. Desde hace tiempo desarrollaba su propia IA que prometía ser la más avanzada, aunque los usuarios no podían comprobarlo directamente. Finalmente, en marzo de 2023, permitió el acceso abierto al sistema Bard que se iría ampliando con el tiempo por países y lenguas, para recibir retroalimentación de la comunidad¹⁶. En diciembre del año pasado, cambiaron el nombre y nos presentaron la familia de modelos Gemini, incluyendo Ultra, Pro y Nano.

Microsoft presenta Copilot, un asistente de IA que se integra al motor de búsqueda, pero la posición dominante de Google continúa consolidada. La competencia también alcanza los grandes modelos de lenguaje, donde Microsoft obtiene revancha.

¹⁵ Al modelo GPT-3 podían acceder desarrolladores y personas autorizadas por la empresa mediante una “lista blanca” de interesados. BROCKMAN Greg, MURATI Mira y WELINDER Peter, “OpenAI API. We’re releasing an API for accessing new AI models developed by OpenAI”, OpenAI Blog, junio 2020, disponible en <https://openai.com/blog/openai-api> (consultado el 21/03/2024).

¹⁶ Hsiao Sissie, Collins Ellie, “Try Bard and Share your Feedback” Google, marzo 2023, disponible en: <https://blog.google/technology/ai/try-bard/> (consultado el 22 de marzo de 2024).

IV

Lanzamiento de la API: escalando la integración de ChatGPT

IV. Lanzamiento de la API: escalando la integración de ChatGPT

El 1 de marzo de 2023 OpenAI lanzó la API de ChatGPT, otro avance significativo en materia de accesibilidad y democratización de la IA. Hasta ese momento, el modelo de lenguaje solo estaba disponible a través de su interfaz de usuario. La API, o interfaz de programación de aplicaciones, actúa como un puente que conecta distintos sistemas de software, facilitando su integración¹⁷. Esta nueva vía de acceso permite a los programadores crear nuevos desarrollos aprovechando los existentes y optimizando recursos.

En el caso de ChatGPT el impacto fue profundo: empresas de todos los tamaños, incluso pequeñas, ahora tenían a su alcance la tecnología más poderosa para construir sobre ella y explotarla comercialmente por un precio accesible. A través de la API pueden optimizar modelos que rivalizan en funcionalidad y precisión con los de gigantes tecnológicos. Ello no solo nivela el campo de juego, sino que acelera la adopción y la creación de sistemas generativos personalizados y agentes artificiales, capaces de transformar industrias completas. Veamos con más detalle cómo todo esto es posible.

La API da acceso al modelo base o *foundation model* (modelo fundacional en español). Se trata de una versión del modelo de lenguaje pre-entrenado por OpenAI con astronómicas cantidades de datos generales, no adaptado a ningún contexto específico, lo que permite que sea lo suficientemente flexible y se constituya como base o soporte para construir aplicaciones especializadas¹⁸. El adjetivo “*foundation*” hace referencia al fundamento de algo, a la idea de un sistema de IA que, dadas ciertas condiciones, podrá ser optimizado para mejorar su precisión en una tarea o ámbito específico, como la traducción de idiomas, la interpretación de normas, o la elaboración de recetas de cocina.

Aquí también se suman dos ingredientes clave para optimizar los modelos: una técnica de adaptación y más datos. En relación a lo primero, se destacan dos enfoques populares: por un lado, *transfer learning* (aprendizaje por transferencia) y *fine-tuning* (ajuste fino)¹⁹ y por otro, RAG (generación aumentada de recuperación).

¹⁷ Se trata de un protocolo que permite la comunicación entre dos aplicaciones a través de un conjunto de reglas y que se utiliza para desarrollar e integrar su software. Fernández Yúbal, “API: qué es y para qué sirve” Xataka, agosto 2019, disponible en: <https://www.xataka.com/basics/api-que-sirve> (consultado el 25 de marzo de 2024).

¹⁸ Sobre el concepto de foundation models ampliar en: Bommasani, Rishi, Liang, Percy, “Reflections on Foundation Models” Stanford HAI, Octubre, 2021, Disponible en: <https://hai.stanford.edu/news/reflections-foundation-models> (consultado el 8/1/24).

¹⁹ Hemos observado que la diferencia entre *transfer learning* y *fine-tuning* no está clara. Lo cierto es que no existe un consenso general sobre si se trata de términos intercambiables, de una relación género-especie o simplemente, de técnicas similares con marcadas diferencias. Luego de una revisión bibliográfica exhaustiva nos inclinamos por la última opción.

Un ejemplo para ilustrar este fenómeno. Imaginemos que es abogado y ahora quiere empezar una nueva carrera de grado: Licenciatura en Relaciones Internacionales. Seguramente, habrá materias que no necesitará estudiar de nuevo, como Derecho Constitucional. No hará falta que nadie le explique qué es una Constitución, cuál es su importancia o cuál la jerarquía que le da a un orden jurídico. Lo que usted está haciendo es aprovechar conocimientos previamente adquiridos, tal como lo hace el *transfer learning*. De acuerdo con esta definición, se utiliza un modelo pre-entrenado como punto de partida para generar un nuevo modelo que realice otras tareas, diferentes pero relacionadas²⁰. Esta similitud implica la existencia de conocimiento común para realizar ambas actividades, el cual no debe ser aprendido desde cero, sino reaprovechado²¹.

En términos técnicos, para hacer *transfer learning* solo deben ajustarse las capas finales de una red neuronal en que se basa el modelo y sus parámetros asociados. Las capas iniciales que representan la gran mayoría y suelen capturar ciertas características como la comprensión básica del lenguaje, son aquellas útiles para una amplia gama de tareas y se mantienen congeladas e inalterables²².

Ahora volvamos al ejemplo anterior. Sigue interesado en Derecho Internacional, pero en lugar de hacer otra carrera de grado, se inclina por una especialización. Una maestría en la materia le permite profundizar sus conocimientos y volverse más experto en esta rama de la profesión. Algo similar se logra haciendo *fine-tuning* a un modelo de IA, al re-entrenarlo con un conjunto de datos más específico²³. En este caso, también se aprovechan las capacidades originales, mientras se satisfacen casos de uso más especializados maximizando el rendimiento en un ámbito específico. Ello puede lograrse proporcionando ejemplos adicionales de la tarea que se quiere realizar o mejorar, que ajustarán algunas o todas las capas de la red neuronal²⁴.

²⁰ Chaba Victor, “*Understanding the Differences: Fine-Tuning vs. Transfer Learning*”, Lux Tech Academy, Agosto 2023, Disponible en: <https://dev.to/luxacademy/understanding-the-differences-fine-tuning-vs-transfer-learning-370> (consultado el 4 de mayo de 2024).

²¹ La idea está en que si ambos se desarrollan para realizar tareas similares, entonces el conocimiento común se puede compartir entre ellos. Ver: *¿Qué es el transfer learning y qué ventajas tiene?*, Revista Unir, Noviembre 2023, Disponible en: <https://www.unir.net/ingenieria/revista/transfer-learning/#:~:text=El%20transfer%20learning%20se%20usa,solventar%20un%20problema%20nuevo%20similar> (consultado el 21 de mayo de 2024).

²² Kumar Ajistesh, “*Transfer Learning vs Fine Tuning LLMs: Differences*”, Analytics Yogi, Enero 2024, Disponible en: <https://vitalflux.com/transfer-learning-vs-fine-tuning-differences/> (consultado el 21 de mayo de 2024).

²³ Lev Craig, “*fine-tuning*”, TechTarget, Disponible en: <https://www.techtarget.com/searchenterpriseai/definition/fine-tuning> (consultado el 21 de mayo de 2024).

²⁴ Kumar Ajistesh, “*Transfer Learning vs Fine Tuning LLMs: Differences*”, Analytics Yogi, Enero 2024, Disponible en: <https://vitalflux.com/transfer-learning-vs-fine-tuning-differences/> (consultado el 21 de mayo de 2024).

Finalmente para comprender RAG, imaginemos que se da cuenta que el objetivo de aprender Derecho Internacional era ampliar los servicios de su estudio, y decide que lo más conveniente es buscar un socio que cubra esta rama del derecho. En este caso estará complementando su conocimiento general con una fuente externa específica, extrayendo información relevante y contextualizando mejor su asesoramiento. Algo parecido es lo que ocurre con esta última técnica, que reúne la generación de texto con información de otras fuentes de datos, ya sean privadas o propietarias, sin alterar el modelo base²⁵. En este sentido, se vuelve más flexible o actualizable. A diferencia del *fine-tuning* en que el modelo queda estático una vez re-entrenado con datos específicos, la fuente externa a la que el modelo generativo consulta a través de RAG, puede ser fácilmente modificada o ampliada.

En síntesis, estos enfoques tienen el potencial de mejorar la precisión de los modelos de lenguaje en entornos, industrias o flujos de trabajo específicos, ya sea proveyendo información contextual que no está dentro de la base de datos de entrenamiento u optimizando el formato, el estilo o el razonamiento de las salidas producidas²⁶. Asimismo, al construir sobre el modelo base, permiten ahorrar costos, tiempo e infraestructura computacional requerida para desarrollar una IA sofisticada de cero.



La API de ChatGPT tuvo un impacto profundo: empresas de todos los tamaños ahora tienen a su alcance la tecnología más poderosa para construir sobre ella y explotarla comercialmente por un precio accesible. Pueden optimizar modelos que realicen en funcionalidad y precisión con los de gigantes tecnológicas.



²⁵ Sobre RAG ampliar en: OpenAI “Optimizing LLM Accuracy” disponible en: <https://platform.openai.com/docs/guides/optimizing-llm-accuracy> (consultado el 25 de noviembre de 2024). Asimismo ver: AWS, “¿Qué es la RAG (generación aumentada por recuperación)?”, disponible en: <https://aws.amazon.com/what-is/retrieval-augmented-generation/> (consultado el 25 de noviembre de 2024).

²⁶ OpenAI “Optimizing LLM Accuracy” disponible en: <https://platform.openai.com/docs/guides/optimizing-llm-accuracy> (consultado el 25 de noviembre de 2024).

GPT-4: de la promesa open a un presente profit



V. GPT-4: De la promesa *open* a un presente *profit*

A sólo dos meses del lanzamiento de ChatGPT, el 1 de febrero de 2023, OpenAI introdujo ChatGPT Plus. Como si el furor del nuevo sistema por sí solo no nos tuviera lo suficientemente entretenidos, la empresa desarrolladora supo aprovechar la atención que acaparaba para introducir nuevas innovaciones. Sin embargo, esta vez no serían gratuitas.

A principios de diciembre de 2022, pocos días después del lanzamiento inicial, Sam Altman ya había anunciado en su cuenta de X que OpenAI monetizaría en algún momento y de alguna manera el servicio, ya que los costos de computación eran excesivos²⁷. Se volvió a un clásico del marketing: cuando las herramientas se convierten en hábito, los grandes descuentos tienden a desaparecer²⁸.

Aquí había dos caminos posibles sobre la gratuidad del uso. Restringir el acceso por alta demanda y enfocar el negocio en la recopilación de datos, información y finalmente perfilar a los usuarios como modelo de negocios²⁹. Sin embargo, no quedaba claro con qué frecuencia o en qué momentos este grupo de usuarios quedarían excluidos del servicio. La segunda versión, sería paga y podría llegar a costar hasta 42 dólares mensuales.

Finalmente se lanzó la versión Plus que coexistió con la gratuita, con un servicio de suscripción de 20 dólares al mes y que ofreció ventajas exclusivas: acceso ininterrumpido a ChatGPT en cualquier momento, incluyendo períodos de mayor demanda; respuestas más rápidas y completas garantizadas, junto con prioridad de acceso a las últimas funciones y mejoras disponibles³⁰. Con esta novedad, la empresa liderada por Sam Altam abandonó definitivamente su carácter de “*non-profit*”.

La nueva versión de suscripción sirvió como precursora de otro lanzamiento posterior aún más importante. El reemplazo de GPT-3.5 por GPT-4 sería inminente y, otra

²⁷ El tuit de Sam Altman se encuentra disponible en: https://twitter.com/sama/status/1599669571795185665?ref_src=twsr%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1599669571795185665%7Ctwgr%5Ece12fbf16408216c8a6bb4b0d3f9e496ed3028d0%7Ctwcon%5Es1&ref_url=https%3A%2F%2Fwww.pcworld.com%2Farticle%2F1481398%2Fchatgpt-may-charge-42-mo-for-a-paid-tier-as-microsoft-invests-again.html (consultado el 31 de marzo de 2024).

²⁸ Sundar Sindhu, “*ChatGPT is testing a paid version — here's what that means for free users*” Business Insider, Enero 2023, disponible en: <https://www.businessinsider.com/will-chatgpt-always-be-free-openai-paid-professional-version-2023-1> (consultado el 10 de abril de 2024).

²⁹ Sobre cómo OpenAI y otras empresas de GenAI recopilan y tratan los datos de los usuarios ver: CORVALÁN Juan G., PAPINI, Carina M. y HELEG Giselle, “*Privacidad y tratamiento automatizado de datos personales en la IA generativa. Los grandes modelos de lenguaje (LLM): ChatGPT, Bing y Bard*” en Corvalán Juan G. “*Tratado de inteligencia artificial y derecho*”, Thompson Reuters La Ley, 2da edición, 2023.

³⁰ Página oficial de Open AI “*Presentamos ChatGPT Plus*”, disponible en: <https://openai.com/blog/chatgpt-plus> (consultado el 04/01/24).

vez, el público se sumergió en una nueva ola de especulaciones que se centraron en capacidades multimodales.

Era evidente que esperábamos un modelo más capaz. Y ello casi siempre es sinónimo de un modelo más grande. La pregunta era: ¿cuánto más grande? Teniendo en cuenta que GPT-3.5 contaba con 175.000 millones de parámetros³¹, los rumores sobre GPT-4 lo llevaban a entre 100 billones³² y 1 trillón de parámetros³³. Sin embargo, los más osados en sus apuestas no hablaron de números, sino que introdujeron en el debate a la inteligencia artificial general (AGI)³⁴. Según esta corriente, el nuevo modelo sería tan impresionante que nos conduciría a una nueva generación de sistemas inteligentes con capacidades diversas. Y como veremos más adelante, no estaban tan equivocados.

El 14 de marzo de 2023, GPT-4 finalmente salió a la luz mediante un comunicado de OpenAI en su página oficial³⁵. Podrían probar sus capacidades los suscriptores de ChatGPT Plus que se anotaran en una lista de espera³⁶. Según las pruebas realizadas por la empresa, la nueva versión era significativamente más avanzada que su antecesora: GPT-4 tiene un 82% menos de probabilidades de responder a solicitudes de contenido no permitido o malicioso, como discurso del odio, y un 40% más de probabilidades de arrojar respuestas sin reflejar opiniones subjetivas³⁷. Además, OpenAI puso a prueba a GPT-4 en varios exámenes académicos, incluyendo el Uniform Bar Exam (utilizado en Estados Unidos para ejercer la abogacía) y el LSAT (necesario para ingresar a Columbia Law School).

³¹ El aumento del tamaño del modelo, depende de la cantidad de parámetros que tiene, lo que permite absorber más información de los datos de entrenamiento y obtener resultados más precisos sobre los datos nuevos. Ver: Corvalan, Gustavo Juan; Estevez, Elsa; Le Fevre, Maria Enzo; Shapira, Debora; Simari, Gerardo (dirección), "ChatGPT vs GPT-4 ¿Imperfecto por diseño? Explorando los límites de la inteligencia artificial conversacional", UBA IALAB, 2023, disponible en: <https://ialab.com.ar/wp-content/uploads/2023/03/Libro-ChatGPT-vs-GPT-4.-UBA-Thomson-Reuters-La-Ley.pdf> (consultado el 31 de marzo de 2024).

³² "What to expect from GPT-4" TechCircle, marzo 2023, disponible en: <https://www.techcircle.in/2023/03/13/what-to-expect-from-gpt-4> (consultado el 10 de abril de 2024).

³³ Ver: Corvalan, Gustavo Juan; Estevez, Elsa; Le Fevre, Maria Enzo; Shapira, Debora; Simari, Gerardo (dirección), "ChatGPT vs GPT-4 ¿Imperfecto por diseño? Explorando los límites de la inteligencia artificial conversacional", UBA IALAB, 2023, disponible en: <https://ialab.com.ar/wp-content/uploads/2023/03/Libro-ChatGPT-vs-GPT-4.-UBA-Thomson-Reuters-La-Ley.pdf> (consultado el 31 de marzo de 2024).

³⁴ "What to expect from GPT-4" TechCircle, marzo 2023, disponible en: <https://www.techcircle.in/2023/03/13/what-to-expect-from-gpt-4> (consultado el 10 de abril de 2024).

³⁵ Información extraída de la página oficial de OpenAI, disponible en: <https://openai.com/gpt-4> (consultada el 05/01/24)

³⁶ Will Douglas Heaven "OpenAI guarda silencio sobre por qué GPT-4 es mejor que ChatGPT", Marzo 2023, disponible en: <https://www.technologyreview.es/s/15195/openai-guarda-silencio-sobre-por-que-gpt-4-es-mejor-que-chat-gpt> (consultado el 10 de abril de 2024).

³⁷ Sobre estas pruebas y sus resultados ampliar en la página oficial de OpenAI, disponible en: <https://openai.com/gpt-4> (consultada el 06/01/2024)

También GPT-4 fue evaluado en la sección cuantitativa del GRE para comprender y razonar con conceptos matemáticos. En la mayoría de estas pruebas, GPT-4 obtuvo calificaciones sobresalientes, mostrándose superior a la mayoría de los estudiantes humanos³⁸.

De manera similar, en la publicación “*ChatGPT vs GPT-4: Imperfecto por diseño*” advertimos una mejora significativa de hasta el 30% del modelo de última generación respecto a su antecesor, en todos los ámbitos de evaluación: funcionalidades, ciertos tipos de razonamiento, preguntas de derecho argentino, salud y sesgos³⁹.

¿A qué se deben estas mejoras? ¿Cuán grande es el modelo que logra estas métricas superiores? Por primera vez, OpenAI se guardó el secreto y se volvió menos “open” que nunca. ¿El motivo? La feroz competencia y creciente rivalidad que se desata en el campo de desarrollo de la GenAI⁴⁰.

Como suele ocurrir cada vez que se anuncia un modelo de este tipo, un *paper* o reporte con ciertas especificaciones técnicas es publicado. GPT-4 no fue la excepción⁴¹. El problema es que sus páginas no contienen la información que esperábamos. Sabemos que es una de las IA más potentes, pero no cómo “razona” o por qué se volvió más robusta. Como si la opacidad intrínseca en el funcionamiento de las redes generativas pre-entrenadas (basadas en *Transformers*)⁴² y la falta de información sobre los datos de entrenamiento fuera poco, esta nueva forma de hermetismo nos impide acceder a ciertos aspectos de su arquitectura interna y entender el motivo por el cual este modelo es superior. Después de todo, OpenAI abandonó la lógica del mago enmascarado, famoso por revelar los trucos.

³⁸ OpenAI, *GPT-4 Technical Report*, Marzo 2023, disponible en: <https://cdn.openai.com/papers/gpt-4.pdf>

³⁹ Corvalán, JG, Estevez, E., Le Fevre Cervini, EM, Schapira, D. y Simari, G. (2023). *ChatGPT vs GPT-4: ¿Imperfecto por diseño? Explorando los límites de la inteligencia artificial conversacional* . Buenos Aires: Thomson Reuters La Ley. Disponible en: <https://ialab.com.ar/webia/wp-content/uploads/2023/03/Libro-ChatGPT-vs-GPT-4-UBA-Thomson-Reuters-La-Ley.pdf> (consultado el 21/11/2024).

⁴⁰ Will Douglas Heaven “OpenAI guarda silencio sobre por qué GPT-4 es mejor que ChatGPT”, Marzo 2023, disponible en: <https://www.technologyreview.es/s/15195/openai-guarda-silencio-sobre-por-que-gpt-4-es-mejor-que-chat-gpt> (consultado el 6 de enero de 2024).

⁴¹ El reporte se encuentra disponible en: “*GPT-4 Technical Report*” OpenAI, marzo 2023, disponible en: <https://cdn.openai.com/papers/gpt-4.pdf> (consultado el 21 de mayo de 2024).

⁴² Las redes neuronales de las IAs imitan el funcionamiento del cerebro humano, pero una de las principales características de estas es que funcionan como cajas negras. Esto significa que, si bien podemos observar los datos de entrada y la salida final, no podemos comprender completamente cómo la red llega a esa salida. Esto sucede por la complejidad de las redes neuronales. Las mismas están formadas por miles o incluso millones de neuronas artificiales interconectadas. Cada neurona realiza cálculos matemáticos simples, y la combinación de estas acciones genera el resultado final. La interacción entre estas neuronas es tan compleja que resulta difícil, o incluso imposible, para los humanos seguir el proceso paso a paso.

GPT-4 Turbo. La primera gran actualización de GPT-4 llegó el 6 de noviembre de 2023 con GPT-4 Turbo. Esta nueva versión del modelo de lenguaje más sofisticado amplía la ventana de contexto de 32.000 tokens a 128.000, lo que le permite mantener el hilo de la conversación durante una mayor cantidad de interacciones con el usuario y retener el contexto⁴³. Además, sus datos de entrenamiento se actualizaron hasta abril de 2023, y se redujeron los precios de la API para hacerlo más accesible a los desarrolladores.

Divide y vencerás: MoE (Mixture of experts). Un mes más adelante, en diciembre de 2023, OpenAI publicó un nuevo reporte técnico sobre GPT-4, ahora sí revelando algunos detalles más precisos. Pareciera ser que la superioridad del modelo respecto a su competencia no reside en el tamaño, si no en la arquitectura “Mezcla de expertos” (MoE, por sus siglas en inglés)⁴⁴.

Este diseño, creado por otros, fue mejorado para GPT-4⁴⁵. En esencia, se trata de parámetros del modelo que se dividen en distintos conjuntos y se especializan en tareas determinadas. Cada uno de estos grupos se denomina “experto” y solo se activan aquellos que son útiles para resolver la petición del usuario en cada caso. Ello conduce a un menor costo computacional, porque no es necesario que trabaje todo el sistema frente a cada entrada. Por ejemplo, puede haber un experto en un lenguaje de programación como Python, otro en verificación fáctica y otro en seguridad. Se ha dicho que GPT-4 alberga 16 modelos expertos, cada uno con alrededor de 111 mil millones de parámetros⁴⁶. Todo esto no ha sido confirmado oficialmente por la empresa.

⁴³ Ver: Corvalan, Gustavo Juan; Estevez, Elsa; Le Fevre, Maria Enzo; Shapira, Debora; Simari, Gerardo (dirección), “ChatGPT vs GPT-4 ¿Imperfecto por diseño? Explorando los límites de la inteligencia artificial conversacional”, UBA IALAB, 2023, disponible en: <https://ialab.com.ar/wp-content/uploads/2023/03/Libro-ChatGPT-vs-GPT-4.-UBA-Thomson-Reuters-La-Ley.pdf> (consultado el 15/01/24)

⁴⁴ O.V.Gendelman, “Escape of a harmonically forced particle from an infinite-range potential well: a transient resonance” Abril de 2017, arXiv:1704.06924, disponible en: <https://arxiv.org/pdf/1704.06924.pdf> (consultado el 21 de junio de 2024).

⁴⁵ Carlos Santana Vega. “Lo que OpenAI NO quería que supieras sobre GPT4 - (De los MoEs a Mixtral)” Diciembre, 2023. Disponible en: <https://www.youtube.com/watch?v=Sfnu5OmAITA> (consultado el 08/01/24).

⁴⁶ Betts Sean, “Peering Inside GPT-4: Understanding Its Mixture of Experts (MoE) Architecture” Medium, Julio 2023, disponible en: <https://medium.com/@seanbetts/peering-inside-gpt-4-understanding-its-mixture-of-experts-moe-architecture-2a42eb8bdcb3> (consultado el 21 de junio de 2024).

VI

¿La nueva apuesta de Pascal?
Riesgos catastróficos de la IA



VI. ¿La nueva apuesta de Pascal? Riesgos catastróficos de la IA

A fines de marzo de 2023 y como consecuencia del lanzamiento de GPT-4, se publicó una carta abierta firmada por más de miles de expertos y empresarios que pidió pausar por seis meses los “*modelos de caja negra impredecibles cada vez más grandes con capacidades emergentes*”⁴⁷. los experimentos sobre los modelos equivalentes a GPT-4. Personalidades como Elon Musk, Yoshua Bengio, Steve Woznia, Stuart Russell, Tristan Harris, Gary Marcus, Marc Rotemberg y Yuval Noah Harari consideraron que los sistemas inteligentes avanzados pueden representar cambios profundos en la historia de la vida en la tierra y por lo tanto, deberían ser desarrollados solo cuando estemos seguros de que sus efectos son positivos y sus riesgos controlables⁴⁸.

Pero esa no fue la única advertencia. “*La mitigación del riesgo de extinción por la IA debería ser una prioridad global junto con otros riesgos a escala social, como las pandemias y la guerra nuclear*” dice la declaración publicada dos meses después por el Centro de Seguridad de la IA (*Center for AI Safety*) sin fines de lucro⁴⁹. Esta vez, firmada por Sam Altman, Dario Amodei, Demis Hassabis (ganador del nobel de química 2024), Bill Gates y Geoffrey Hinton (ganador del nobel de física 2024), entre muchas otras figuras.

Por otro lado, los más escépticos sostienen que estas advertencias son exageradas y los riesgos de la IA se están convirtiendo en la nueva apuesta de Pascal. Recordemos que esto se refiere al siguiente fenómeno: aunque no tengamos evidencia certera de la existencia de Dios, e incluso esto parezca improbable, es mejor creer en él, porque si existe y no lo hacemos corremos el riesgo de ir al infierno. Algo similar ocurre con los sistemas inteligentes avanzados⁵⁰. Los riesgos potenciales son tan graves y catastróficos, que aunque sean remotos e inciertos, qué mejor que irnos preparando, aunque esto implique esfuerzos desproporcionados y dejar de lado otros problemas más inmediatos.

⁴⁷ La carta abierta puede encontrarse en: “*Pause Giant AI Experiments: An Open Letter*” Future of Life Institute, 22 de marzo de 2023, disponible en: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (consultado el 1 de abril de 2024).

⁴⁸ La carta abierta puede encontrarse en: “*Pause Giant AI Experiments: An Open Letter*” Future of Life Institute, 22 de marzo de 2023, disponible en: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (consultado el 1 de abril de 2024).

⁴⁹ La declaración puede verse en: “*Statement on AI Risk*”, Center for AI Safety, Mayo 2023, disponible en: <https://www.safe.ai/work/statement-on-ai-risk> (consultado el 5 de abril de 2024).

Sobre esta declaración ampliar en: Roose Kevin, “*A.I. Poses ‘Risk of Extinction,’ Industry Leaders Warn*”, The New York Times, Mayo 2023, disponible en: <https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html> (consultado el 5 de abril).

⁵⁰ Blake Richards, Blaise Aguera y Arcas, Guillaume Lajoie y Dhanya Sridhar, *The Illusion Of AI’s Existential Risk*, Noema, Julio 2023, disponible en: <https://www.noemamag.com/the-illusion-of-ais-existential-risk/> (consultado el 28 de noviembre de 2024).

También es curioso que las mismas personas que apoyan estas cartas sean quienes, al mismo tiempo, desarrollan la tecnología. El magnate sudafricano CEO de Tesla no pudo esperar ni un tercio de año para ignorar su propia petición: ya en julio anunció la creación de una nueva empresa xAI que trataría de ir más allá de los sistemas existentes para competir con Google, Microsoft y OpenAI⁵¹. Por lo demás, muchos otros de los firmantes, admitieron que a pesar de sus esfuerzos creían que la investigación no iba a detenerse realmente⁵².

Pero la contradicción de Elon Musk queda en segundo plano si tenemos en cuenta la controversia que protagonizó Sam Altman unos meses después. No se sabe bien si los riesgos catastróficos de la IA estuvieron en el centro del caos empresarial que culminó con su despido el 18 de noviembre de 2023. Ese día, el consejo directivo de OpenAI lo convocó para informarle de su cese inmediato como CEO de la compañía.

Según medios de comunicación, uno de los principales desencadenantes de la desafectación fue una carta que algunos ingenieros de OpenAI escribieron a la junta directiva advirtiendo sobre un poderoso descubrimiento⁵³ relacionado con AGI que podría amenazar seriamente a la humanidad. Esta advertencia se sumó a las preocupaciones que el consejo ya tenía sobre la comercialización de los avances de la empresa antes de comprender las consecuencias que los mismos acarrearían. Al parecer, las cosas estaban sucediendo de manera frenética y la supervisión no iba a la par del desarrollo. Al menos para el directorio, una actitud reactiva no era suficiente.

Cuando OpenAI decidió reorganizarse internamente en 2019 para adoptar una estructura híbrida, se crearon tensiones entre dos grupos dentro de la compañía. Por un lado, una posición más conservadora, liderada por el científico jefe y miembro de la junta directiva Ilya Sutskever, quien creía que la empresa debía ceñirse a su propósito fundacional y solo implementar la IA con cuidado. Por el otro, una postura más pragmática y orientada al mercado, enfocada en la explotación comercial de la tecnología⁵⁴.

⁵¹ Knight Will, "Hace seis meses, Elon Musk firmó una carta para frenar la IA, pero de hecho se aceleró" The Wired, septiembre 2023, disponible en: <https://es.wired.com/articulos/hace-seis-meses-elon-musk-firmo-una-carta-para-frenar-ia-pero-se-acelero> (consultado el 1 de abril de 2024).

⁵² Knight Will, "Hace seis meses, Elon Musk firmó una carta para frenar la IA, pero de hecho se aceleró" The Wired, septiembre 2023, disponible en: <https://es.wired.com/articulos/hace-seis-meses-elon-musk-firmo-una-carta-para-frenar-ia-pero-se-acelero> (consultado el 1 de abril de 2024).

⁵³ TONG Anna, DASTIN Jeffrey y HU Krystal, "OpenAI researchers warned board of AI breakthrough ahead of CEO ouster, sources say", Reuters, publicado el 23/11/2023, disponible en: <https://www.reuters.com/technology/sam-altmans-ouster-openai-was-precipitated-by-letter-board-about-ai-breakthrough-2023-11-22/> (consultado el 26/07/2024).

⁵⁴ GASCÓN, Marta, "Cronología del caos en OpenAI: del despido de Altman y su reincorporación a la tecnología secreta que pudo detonarlo todo", 20minutos, publicado el 24/11/2023, disponible en:

Lo que parecía en principio una victoria para el primer grupo, pronto se convirtió en un caos empresarial. Para empezar, al despedir a Sam Altman, OpenAI dejó libre a uno de los referentes del rubro para poder emprender proyectos con potencial para convertirse en competencia. Tanto el ex CEO como otras instituciones tecnológicas lo notaron desde el primer momento. Así fue que Microsoft no tardó más de 2 días en anunciar⁵⁵ que había contratado a Sam Altman para que lidere un nuevo equipo de investigación de IA.

A ello, se sumarían más de 500 empleados de OpenAI que pidieron la restitución inmediata de Altman y la renuncia de todos los integrantes del consejo. Si OpenAI no accedía, este grupo de expertos que representaba más del 70% del plantel de la empresa, tenía asegurado un lugar en Microsoft para continuar trabajando en proyectos liderados por sus ex-jefes. Y estaban dispuestos a irse. Así fue que a pesar de haber sido despedido un viernes, para el día miércoles 22 de noviembre OpenAI ya anunciaba en redes sociales⁵⁶ que había llegado a un acuerdo para el regreso de Sam Altman como CEO de la empresa.



Los riesgos potenciales de la IA son tan graves y catastróficos, que aunque sean remotos e inciertos, qué mejor que irnos preparando, aunque esto implique esfuerzos desproporcionados y dejar de lado otros problemas más inmediatos.

<https://www.20minutos.es/tecnologia/actualidad/todo-lo-que-tienes-que-saber-sobre-drama-de-openai-5192993/> (consultado el 26/07/2024).

⁵⁵ NADELLA, Satya, red social X, publicado el 20/11/2023. Disponible en: <https://x.com/satyanadella/status/1726509045803336122> (consultado el 26/07/2024).

⁵⁶ OpenAI, red social X, publicado el 22/11/2023. Disponible en: <https://x.com/OpenAI/status/1727206187077370115?s=20> (consultado el 26/07/2024).

VII

Opciones para personalizar y
editar tu ChatGPT sin programar

VII. Opciones para personalizar y editar tu ChatGPT sin programar

Plugins. Casarse nunca ha sido tan fácil como con la ayuda de ChatGPT. Así lo cree un informático ruso, quien utilizó el modelo de lenguaje para interactuar con más de cinco mil mujeres en Tinder⁵⁷. El candidato logró filtrar perfiles según sus preferencias personales y luego simular su manera de comunicarse, lo que le permitió concertar citas y finalmente contraer matrimonio con una de sus coincidencias. Sin embargo, necesitó de un extra: los plugins.

En marzo de 2023 OpenAI presentó la posibilidad de introducir plugins en ChatGPT⁵⁸. Se trataba de una especie de extensiones que permitían al modelo interactuar con servicios o bases de datos externas para enriquecer sus respuestas o realizar tareas para las cuales fuera necesario acceder a información en tiempo real. El objetivo consistía en que los desarrolladores pudieran introducir funcionalidades extras que no estaban presentes en la versión original del modelo.

Desde recomendaciones de productos en Amazon hasta la reserva de alojamientos en Booking, los plugins abrieron un enorme abanico de posibilidades. Los usuarios contaban con la posibilidad de usar hasta tres de estos accesorios al mismo tiempo. Con esta iniciativa se buscó fomentar una comunidad activa de desarrolladores que colaboraban, compartían ideas y soluciones, contribuyendo al crecimiento y mejora continua de la plataforma. Sin embargo, un ambiente tan dinámico y abierto se vuelve difícil de controlar.

Rápidamente emergieron serias cuestiones de seguridad. Los piratas informáticos comenzaron a explotar vulnerabilidades en estos plugins, lo que aumentó los incidentes de robo y acceso no autorizado a información sensible. Algunos plugins podían contener malware, virus u otros elementos maliciosos si se descargaban desde fuentes no confiables y OpenAI no tenía forma de controlarlo. Se solicitaban más permisos de los necesarios para su funcionamiento, permitiendo que los atacantes accedan al historial de chat, obtengan información personal y ejecuten código de forma remota en la máquina de la víctima.⁵⁹

⁵⁷ Diego Barbera, “Un hombre utilizó ChatGPT para ligar con cinco mil mujeres a la vez en Tinder. Logró casarse con una” Wired, Febrero 2024, disponible en <https://es.wired.com/articulos/hombre-utilizo-chatgpt-para-ligar-con-5-mil-mujeres-en-tinder-logro-casarse-con-una> (consultado el 21/06/2024)

⁵⁸ OpenAI, “ChatGPT plugins”, Blog OpenAI, Marzo de 2023, disponible en <https://openai.com/blog/chatgpt-plugins#OpenAI> (consultado el 30/04/2023).

⁵⁹ Matt Burgess, ChatGPT Has a Plug-In Problem, Wired, Julio 2023, disponible en <https://www.wired.com/story/chatgpt-plugins-security-privacy-risk/> (consultado el 21 de Junio de 2024).

De esta forma, la confianza de los usuarios se fue erosionando. En las nuevas funciones la seguridad no pudo ser garantizada y ello desestimula el uso del servicio más de lo que lo promociona. Por ello, OpenAI ha decidido, un año después, el 19 de marzo de 2024, que la creación de nuevas conversaciones con plugins quedaría deshabilitada. Si bien los usuarios podrían continuar sus conversaciones existentes ya no sería posible instalar nuevos complementos.

Instrucciones personalizadas. Para julio de 2023 los usuarios experimentados ya habían comprendido la importancia de brindar al modelo una instrucción completa, con el contexto suficiente para que la herramienta entienda cómo direccionar su respuesta. Sucedía que a veces ese contexto era tan amplio, tan complejo o tan ambiguo que muchos comenzaron a recurrir a aplicaciones de notas de texto con esta información ya redactada para copiar y pegar manualmente al escribir cada prompt. Para quien utiliza ChatGPT a diario, la tarea se volvió desgastante.

En respuesta a esto, OpenAI Lanzó “*Custom Instructions*” o instrucciones personalizadas⁶⁰, una función que permite añadir preferencias o requisitos que ChatGPT debería tener en cuenta al formular sus respuestas. Una vez introducida esta información, la misma es aplicada a todos los chats futuros.

Así, se volvió común informar si somos estudiantes, si trabajamos, a qué rubro nos dedicamos, para qué usamos el chat y a quiénes se dirige lo que estamos redactando para ajustar el estilo de respuesta. OpenAI presentó esta innovación como fruto de un proceso de retroalimentación realizado con usuarios de más de veintidós países⁶¹, cuyo resultado fue una comprensión más profunda del papel esencial que desempeña la direccionalidad para permitir que los modelos reflejen de manera efectiva los diversos contextos y las necesidades únicas de cada persona.

GPTs. Entramos a finales de 2023. El 6 de noviembre, OpenAI anunció una funcionalidad que hizo accesible la personalización y especialización de modelos para el usuario común. El lanzamiento de los llamados GPTs fue una especie de mezcla entre algunos beneficios de los plugins y *custom instructions*, permitiendo a los usuarios aprovecharlos desde la comodidad de su computadora personal y abonando únicamente los costos de una suscripción Plus a ChatGPT.

Cualquier persona o empresa puede crear fácilmente su propio GPT para uso propio o institucional. También puede ofrecerlo al público en general. Con esta innovación

⁶⁰ OpenAI, “*Custom instructions for ChatGPT*”, OpenAI Blog, 20 de julio de 2023, disponible en: <https://openai.com/blog/custom-instructions-for-chatgpt> (consultado el 02/05/2024).

⁶¹ OpenAI, “*ChatGPT plugins*”, Blog OpenAI, Marzo de 2023, disponible en <https://openai.com/blog/chatgpt-plugins#OpenAI> (consultado el 30/04/2023).

es posible customizar una versión para volverla idónea en la ejecución de ciertas tareas, o especialista en determinado tema. El usuario elige el propósito y lo construye en base a ello. Es tan fácil como iniciar una conversación, darle instrucciones y elegir qué puede hacer: buscar en la web, crear imágenes, analizar datos o generar texto.

Esta iniciativa ofrece la posibilidad de programación con lenguaje natural mediante una plataforma *no code* que da OpenAI llamada GPT Editor⁶². Desde entonces es posible crear una suerte de modelo propio sin programación, conocimientos técnicos ni gastos extras y a través del formato conversacional familiar para el usuario de a pie. El creador del GPT puede, también, cargar su propia base de datos en diversos formatos, incluso PDF, y así volver más específica a su versión. Esto es lo más cercano a la generación de software a medida que las personas ajenas al ámbito tecnológico han estado jamás. Poder hacerlo con tanta facilidad, en pocos pasos y eligiendo entre tantas tareas y objetivos es algo sin precedentes. Si bien los resultados no serán tan específicos como los obtenidos con un modelo que pasó por el proceso de *fine-tuning*, son suficientes para superar el entrenamiento genérico de ChatGPT y responder a demandas individuales.

De este modo, se amplía el grupo de personas que deciden "qué construir" más allá de aquellos con acceso a tecnología avanzada. Mediante el involucramiento de usuarios tan diversos, son las mismas personas quienes diseñan aquello que necesitan.

Para dimensionar lo anterior basta con analizar ejemplos de GPTs disponibles en la *GPTStore*⁶³, una tienda online exclusiva para estos modelos personalizados, los cuales cumplen funciones variadas que van desde lo más técnico hasta lo más cotidiano. Por ejemplo, hay GPTs de uso gratuito que generan resúmenes educativos escritos a partir de videos extensos en cualquier idioma⁶⁴ y otros que escriben recetas basadas en las preferencias del usuario y los ingredientes que tenga en su heladera⁶⁵, o que son capaces de crear y alojar sitios webs completos⁶⁶. A dos meses del lanzamiento de esta función, ya habían sido creados más de tres millones de modelos personalizados⁶⁷.

⁶² GPT Editor es el nombre que recibe la plataforma creada por OpenAI para que los usuarios desarrollen sus GPT personalizados en un formato amigable y conversacional similar al de ChatGPT. Más información disponible en OpenAI, *GPT Editor*, disponible en <https://help.openai.com/en/articles/8770868-gpt-builder> . Consultado el 27 de noviembre de 2024.

⁶³ Más información y enlace a la tienda en OpenAI, *Introducing the GPT Store*, disponible en <https://openai.com/index/introducing-the-gpt-store/> . Consultado el 27 de noviembre de 2024.

⁶⁴ *YouTubeΣ*, por Video Summarizer, disponible en <https://chat.openai.com/g/g-GvcYCKPIH-youtube> (consultado el 02/05/2024).

⁶⁵ *Sous Chef*, por ChatGPT, disponible en <https://chat.openai.com/g/g-3VrgJ1GpH-sous-chef> (consultado el 02/05/2024).

⁶⁶ *DesignerGPT*, por Pietro Schirano, disponible en <https://chat.openai.com/g/g-2Eo3NxuS7-designergpt> (consultado el 02/05/2024).

⁶⁷ OpenAI, "Introducing the GPT Store", OpenAI Blog, 10 de enero de 2024, disponible en: <https://openai.com/blog/introducing-the-gpt-store> (consultado el 02/05/2024).

Dentro de este océano de opciones, OpenAI ha resaltado que serán destacados los trabajos que sean de utilidad para las demás personas. Por eso los GPTs generados por los socios o la comunidad pueden, previo control -y esta es una de las principales diferencias respecto de los plugins-, estar disponibles en la tienda dentro de categorías como escritura, investigación, programación, educación y estilo de vida.

Además, al lanzar este nuevo espacio, desde la compañía han hablado de la monetización para quienes cuenten con un perfil de creador verificado, habilitando la posibilidad de publicar su elaboración y recibir comisiones por su uso⁶⁸. Esto último, al menos hasta el momento en que escribimos este documento, no ha recibido actualizaciones oficiales en la página de la empresa desde enero de 2024.

AI no requiere habilidades de programación, los GPTs amplían el grupo de personas que deciden "qué construir" más allá de aquellos con acceso a tecnología avanzada.

⁶⁸ OpenAI, "Introducing the GPT Store", OpenAI Blog, 10 de enero de 2024, disponible en: <https://openai.com/blog/introducing-the-gpt-store> (consultado el 02/05/2024).

VIII

Un nuevo ecosistema para
mejorar la productividad

VIII. Un nuevo ecosistema para mejorar la productividad

Para maximizar el potencial de ChatGPT en la transformación de la forma en que trabajamos y accedemos al conocimiento, OpenAI ha implementado mejoras a varios niveles, específicamente orientadas a las organizaciones y sus flujos de trabajo. Estas actualizaciones buscan establecer un nuevo ecosistema que optimice la productividad de manera cada vez más intuitiva, simplificada y automatizada. Veamos algunas de estas iniciativas.

ChatGPT Enterprise. Esta versión fue lanzada a finales de agosto de 2023. ChatGPT Enterprise ofrece a las empresas la versión más avanzada del modelo de lenguaje al mismo tiempo que se centra en mejorar la privacidad y la seguridad. Elimina los límites de uso, funciona más rápido, amplía la ventana de contexto, y lo más importante, promete que los datos o conversaciones comerciales de las organizaciones no se utilizarán para entrenar los modelos⁶⁹.

ChatGPT Teams. En enero de 2024 se presenta ChatGPT Teams, que cuesta 25 USD por persona, por mes⁷⁰. Cuenta con funcionalidades específicas como analizar datos, generar contenido y crear mejor código. Está pensado para todos los equipos de empresas de cualquier tamaño, permite crear y compartir con los otros miembros del espacio de trabajo y el administrador cuenta con una consola de control. En cuanto a la privacidad, se aplican los mismos recaudos que en Enterprise, y se incluyen los beneficios de la versión Plus.

Canvas. Recientemente, en octubre de 2024, OpenAI lanzó Canvas, una nueva interfaz para trabajar con ChatGPT que va más allá del simple chat⁷¹. Básicamente esta herramienta abre en una ventana separada, que permite interactuar con el modelo de lenguaje, no a través de la conversación, sino creando y refinando ideas a los dos lados de la pantalla. Está pensado para darle al usuario un colaborador creativo en proyectos que requieren edición y revisión, como escribir documentos o código informático.

De esta forma, ChatGPT detecta de forma automática situaciones en las que podría ser útil Canvas y abre la nueva ventana, o puede indicarlo el usuario en el prompt.

⁶⁹ La información sobre la versión Enterprise fue extraída de: OpenAI, “*Introducing ChatGPT Enterprise*” Página oficial de OpenAI, agosto 2023, disponible en: <https://openai.com/index/introducing-chatgpt-enterprise/> (consultado el 26 de julio de 2024).

⁷⁰ La información sobre la versión Teams fue extraída de: OpenAI, “*An always-improving superassistant for every member of your team*” Página Oficial de OpenAI, disponible en: <https://openai.com/chatgpt/team/> (consultado el 26 de julio de 2024).

⁷¹ La información sobre Canvas se encuentra disponible en: OpenAI, *Introducing Canvas*, 3 de octubre de 2024, disponible en: <https://openai.com/index/introducing-canvas/> (consultado el 29 de noviembre de 2024).

Teniendo en cuenta el contexto del trabajo, el modelo sugiere realizar modificaciones específicas o reescribir por completo. Asimismo, el usuario puede indicar secciones específicas para que ChatGPT se enfoque en ellas. Si se trata de escribir código, el sistema puede traducirlo a otros lenguajes de programación.



Para maximizar el potencial de ChatGPT en la transformación de la forma en que trabajamos y accedemos al conocimiento se han implementado actualizaciones que buscan establecer un nuevo ecosistema, optimizando la productividad de manera cada vez más intuitiva, simplificada y automatizada.

IX

Multimodalidad y nuevas formas de interacción



IX. Multimodalidad y nuevas formas de interacción

En los primeros meses de 2024 la multimodalidad se convirtió en protagonista. Esto implica que el modelo puede “razonar” tanto a través de texto como de imágenes, audio y videos procesando varias instrucciones respecto de ellas. De este modo, una IA que posee diversas capacidades para intentar “comprender” el mundo y manejar conceptos a través de distintas modalidades representa un avance hacia la creación de IA más general.

También, la multimodalidad abre la puerta a una nueva dimensión de interacción con los usuarios, donde las palabras se combinan con imágenes para enriquecer la comunicación y la generación de contenido en una variedad de contextos y en distintos formatos de salida. Según OpenAI, proporcionar imágenes resulta muy útil para contextualizar los temas discutidos con la herramienta⁷².

Los primeros avances de multimodalidad se publicaron en septiembre de 2023⁷³. Allí la empresa declaró que habilitaría a los usuarios Plus y Enterprise a interactuar a través de voz e imágenes. En cuanto a la primera modalidad, los usuarios podrían enviar audios al modelo. Por su parte, con la entrada de imagen, podrías pedirle que examine el contenido de tu heladera para organizar la receta de una comida o si te vas de vacaciones, mandarle la foto de un menú de un restaurante para que no solo lo traduzca, sino también te cuente el origen de cada plato, su historia y hasta te haga recomendaciones en base a tus preferencias.

Sora. En febrero de 2024 OpenAI presentó Sora⁷⁴, su nuevo modelo capaz de generar videos de hasta un minuto de duración manteniendo la calidad de la visualización a partir de comandos textuales. Actualmente, este modelo está disponible únicamente para investigadores dentro de la empresa, pero promete ser la próxima gran revolución. Según declaraciones de Mira Murati, jefa de tecnología, OpenAI planeaba lanzar Sora al público en general en "pocos meses". Sin embargo, llegando a finales de 2024, esto aún no ha ocurrido.

⁷²“ChatGPT ahora puede ver, oír y hablar”, OpenAI, publicado el 25/09/23. Información extraída de <https://openai.com/blog/chatgpt-can-now-see-hear-and-speak> (consultada el 20/02/24)

⁷³“ChatGPT ahora puede ver, oír y hablar”, OpenAI, publicado el 25/09/23. Información extraída de <https://openai.com/blog/chatgpt-can-now-see-hear-and-speak> (consultada el 20/02/24)

⁷⁴ OpenAI “Creating video from text” Página Oficial de OpenAI. Sora primeras impresiones, publicado el 15/02/24. Disponible en: <https://openai.com/index/sora/> (consultado el 23/06/24)

GPT-4o. El 19 de mayo de 2024 OpenAI lanzó GPT-4o, un modelo que se destaca por ser mucho más rápido y multimodal de forma nativa⁷⁵, es decir que esta capacidad está integrada desde el diseño y entrenamiento inicial del sistema.

Para interactuar por audio, el usuario cuenta con la posibilidad de seleccionar entre un número limitado de voces sintéticas que le permite entablar conversaciones con esta herramienta de manera similar a una llamada telefónica, recibiendo respuestas a sus preguntas en tiempo real. En la demostración de este modelo la empresa liderada por Sam Altman se enfrentó a un nuevo conflicto, esta vez con Scarlett Johansson.

Ocurrió que la actriz había rechazado la propuesta de OpenAI de proporcionar su voz para el nuevo asistente de IA⁷⁶. Sin embargo, la compañía siguió adelante con un sonido parecido y lo llamó “Sky”, lo que despertó el enojo de Johansson al sentir que habían usado su voz sin autorización. Ante la posibilidad de ser demandada, la organización creadora de ChatGPT emitió un comunicado explicando cómo se eligieron las voces, afirmando que no tuvieron intención de parecerse a la reconocida artista⁷⁷ y anunciando la suspensión del agente en cuestión. Para evitar problemas similares, retrasaron el lanzamiento de “Voice Engine”, una tecnología de clonación de voz que permite crear voces sintéticas a partir de instrucciones escritas basadas en un segmento de 15 segundos de audio grabado⁷⁸.

ChatGPT Search. Finalmente, uno de los lanzamientos más recientes de la empresa, involucra la posibilidad de buscar en internet⁷⁹, sea textos o imágenes. Gran parte de las preocupaciones que mencionamos al introducir Copilot fueron abordadas con esta funcionalidad. Ahora, ChatGPT puede darte las referencias de las respuestas que produce y decidir por sí mismo si buscar información en internet a partir de una pregunta o hacerlo a raíz de una petición expresa del usuario. También la empresa se asoció con ciertas organizaciones para agregar información actualizada y nuevos diseños visuales en

⁷⁵ Fernández Yúbal, “GPT-4o: qué es y en qué se diferencia con GPT-4”, Xataka, mayo de 2024, disponible en: <https://www.xataka.com/basics/gpt-4o-que-que-se-diferencia-gpt-4> (consultado el 21 de junio de 2024).

⁷⁶ Fung Brian, “¿Por qué OpenAI debería temer una demanda de Scarlett Johansson?” CNN, mayo 2024, disponible en: <https://cnnespanol.cnn.com/2024/05/22/openai-demanda-scarlett-johansson-chatgpt-trax/> (consultado el 26 de julio de 2024).

⁷⁷ OpenAI, “Navigating the Challenges and Opportunities of Synthetic Voices” Página Oficial de OpenAI, Marzo 2024, disponible en: <https://openai.com/index/navigating-the-challenges-and-opportunities-of-synthetic-voices/> (consultado el 26 de julio de 2024).

⁷⁸ Edwards Benj, “OpenAI retrasa el lanzamiento de Voice Engine, su tecnología de clonación de voz, por temor a usos indebidos” The Wired, abril 2024, disponible en: <https://es.wired.com/articulos/openai-retrasa-lanzamiento-de-voice-engine-tecnologia-de-clonacion-de-voz-por-temor-a-usos-indebidios> (consultado el 26 de julio de 2024).

⁷⁹ Sobre ChatGPT Search ver: OpenAI, *Introducing ChatGPT Search*, 31 de octubre 2024, disponible en: <https://openai.com/index/introducing-chatgpt-search/> (consultado el 29 de noviembre de 2024).

categorías como clima, deportes, noticias y mapas. En el corto plazo estará disponible para los usuarios pagos, mientras que para los gratuitos se habilitará en los próximos meses.



La multimodalidad abre la puerta a una nueva dimensión de interacción con los usuarios, donde las palabras se combinan con imágenes, videos y audio para enriquecer la comunicación y la generación de contenido.



X

Agentes de IA y el camino hacia la AGI
(inteligencia artificial general)

X. Agentes de IA y el camino hacia la AGI (inteligencia artificial general)

Demanda de Elon Musk a OpenAI. El 29 de febrero de 2024 Elon Musk demandó a OpenAI⁸⁰. Entre las cuestiones que plantea el magnate sudafricano en su escrito, afirma que la empresa se apartó de su misión original de desarrollar IA en beneficio de la humanidad. Sin embargo, lo más importante se relaciona con la alianza entre ésta y Microsoft y la supuesta creación de AGI.

Aunque poco se sabe de los términos del acuerdo entre los dos gigantes tecnológicos, se ha dado a conocer que se excluye automáticamente AGI de cualquier licencia de propiedad intelectual o cesiones de uso comercial acordadas con Microsoft. Es decir, si bien Microsoft tiene licencia exclusiva sobre los productos que vaya lanzando OpenAI, se supone que, cuando la empresa dirigida por Sam Altman logre finalmente desarrollar un sistema tan capaz e inteligente para ser considerado AGI, este estará fuera del alcance de Microsoft.

Lo que ocurre según Musk es que este nivel de inteligencia ya se ha logrado. Sostiene ante el Tribunal Superior de California que GPT-4 es una AGI y al ser explotada por Microsoft comercialmente, se han violado los términos del acuerdo. Según declara, solo desea que la AGI desarrollada sea beneficiosa para toda la humanidad, no solo rentable para Altman y Microsoft⁸¹.

Ahora, parece ser que se encuentra en manos de unos pocos jueces determinar si nuestra sociedad ha alcanzado AGI o no. Esto supone un problema porque, como punto de partida, no hay consenso en el propio campo de la IA sobre lo que AGI significa. En la demanda, Elon dice que GPT-4 por su capacidad de razonamiento y habilidad para realizar tareas de manera competente en una amplia gama de dominios, es una forma temprana, aunque incompleta de AGI. Se trata de una IA de propósito general, que posee inteligencia para muchas actividades, al igual que el ser humano.

Volviendo a la demanda, el punto a favor de Elon Musk es que los propios expertos de Microsoft habrían admitido que GPT-4 constituye AGI. En una investigación publicada

⁸⁰ Es importante tener en cuenta que cuando OpenAI nació, en 2015, Elon Musk fue uno de sus cofundadores junto con Sam Altman y Greg Brockman. A principios del año 2018 el Sr. Elon Musk, decidió dar un paso al costado de la Junta Directiva de la organización para, según sus afirmaciones, evitar cualquier tipo de conflicto con los intereses de sus propias empresas. A pesar de ello, Musk siguió durante mucho tiempo asesorando y participando económicamente de OpenAI, realizando inversiones de una gran cantidad de dinero en ella.

⁸¹ Hernández Gonzalo, "Elon Musk demanda a OpenAI por una razón tan absurda que, en vez de ser sonar justa, parece rencor" Xataka, marzo 2024, disponible en: <https://www.xataka.com.mx/robotica-e-ia/elon-musk-demanda-a-openai-razon-absurda-que-vez-ser-sonar-justa-parece-rencor> (consultado el 27 de junio de 2024).

por un equipo de investigación de la empresa, se afirma que las pruebas del modelo demuestran una especie de chispas de AGI⁸² y se define a este tipo de IA como aquellos sistemas que exhiben amplias capacidades de inteligencia, incluido el razonamiento, la planificación y la capacidad para aprender de la experiencia, y que alcanzan estas capacidades al mismo nivel que un humano o superior. Sin embargo, aclara que ello no implica que el sistema sea perfecto, que pueda hacer cualquier cosa que un humano puede hacer, y que tenga motivaciones y objetivos internos.

Por su parte, si bien OpenAI efectuó un descargo a través de su página oficial respecto de la demanda en su contra, no emitió ninguna declaración en relación con la AGI. Además, recientemente ha publicado su propia teoría sobre el camino hacia la “Super AI”. Se trata de 5 niveles o pasos que nos acercan progresivamente a la creación de una inteligencia de nivel humano. Sus modelos actuales se encuentran en el primer escalón (conversacional) al que le siguen 2-razonadores; 3-agentes; 4-innovadores y 5-organizacionales⁸³.

“Modelo frutilla” o 01 preview de razonamiento avanzado. En septiembre de 2024, se presentó OpenAI o1, un modelo diseñado para mejorar las capacidades de razonamiento y resolución de problemas complejos⁸⁴. Conocido internamente como “Strawberry”, OpenAI afirma que este sistema es capaz de aprender a reconocer y corregir sus errores y probar un enfoque diferente cuando el actual no funciona. Se trata de uno de los mejores esfuerzos para lograr verdaderas IA “razonadoras” que ocupan el segundo lugar en los 5 pasos hacia la superinteligencia.

Sabemos que nuestra especie naturalmente necesita más tiempo para reflexionar cuanto más difícil sea una tarea. Así, el modelo o1 retrasa la generación de respuestas unos segundos para simular ciertos patrones del pensamiento humano. Mientras tanto, recurre a las *Chain-of-thoughts* (cadenas de pensamiento) para mejorar la precisión y coherencia de sus salidas. La famosa técnica de *prompting*, conocida por hacer que los modelos de lenguaje lleguen a mejores resultados, ahora está embebida en esta nueva IA y resulta su principal estrategia, para, según la empresa, superar a cualquier estudiante de

⁸² Bubeck Sébastien, Chandrasekaran Varun, Eldan Ronen, Gehrke Johannes, Horvitz Eric, Kamar Ece, Lee Peter, Lee Yin Tat, Li Yuanzhi, Lundberg Scott, Nori Harsha, Palangi Hamid, Tulio Ribeiro Marco and Zhang Yi, “*Sparks of Artificial General Intelligence: Early experiments with GPT-4*” Microsoft Research, Abril 2023, disponible en: <https://arxiv.org/pdf/2303.12712.pdf> (consultado el 27 de junio de 2024).

⁸³ Lagos Anna, “*OpenAI traza el mapa hacia la superinteligencia artificial*” The Wired, julio 2024, disponible en: <https://es.wired.com/articulos/openai-traza-el-mapa-hacia-la-super-inteligencia-artificial> (consultado el 25 de julio de 2024).

⁸⁴ OpenAI, *Introducing OpenAI o1-preview*, 12 de septiembre de 2024, disponible en: <https://openai.com/index/introducing-openai-o1-preview/> (consultado el 29 de noviembre de 2024).

doctorado en problemas de física, química y biología. La clave del éxito radica en descomponer los problemas complejos en pasos más simples⁸⁵.

Por ahora, se han presentado dos versiones: o1-preview, que es la principal con “capacidades completas de razonamiento” y o1-mini, más ligera y económica, diseñada para aplicaciones que requieren razonamiento profundo, pero con menor carga computacional, siendo un 80% más barata que o1-preview. Asimismo, el modelo frutilla está disponible para usuarios de ChatGPT Plus y Team, con planes de expandirse al resto en el futuro.

Agentes de IA. Los agentes de IA constituyen el tercer nivel en la escala de OpenAI hacia la superinteligencia. A pesar de ello, ya hay muchísimos proyectos pilotos que los han desarrollado. De hecho, se predice que para el año que viene, el 25% de las empresas que utilizan IA generativa lanzarán pruebas de concepto de IA basada en agentes, cifra que aumentará al 50% en 2027⁸⁶. Esto sin duda, debido al potencial que tiene esta tecnología de mejorar exponencialmente la productividad de los trabajadores y automatizar procesos completos.

Los agentes de IA se caracterizan por la posibilidad de interactuar con su entorno, procesar información y tomar decisiones en tiempo real. Mientras que los creadores humanos establecen sus objetivos en términos generales, los agentes eligen de forma autónoma qué tareas ejecutar para alcanzarlos⁸⁷. Se basan en modelos fundacionales como los de lenguaje, pero agregan técnicas que les permiten actuar de forma independiente, autónoma y sin supervisión humana.

Cada innovación que hemos visto hasta aquí, contribuyó a allanar el camino para que esta tecnología fuera posible. Desde las búsquedas en internet, la posibilidad de personalizar los modelos, la mejora en el razonamiento y las capacidades multimodales. Incluso los esfuerzos para ampliar la ventana de contexto: una característica distintiva de la IA basada en agentes es que puede acceder a la memoria de corto plazo para mantener

⁸⁵ OpenAI, “*Learning to Reason with LLMs*” 12 de septiembre de 2024, disponible en: <https://openai.com/index/learning-to-reason-with-langs/> (consultado el 29 de noviembre de 2024).

⁸⁶ Jeff Loucks, Gillian Crossan, Baris Sarer, China Widener, “*Autonomous generative AI agents: Under development*”, 19 de noviembre 2024, disponible en: <https://www2.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2025/autonomous-generative-ai-agents-still-under-development.html> (consultado el 29 de noviembre de 2024).

⁸⁷ AWS “*¿Qué son los agentes de IA?*”, disponible en: <https://aws.amazon.com/what-is/ai-agents/#:~:text=AI%20agents%20are%20autonomous%20intelligent,repetitive%20tasks%20to%20AI%20agents.> (Consultado el 29 de noviembre de 2024).

el la información contextual mientras realiza una tarea específica y a la memoria de largo plazo para aprender y mejorar a partir de la experiencia⁸⁸.

Algunos de los casos de uso más interesantes van desde escribir código, detectar y gestionar ciberataques, jugar videojuegos y hasta organizar una agenda completa. De hecho, los agentes pueden comunicarse con otros y así formar una colaboración multiagente.

Los agentes de IA constituyen el tercer nivel en la escala de OpenAI hacia la superinteligencia. Ya existen muchísimos proyectos pilotos de esta tecnología que no solo procesa información, sino que es capaz de tomar decisiones de forma autónoma

⁸⁸ Jeff Loucks, Gillian Crossan, Baris Sarer, China Widener, “Autonomous generative AI agents: Under development”, 19 de noviembre 2024, disponible en: <https://www2.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2025/autonomous-generative-ai-agents-still-under-development.html> (consultado el 29 de noviembre de 2024).

Línea de tiempo

2022

30 DE NOVIEMBRE • **Lanzamiento ChatGPT**

2023

1 DE FEBRERO • **Lanzamiento ChatGPT Plus**7 DE FEBRERO • **Lanzamiento de Copilot**1 DE MARZO • **Lanzamiento de la API**14 DE MARZO • **Lanzamiento GPT-4**22 DE MARZO • **Publicación de la primera carta abierta para pausar el desarrollo de la IA**23 DE MARZO • **Lanzamiento Plugins**30 DE MAYO • **Publicación de la segunda carta abierta advirtiendo los riesgos de la IA**20 DE JULIO • **Lanzamiento Custom Instructions o instrucciones personalizadas**28 DE AGOSTO • **Lanzamiento ChatGPT Enterprise**25 DE SEPTIEMBRE • **Ahora ChatGPT puede observar, escuchar y hablar: primeros avances en materia de multimodalidad**6 DE NOVIEMBRE • **Lanzamiento GPTs y GPT-4 Turbo**17 DE NOVIEMBRE • **OpenAI despide a Sam Altman**

2024

10 DE ENERO • **Lanzamiento ChatGPT Teams**15 DE FEBRERO • **Anuncio de Sora**29 DE FEBRERO • **Elon Musk demanda a OpenAI**13 DE MAYO • **Lanzamiento GPT-4o**12 DE JULIO • **Publicación de los niveles para alcanzar la superinteligencia de OpenAI**12 DE SEPTIEMBRE • **Lanzamiento de OpenAI o1-preview o “modelo frutilla”**3 DE OCTUBRE • **Lanzamiento de Canvas**31 DE OCTUBRE • **Lanzamiento de ChaGPT Search**

.UBAderecho



IALAB FAIR