

Las 'Big Tech' recelan ahora de la IA de código abierto que impulsaron

Por: Will Douglas Heaven. 28/06/2023

Un mayor acceso al código de los modelos generativos de lenguaje está impulsando la innovación. Pero si las grandes empresas se asustan, podrían cerrar sus puertas

Hace unas semanas [se filtró un memorándum](#) redactado por **Luke Sernau**, ingeniero jefe de Google. En este se decía claramente lo que muchos otros llevaban semanas susurrando en Silicon Valley: una batalla campal sobre el código abierto está amenazando el control de la IA por parte de las grandes empresas tecnológicas.

Los grandes modelos lingüísticos de código abierto, que son alternativas a [Bard de Google](#) u [OpenAI de ChatGPT](#), están cayendo como caramelos de una piñata para que los investigadores y desarrolladores de *apps* los puedan estudiar, desarrollar y modificar. Se trata de **versiones más pequeñas y baratas** de los mejores modelos de IA creados por grandes empresas, que (casi) los igualan en rendimiento, y se comparten de forma gratuita.

Google [reveló en su presentación anual](#) que lanzará IA generativa a todos sus productos, desde Gmail hasta Fotos y Maps. Según Sernau, las empresas estaban demasiado ocupadas mirando por encima del hombro como para ver venir la verdadera competencia: **“Mientras discutíamos, una tercera facción se ha comido nuestra merienda en silencio”**.

En muchos sentidos, esto es bueno. Un mayor acceso a estos modelos ha contribuido a impulsar la innovación, y también puede ayudar a detectar sus defectos. **La IA no prosperará si unas pocas empresas ricas controlan la tecnología** o deciden cómo se utiliza.

Sin embargo, este auge del código abierto es algo precario. La mayoría de lanzamientos de código abierto se apoyan en grandes modelos creados por empresas enormes con mucho capital. Si OpenAI y Meta deciden cerrar, esta enorme ola podría convertirse en agua estancada.

Por ejemplo, muchos de estos modelos se basan en **LLaMA**, un gran modelo lingüístico de código abierto que ha sido publicado por **Meta AI**. Otros utilizan un conjunto de datos públicos llamado **Pile**, creado por EleutherAI, una organización de código abierto sin ánimo de lucro. Aunque [\[CCB1\]](#) EleutherAI solo existe porque la apertura de OpenAI permitió a un grupo de programadores aplicar ingeniería inversa a Chat [GPT-3](#), y crear su propia tecnología durante su tiempo libre.

“Meta AI ha hecho un estupendo trabajo formando y liberando modelos para la comunidad investigadora”, afirma Stella Biderman, directora ejecutiva y responsable de Investigación en EleutherAI, y científica en la consultora **Booz Allen Hamilton**. Sernau también destaca el papel crucial de Meta AI en su informe de Google. La empresa confirmó a *MIT Technology Review* que el memorándum fue escrito por uno de sus empleados, pero señala que no es un documento estratégico oficial.

Todo eso podría cambiar. OpenAI ya está [dando marcha atrás a su anterior política de apertura](#) por miedo a la competencia. Y Meta podría frenar debido **al riesgo de que los advenedizos hagan cosas desagradables** con su código abierto. “Es lo que hay que hacer ahora mismo”, asegura Joelle Pineau, directora general de Meta AI, sobre la apertura del código a personas ajenas a la empresa. “¿Adoptaremos esta misma estrategia durante los próximos cinco años? No lo sé, la IA avanza muy deprisa”.

Si la tendencia a cerrar el acceso continúa, no solo se abandonará a muchas personas interesadas en el código abierto, sino que **la próxima generación de avances en IA volverá a estar totalmente en manos de los laboratorios de IA más grandes y ricos del mundo**.

Por tanto, el futuro de la creación y el uso de la inteligencia artificial se encuentra en una encrucijada.

Bonanza de código abierto

El *software* de código abierto existe desde hace décadas, de hecho, **es la base de internet**. Pero el coste de crear modelos potentes hizo que la IA de código abierto no despegara hasta hace un año. Y se ha convertido rápidamente en un filón.

Basta con echar un vistazo a las últimas semanas. El 25 de marzo, **Hugging Face**,

una *startup* que defiende el acceso libre y abierto a la IA, presentó la primera alternativa de código abierto a ChatGPT, el chatbot que OpenAI lanzó en noviembre.

HuggingChat, el chatbot de Hugging Face, se basa en **Open Assistant**, un gran modelo lingüístico de código abierto y perfeccionado para la conversación. Este fue entrenado con la ayuda de unos **13.000 voluntarios** y publicado hace un mes. Pero el propio Open Assistant se basa en el LLaMA de Meta.

Luego está **StableLM**, un gran modelo lingüístico de código abierto publicado el 19 de marzo por **Stability AI**, que está detrás de [Stable Diffusion](#), el exitoso modelo que traduce texto a imagen. Una semana después, el 28 de marzo, Stability AI lanzó **StableVicuna**, una versión de StableLM que, como Open Assistant o HuggingChat, está optimizada para la conversación. Así, StableLM sería la respuesta de Stability a GPT-4 y StableVicuna una respuesta a ChatGPT.

Estos nuevos modelos de código abierto se unen ya a los publicados en los últimos meses, como [Alpaca](#) (de un equipo de la Universidad de Stanford), [Dolly](#) (de Databricks, empresa de *software*) y [Cerebras-GPT](#) (de la empresa **Cerebras**). La mayoría de estos modelos se basan en LLaMA, o en conjuntos de datos y modelos de EleutherAI; Cerebras-GPT, por ejemplo, sigue una plantilla establecida por **DeepMind**. Y se puede apostar a que vendrán muchas más.

Para algunos, el código abierto es una cuestión de principios. “Colocar el poder de la IA conversacional al alcance de todos es un **esfuerzo comunitario global** para que no esté en manos de unas pocas grandes corporaciones”, afirma **Yannic Kilcher**, investigador de IA y YouTuber, en un [vídeo de presentación de Open Assistant](#).

“**Nunca abandonaremos la lucha por la IA de código abierto**”, [tuiteó](#) el mes pasado Julien Chaumond, cofundador de Hugging Face.

Para otros, es una cuestión de beneficios. Stability AI espera repetir con los chatbots el mismo truco que utilizó con las imágenes: alimentar y luego beneficiarse de un estallido de innovación entre los desarrolladores que utilizan sus productos. La empresa planea aprovechar lo mejor de esa innovación y convertirlo en productos personalizados para una amplia gama de clientes. “Avivamos la [llama de la] innovación y luego elegimos. **Es el mejor modelo de negocio del mundo**”, afirma Emad Mostaque, CEO de Stability AI.

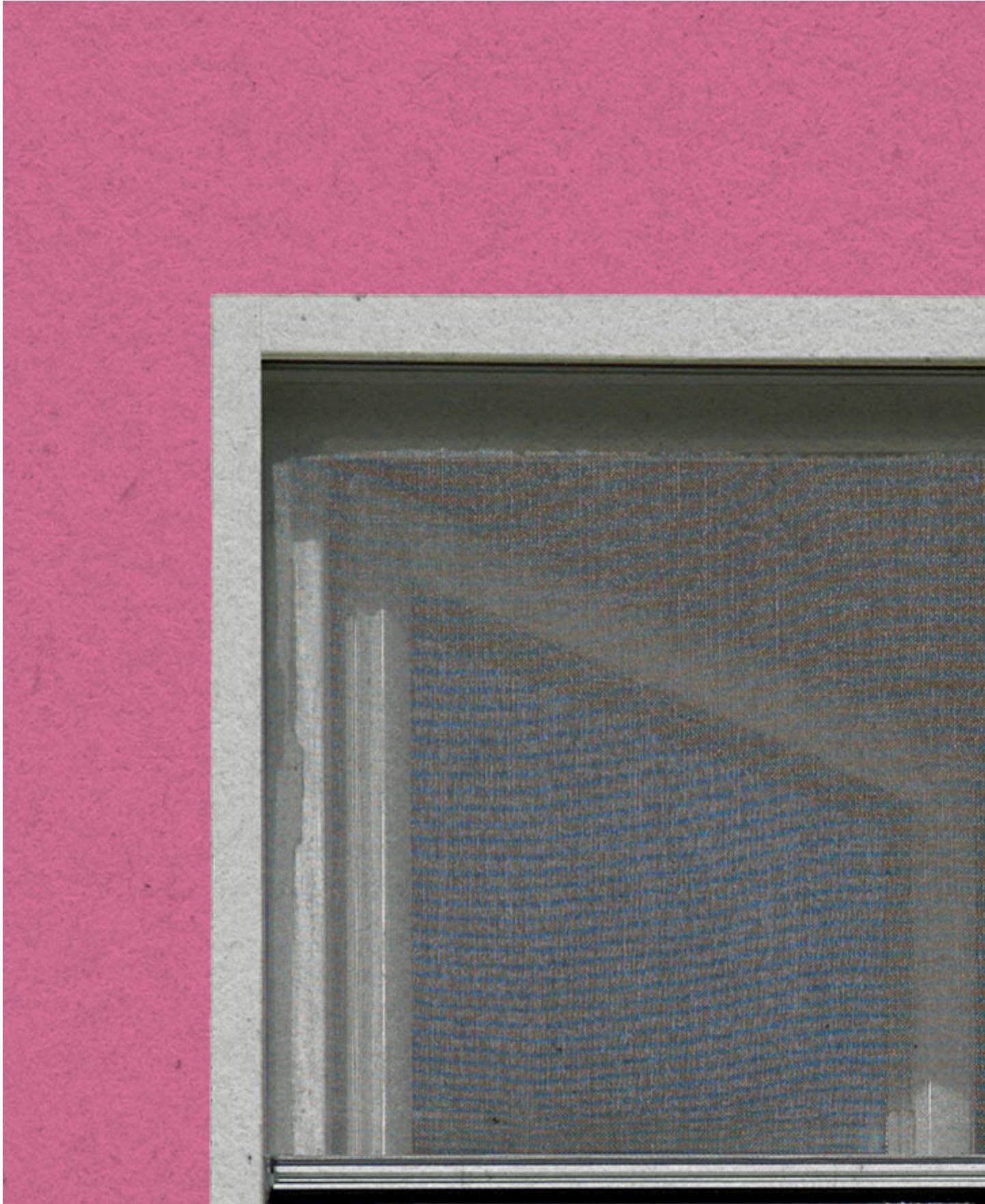
En cualquier caso, la abundancia de grandes modelos de lenguaje gratuitos y de código abierto pone esta tecnología en manos de millones de personas de todo el mundo, lo que inspira a muchos a crear nuevas herramientas y explorar su funcionamiento. “**El acceso a esta tecnología es mucho mayor que antes**”, afirma Biderman.

“El increíble número de formas en que la gente ha estado utilizando esta tecnología es alucinante”, dice Amir Ghavi, abogado de la firma Fried Frank que representa a varias empresas de IA generativa, incluida Stability AI. “Es una muestra de la creatividad humana, que es de lo que trata el código abierto”.

GPUs que se derritieron

No obstante, entrenar grandes modelos lingüísticos desde cero, en lugar de basarse en ellos o modificarlos, es difícil. “Todavía está fuera del alcance de la mayoría de personas”, afirma Mostaque. “**Fundimos un montón de GPUs en la construcción de StableLM**”.

La primera versión de Stability AI, [el modelo de conversión de texto a imagen](#) de Stable Diffusion, **funcionaba tan bien (si no mejor) que sus homólogos de código cerrado**, como Imagen de Google y [DALL-E](#) de OpenAI. No solo era gratuito, sino que además funcionaba en un buen ordenador doméstico. Stable Diffusion hizo más que ningún otro modelo para desencadenar el auge del año pasado sobre el desarrollo de código abierto en la IA de generación de imágenes.



Sin embargo, esta vez Mostaque quiere gestionar las expectativas: StableLM no está ni de lejos a la altura de GPT-4. “Todavía queda mucho trabajo por hacer. No es como Stable Diffusion, donde inmediatamente tienes algo con mucha usabilidad. Los modelos lingüísticos son más difíciles de entrenar”.

Otra dificultad es que **los modelos son más difíciles de entrenar cuanto más grandes son**. Esto no solo se debe al coste de la potencia de cálculo. El proceso de entrenamiento se interrumpe con mayor frecuencia cuando se trata de modelos más grandes y hay que reiniciarlo, lo que encarece aún más la construcción de esos modelos.

En la práctica, el número de parámetros que la mayoría de los grupos puede permitirse entrenar tiene un límite, asegura Biderman. Esto se debe a que los modelos de gran tamaño deben entrenarse en varias GPUs diferentes y el cableado de todo ese *hardware* es complejo. “Entrenar con éxito modelos a esa escala es **un campo muy nuevo de la investigación en computación de alto rendimiento**”, afirma Biderman.

A medida la tecnología avanza, el número exacto cambia, pero Biderman sitúa ese techo entre **6.000 y 10.000 millones de parámetros** ahora. Como comparación, **GPT-3 tiene 175.000 millones de parámetros y LLaMA, 65.000**. No es una correlación exacta, pero los modelos más grandes suelen funcionar mejor en general.

Biderman espera que la actividad continúe en torno a los grandes modelos lingüísticos de código abierto, pero se centrará más en ampliar o adaptar algunos modelos ya formados y existentes que en impulsar la tecnología fundamental. “**Solo hay un puñado de organizaciones que hayan preformado estos modelos**, y creo que seguirá siendo así en un futuro próximo”, afirma.

Por eso, muchos modelos de código abierto se basan en LLaMA, entrenado desde cero por Meta AI, o en versiones de **EleutherAI**, una organización sin ánimo de lucro única en su contribución a la tecnología de código abierto. Biderman afirma que solo conoce otro grupo similar, y se encuentra en China.

EleutherAI comenzó su andadura gracias a OpenAI. En 2020, la empresa de San Francisco acababa de lanzar un nuevo modelo. “GPT-3 supuso un gran cambio para

mucha gente en su manera de pensar sobre la IA a gran escala”, afirma Biderman. “A menudo, se le atribuye el mérito de un cambio de paradigma intelectual, en términos de lo que la gente espera de estos modelos”.

Entusiasmados por el potencial de esta nueva tecnología, Biderman y un puñado de investigadores más quisieron **jugar con el modelo para comprender mejor** cómo funcionaba. Y entonces, decidieron reproducirlo.

OpenAI no había publicado aún el GPT-3, pero compartió suficiente información sobre cómo se había construido para que Biderman y sus colegas pudieran descifrarlo. Nadie fuera de OpenAI había entrenado antes un modelo así, pero se encontraban en plena pandemia y el equipo tampoco tenía mucho más que hacer. “Estaba trabajando y jugando a juegos de mesa con mi mujer cuando me involucré”, dice Biderman. “Así que **fue fácil dedicarle 10 o 20 horas a la semana**”.

Su primer paso fue reunir un nuevo conjunto de datos masivo, con miles de millones de piezas de texto para competir con lo que OpenAI había utilizado para entrenar a GPT-3. EleutherAI llamó a su conjunto de datos Pile y, a finales de 2020, **lo publicó gratuitamente**.

Entonces, EleutherAI utilizó este conjunto de datos para entrenar su primer modelo de código abierto. EleutherAI tardó tres meses y medio en entrenar su modelo más grande, y fue patrocinado por una empresa de computación en nube. “Si lo hubiéramos pagado de nuestro bolsillo, nos habría costado unos **400.000 dólares** (unos 370.000 euros). Es mucho pedir para un grupo de investigación universitario”.

Una mano amiga

Debido a estos costes, es más fácil construir sobre modelos existentes. LLaMA, de Meta AI, se ha convertido en el punto de partida de muchos nuevos proyectos de código abierto. Meta AI se ha inclinado por el desarrollo de código abierto, desde que [Yann LeCun](#) la fundó hace una década. Esa mentalidad forma parte de su cultura empresarial, afirma Pineau: “Es un enfoque propio del mercado libre, tipo: **muévete rápido, construye cosas**”.

Pineau tiene claras las ventajas. “**Diversifica el número de personas que pueden contribuir al desarrollo de la tecnología**. No solo los investigadores o los empresarios, sino también los gobiernos civiles y otros pueden tener visibilidad en

estos modelos”.

Al igual que la comunidad del código abierto, Pineau y sus colegas creen que **la transparencia debe ser la norma**. “Insto mucho a mis investigadores a que empiecen un proyecto pensando que quieren que sea de código abierto. Porque, cuando lo haces, pones un listón más alto sobre los datos que utilizas y cómo construyes el modelo”.

Sin embargo, también existen graves riesgos. Los grandes modelos de lenguaje arrojan **desinformación, prejuicios y discursos de odio**. Pueden utilizarse para producir propaganda en masa, o alimentar fábricas de *malware*. “Hay que elegir entre transparencia y seguridad”, afirma Pineau.

En el caso de Meta AI, esa compensación puede significar que algunos modelos no lleguen a publicarse. Por ejemplo, si el equipo de Pineau ha entrenado un modelo con datos de usuarios de Facebook, se quedará en casa, ya que el riesgo de filtración de información privada es demasiado grande. De lo contrario, **el equipo podría publicar el modelo con una licencia que especifique que debe utilizarse solo con fines de investigación**.

Este es el enfoque que se adoptó para LLaMA. Sin embargo, a los pocos días de su publicación, **alguien colgó el modelo completo y las instrucciones para ejecutarlo en el foro 4chan**. “Sigo pensando que era la solución adecuada para este modelo en concreto”, afirma Pineau. “Pero me decepciona que la gente haga esto, porque hace más difícil realizar estos lanzamientos”.

“Siempre hemos contado con el **firme apoyo** a este enfoque por parte de los directivos de la empresa, desde **Mark [Zuckerberg]**, pero no es fácil”, afirma Pineau.

Meta AI se juega mucho. “La responsabilidad potencial de cometer una locura es menor cuando se trata de una pequeña start-up y no cuando se trata de una gran compañía”, asegura Pineau. “En este momento, liberamos estos modelos a miles de individuos, pero, si se vuelve más problemático o sentimos que los riesgos de seguridad son mayores, cerraremos el círculo. Y solo lo liberaremos para socios académicos con credenciales sólidas, bajo acuerdos de confidencialidad o NDA que **les impidan construir cualquier cosa con el modelo, incluso con fines de investigación**”.

Si eso ocurre, muchos de los favoritos del ecosistema de código abierto podrían ver **revocada su licencia** para construir sobre lo que Meta AI saque a continuación. **Sin LLaMA, modelos de código abierto como Alpaca, Open Assistant o Hugging Chat no serían tan buenos.** Y la próxima generación de innovadores de código abierto no tendrá la ventaja que ha tenido la generación actual.

En la balanza

Otras personas también están sopesando los riesgos y beneficios de esta batalla campal sobre el código abierto.

Casi a la vez que Meta AI lanzó LLaMA, Hugging Face puso en marcha un **mecanismo de bloqueo** para que la gente tenga que solicitar acceso, y ser aprobado, antes de descargar muchos de los modelos que se encuentran en la plataforma de la empresa. La idea es restringir el acceso a las personas que tienen una **razón legítima**, según lo ha determinado Hugging Face, para poner sus manos sobre el modelo.

“No soy una evangelista del código abierto”, afirma Margaret Mitchell, jefa de Ética de Hugging Face. “Pero sí veo razones por las que cerrarlo tiene sentido”.

Mitchell señala la **pornografía no consentida** como un ejemplo de los inconvenientes de hacer accesibles modelos más potentes. Es uno de los principales usos de la IA para crear imágenes, según afirma.

Mitchell, que anteriormente trabajó en Google y fue cofundadora de su equipo Ethical AI, entiende las tensiones que entran en juego. Y es partidaria de lo que denomina “democratización responsable”, un planteamiento similar al de Meta AI, según el cual los modelos se publican de forma controlada en función de su riesgo potencial de causar daños o ser mal utilizados. “Aprecio mucho los ideales del código abierto, pero creo que es útil contar con algún tipo de **mecanismo de rendición de cuentas**”.

OpenAI también está cerrando el grifo. El mes pasado, cuando anunció GPT-4, la nueva versión del modelo de lenguaje que impulsa ChatGPT, hubo una frase llamativa en el informe técnico: “Dado el panorama competitivo como las implicaciones para la seguridad de modelos a gran escala como GPT-4, **este informe no contiene más detalles sobre la arquitectura**”.

(incluido el tamaño del modelo), el *hardware*, el cálculo de entrenamiento, **la construcción del conjunto de datos, el método de entrenamiento** o similares”.

En parte, estas nuevas restricciones se deben a que OpenAI ahora es una empresa con ánimo de lucro, que compite con empresas como Google. Pero también reflejan un cambio de actitud. Ilya Sutskever, cofundador y director científico, ha declarado en una [entrevista con The Verge](#) que la apertura de su empresa en el pasado fue un error.

En opinión de Sandhini Agarwal, investigadora de Políticas de OpenAI, “antes, si algo era de código abierto, tal vez le importara a un pequeño grupo de aficionados. Ahora, todo el entorno ha cambiado. El código abierto puede acelerar mucho el desarrollo, y provocar un hundimiento”.

Sin embargo, no siempre ha sido así. Si OpenAI se hubiera sentido así hace tres años, cuando publicó detalles sobre GPT-3, no existiría EleutherAI, por ejemplo.

En la actualidad, EleutherAI desempeña un papel fundamental en el ecosistema de código abierto. Desde entonces, ha creado varios modelos lingüísticos de gran tamaño y Pile se ha utilizado para entrenar numerosos proyectos de código abierto, como StableLM de Stability AI. Además, Mostaque forma parte del consejo de EleutherAI.

Nada de esto habría sido posible si OpenAI hubiera compartido menos información. Al igual que Meta AI, EleutherAI permite una gran innovación de código abierto.

No obstante, con GPT-4 (5 y 6) bloqueados, **el ecosistema del código abierto podría quedar de nuevo a merced de unas pocas grandes empresas.** Podrían producirse nuevas versiones salvajes, e incluso amenazar a algunos de los productos de Google. Pero se quedarán con los modelos de la última generación. Los verdaderos avances, los próximos pasos hacia delante, se producirán a puerta cerrada.

¿Esto importa? Lo que uno piense sobre el impacto del bloqueo por parte de las grandes empresas tecnológicas y el impacto que tendrá en el código abierto, depende de lo que uno piense sobre **cómo debe crearse la IA y quién debe hacerlo**.

“Puede que la IA impulse cómo se organizará la sociedad en las próximas décadas”, afirma Ghavi. **“Tener un sistema más amplio de controles y transparencia es mejor que concentrar el poder en las manos de unos pocos”**.

Biderman está de acuerdo: “No creo que exista una **necesidad moral** de que todo el mundo trabaje con código abierto. Pero, al fin y al cabo, es muy importante contar con personas que desarrollen e investiguen esta tecnología, sin intereses económicos en su éxito comercial”.

Sin embargo, Biderman tiene claro lo que está en juego. “Soy menos crítica con OpenAI de lo que mucha gente cree. Liberar públicamente esta tecnología conlleva riesgos reales, y potenciales peligros”.

OpenAI afirma que solo está jugando sobre seguro. “No pensamos que la transparencia sea mala”, dice Dave Willner, jefe de los equipos de Confianza y Seguridad de OpenAI. “Más bien intentamos averiguar cómo conciliar la transparencia con la seguridad. Y, a medida que estas tecnologías se hacen más potentes, en la práctica hay cierta tensión entre esos temas”.

“Muchas normas y formas de pensar en la IA se han formado en comunidades de investigación académicas, que valoran **la colaboración y la transparencia** para que la gente pueda basarse en el trabajo de los demás”, asegura Willner. “Quizá eso tenga que cambiar, a medida que se desarrolle esta tecnología”.

[**LEER EL ARTÍCULO ORIGINAL PULSANDO AQUÍ**](#)

Fotografía: Technology review

Fecha de creación

2023/06/28