



Hoja de datos

Adobe Firefly: ética y sesgos



Si se implementa correctamente, la inteligencia artificial (IA) generativa abre nuevas posibilidades y potencia los flujos de trabajo existentes, además de ampliar la creatividad. Sin embargo, como ocurre con la mayoría de las tecnologías innovadoras, la promesa de la IA generativa conlleva desafíos y riesgos que deben abordarse, desde la infracción de los derechos de autoría hasta los impactos dañinos y los sesgos. Basándonos en nuestra trayectoria de más de una década de innovación en IA en todas nuestras tecnologías en la nube, hemos desarrollado e implementado la IA generativa de Adobe Firefly de acuerdo con nuestros principios éticos de la IA de rendición de cuentas, responsabilidad y transparencia.

Adobe se compromete a desarrollar la IA generativa creativa de manera responsable, para que puedas tener confianza en el contenido que generas.

Firefly es seguro para uso comercial

Los modelos de IA generativa patentados de Firefly son seguros para uso comercial, lo que significa que cada cliente puede publicar los resultados con la confianza de que Adobe ha abordado de manera responsable los problemas de licencias y derechos de autoría. Hemos entrenado los modelos de la IA generativa de Adobe Firefly con una gran cantidad de imágenes, audio y vídeo de contenido con licencia, como Adobe Stock, y contenido de dominio público cuyos derechos de autoría han caducado.



Indemnización por propiedad intelectual

Parte de la clientela tendrá la oportunidad de comprar un derecho que incluye una indemnización contractual por propiedad intelectual en el improbable caso de una reclamación que involucre un resultado de Firefly. Esto significa que Adobe asumirá la responsabilidad de la defensa legal y los perjuicios económicos por reclamaciones que aleguen que un resultado de Firefly infringe o viola directamente cualquier patente, derechos de autoría, marca comercial, derechos de publicidad o derechos de privacidad de una empresa externa. La responsabilidad de Adobe por las reclamaciones de resultados de Firefly coincidirá con las mismas limitaciones de las reclamaciones de propiedad intelectual basadas en la tecnología en virtud de los contratos existentes del cliente con Adobe. La indemnización de Adobe no cubrirá las reclamaciones que se basen en o surjan de la acción o el contenido del cliente, tales como cualquier modificación del resultado de Firefly, el contexto en el que se utiliza el resultado de Firefly, el contenido proporcionado para el entrenamiento a medida o cualquier uso de un resultado de Firefly que infrinja el acuerdo con el cliente.

[Más información sobre la seguridad comercial y la indemnización de Firefly](#)






Seguro para las empresas

Antes del lanzamiento de las funciones de IA, nuestros equipos de pruebas siguen un proceso detallado para mitigar los resultados dañinos y adversos, priorizando el lenguaje que necesita ser clasificado y filtrado, creando listas de bloqueo y denegación de indicaciones y reduciendo los casos de contenido inseguro para el trabajo (NSFW).

[Más información sobre la moderación de contenido de Firefly](#)

Responsabilidad

En Adobe, asumimos la responsabilidad de los resultados de nuestras herramientas asistidas por IA y contamos con procesos estandarizados para mantener nuestros compromisos de ética de la IA, tales como:

-  **Establecer** procedimientos de gobernanza para supervisar los datos de entrenamiento y los modelos de IA
-  **Formar** a los equipos de ingeniería y a los responsables de producto sobre cuestiones de ética de la IA
-  **Crear** una junta de revisión de ética de la IA para supervisar la promulgación del desarrollo de la IA y ofrecer un foro de debate para las cuestiones éticas relacionadas con la IA
-  **Exigir** una evaluación del impacto de la IA durante nuestro proceso de desarrollo de servicios para ayudar a garantizar que se realice una revisión de la ética de la IA antes de implementar nuevas tecnologías de IA
-  **Desarrollar** procesos para garantizar la recepción de comentarios y la corrección de cualquier impacto negativo de la IA después de la implementación

Con procesos dedicados a recibir y reaccionar a las preocupaciones sobre nuestra IA tanto interna como externamente, garantizamos que se tomen medidas correctivas cuando sea apropiado. Además de nuestra junta interna de revisión de ética de la IA, Firefly tiene un mecanismo de comentarios integrado para que quienes lo usan puedan informar si una función produce un resultado que se percibe como, por ejemplo, sesgado o inexacto. Este ciclo de comentarios con nuestra comunidad es una forma importante de ayudar a garantizar que nuestras herramientas minimicen los daños y sigan siendo responsables.

[Más información sobre el compromiso de Adobe con la rendición de cuentas](#)

Responsabilidad

En Adobe valoramos mucho asumir la responsabilidad del impacto de nuestra empresa y de la innovación que entregamos al mundo. Creemos que colocar mecanismos de protección bien pensados en torno a nuestro desarrollo y uso de la IA ayudará a aprovechar todo el potencial de la IA para beneficiar a la sociedad y, al mismo tiempo, mitigar los posibles impactos perjudiciales. Nuestro enfoque para el desarrollo responsable de la IA abarca:



Diseñar un sistema de IA de forma meditada



Evaluar cómo interactúa con las personas que la usan



Realizar un análisis exhaustivo para mitigar los sesgos perjudiciales no deseados



Evaluar el impacto humano de la tecnología de IA

Mitigar los daños y los sesgos

La IA es tan buena como los datos con los que se entrena, y reconocemos que todos los datos tienen sesgos. El desarrollo responsable de la IA requiere anticipar posibles sesgos y los daños resultantes, tomar medidas preventivas para mitigar dichas cuestiones, medir y documentar la prevalencia de los daños y establecer sistemas para supervisar y reaccionar ante resultados perjudiciales imprevistos. Aunque diseñamos para la inclusión, evaluamos continuamente el impacto humano de nuestros sistemas de IA para detectar resultados potencialmente injustos, discriminatorios o inexactos que pudieran propagar sesgos o estereotipos dañinos relacionados con atributos humanos (por ejemplo, raza, género, ingresos, religión, discapacidad, etc.). Para ayudar a garantizar que se cumplan nuestros compromisos de responsabilidad y ética, también hemos implementado recursos dedicados y procesos estandarizados para evaluar continuamente el impacto de la IA generativa de Adobe Firefly, tales como:

- Comité de ética de la IA: trabaja con nuestros equipos de desarrollo para defender nuestros principios de ética de la IA en cada etapa del desarrollo y para mitigar de forma proactiva el posible impacto o sesgo perjudicial
- Junta de revisión de ética de la IA: revisa rigurosamente todas las funciones y productos de Adobe impulsados por IA antes de su lanzamiento público

Transparencia

También creemos que la transparencia genera confianza, y nos tomamos muy en serio la confianza que nuestra clientela, partners y la comunidad del sector depositan en nuestros productos. Nos comprometemos a extender esa transparencia a los resultados generados por la IA, aplicando las credenciales de contenido a todas las imágenes generadas por Adobe Firefly. Como parte de un ecosistema creciente de tecnologías disponibles a través de la [iniciativa de autenticidad del contenido \(Content Authenticity Initiative\)](#), de la que Adobe es miembro fundador, las credenciales de contenido son un nuevo tipo de metadatos a prueba de manipulaciones que ayudan a generar credibilidad y confianza, al tiempo que evitan la difusión de información errónea en línea. Esto no solo ayuda a garantizar que el equipo creativo obtenga reconocimiento por su trabajo, sino que también indica cuándo se utiliza la IA generativa en una imagen para que otras personas que usan o consumen el producto estén informadas de cómo se hizo.

[Más información sobre el principio de transparencia de Adobe](#)

Ponte en contacto con un representante para saber cómo Adobe se compromete con el desarrollo responsable de la tecnología de IA a medida que continuamos innovando, aprendiendo y creciendo.