



Folha de dados

Adobe Firefly: ética e viés



Quando a inteligência artificial (IA) generativa é implementada de maneira correta, ela oferece novas possibilidades e eficiência aos fluxos de trabalho existentes, além de impulsionar a criatividade. No entanto, como acontece com a maioria das tecnologias inovadoras, a promessa da IA generativa vem acompanhada de desafios e riscos que precisam ser abordados, de violações de direitos autorais a viés e impactos prejudiciais. Com base em nosso legado de mais de uma década de inovação em IA em nossas tecnologias de nuvem, desenvolvemos e implantamos a IA generativa do Firefly de acordo com nossos [princípios de ética da IA](#): prestação de contas, responsabilidade e transparência.

Para que você possa gerar conteúdo criativo com total confiança, a Adobe faz um trabalho de desenvolvimento meticuloso da IA generativa.

O Firefly é seguro para uso comercial

Os modelos de IA generativa de propriedade do Firefly são seguros para uso comercial, o que significa que os clientes podem publicar resultados com confiança, sabendo que a Adobe tratou de modo responsável questões relativas a licenciamento e direitos autorais. Treinamos os modelos de IA generativa do Firefly com um grande número de imagens, áudio e vídeo provenientes de conteúdo licenciado, como o Adobe Stock, e conteúdo de domínio público cujos direitos autorais já expiraram.



Indenização por violação de propriedade intelectual

Alguns clientes podem adquirir direitos que incluem indenização por violação de propriedade intelectual sob contrato no caso improvável de alegações relativas a resultados gerados pelo Firefly. Isso significa que a Adobe assume a responsabilidade pela defesa jurídica e pelos danos monetários em caso de alegações de que um resultado gerado pelo Firefly infringe ou viola diretamente a patente, direitos autorais, marca registrada, direitos de publicidade ou direitos de privacidade de terceiros. A responsabilidade da Adobe em relação a alegações referentes a resultados gerados pelo Firefly está sujeita às mesmas limitações de responsabilidade das alegações de violação de propriedade intelectual baseadas em tecnologia previstas nos contratos que os clientes já possuem com a Adobe. A indenização da Adobe não cobre alegações baseadas ou decorrentes da ação ou do conteúdo do cliente, incluindo qualquer modificação do resultado gerado pelo Firefly, o contexto no qual o resultado gerado pelo Firefly é usado, o conteúdo fornecido para treinamento personalizado ou qualquer uso de resultados gerados pelo Firefly que viole o contrato do cliente.

[Saiba mais sobre as questões de segurança comercial e indenização do Firefly](#)








Seguro para as empresas

Antes do lançamento dos recursos de IA, nossas equipes de teste seguem um processo detalhado para atenuar resultados prejudiciais e inadequados, priorizando a linguagem que precisa ser classificada e filtrada, criando listas de bloqueios de prompts e reduzindo ocorrências de conteúdo impróprio para o trabalho (NSFW).

[Saiba mais sobre a moderação de conteúdo do Firefly](#)

Prestação de contas

Na Adobe, assumimos a responsabilidade pelos resultados de nossas ferramentas assistidas por IA e temos processos padronizados para manter nossos compromissos de ética da IA, incluindo:

-  **Estabelecer** procedimentos de governança para rastrear dados de treinamento e modelos de IA.
-  **Instruir** engenheiros e gerentes de produto sobre questões relacionadas à ética da IA.
-  **Criar** um conselho de revisão da ética da IA para supervisionar a promulgação do desenvolvimento da IA e oferecer um espaço de consulta para questões sobre a ética da IA.
-  **Exigir** uma avaliação de impacto da IA durante o processo de desenvolvimento de nossos serviços para ajudar a garantir que uma análise da ética da IA seja realizada antes da implantação de novas tecnologias de IA.
-  **Desenvolver** processos para garantir o feedback e a correção de quaisquer impactos negativos da IA detectados após a implantação.

Nossos processos dedicados a receber e responder às preocupações sobre nossa IA, tanto interna quanto externamente, garantem a tomada de medidas corretivas apropriadas. Além do nosso conselho interno de revisão da ética da IA, o Firefly conta com um mecanismo de feedback integrado para que os usuários possam informar se algum recurso produz resultados percebidos como, por exemplo, tendenciosos ou imprecisos. Esse ciclo de feedback com nossa comunidade de usuários é muito importante para ajudar a garantir que nossas ferramentas minimizem os danos e continuem permitindo um uso responsável.

[Saiba mais sobre o compromisso da Adobe com a prestação de contas](#)

Responsabilidade

Assumir a responsabilidade pelo impacto da nossa empresa e pela inovação que oferecemos ao mundo é um aspecto inestimável da Adobe. Ao definir medidas de proteção cuidadosas para o desenvolvimento e o uso da IA, acreditamos ser possível maximizar seus benefícios para a sociedade e minimizar eventuais impactos prejudiciais. A abordagem que usamos para o desenvolvimento responsável da IA inclui:



Desenvolver meticulosamente um sistema de IA.



Avaliar como ela interage com os usuários finais.



Exercer a devida diligência para mitigar preconceitos intoleráveis e prejudiciais.



Avaliar o impacto humano da tecnologia de IA.

Mitigação de danos e preconceitos

A qualidade dos resultados da IA depende diretamente dos dados usados para treiná-la, e reconhecemos que todos os dados são tendenciosos. Desenvolver IA de modo responsável exige antecipar os possíveis vieses e danos resultantes, bem como adotar medidas preventivas para atenuar tais problemas, avaliar e documentar a ocorrência de danos e estabelecer sistemas para monitorar e responder a resultados prejudiciais imprevistos. Apesar de ter sempre em mente a inclusão ao desenvolver nossos sistemas de IA, avaliamos continuamente o impacto humano deles para identificar possíveis resultados injustos, discriminatórios ou imprecisos que possam propagar vieses ou estereótipos prejudiciais relacionados a atributos humanos (por exemplo, raça, gênero, renda, religião, deficiência etc.). Visando garantir o cumprimento de nossos compromissos éticos e de responsabilidade, também implementamos recursos dedicados e processos padronizados para avaliar continuamente o impacto da IA generativa do Firefly, incluindo:

- Comitê de ética da IA: trabalha com as equipes de desenvolvimento para garantir o cumprimento dos nossos princípios de ética da IA em todos os estágios de desenvolvimento e mitigar de modo proativo impactos ou vieses potencialmente prejudiciais.
- Conselho de revisão da ética da IA: revisa rigorosamente todos os recursos e produtos da Adobe viabilizados por IA antes do lançamento público.

Transparência

Também acreditamos que a transparência gera confiança, e levamos a sério a confiança que nossos clientes, parceiros e colegas do setor depositam em nossos produtos. Portanto, faz parte do nosso compromisso estender essa transparência aos resultados gerados por IA, por meio da aplicação de Content Credentials a todas as imagens geradas pelo Adobe Firefly. As Content Credentials correspondem a um novo tipo de metadados invioláveis que ajuda a criar credibilidade e confiança, evitando a disseminação de desinformação online. Elas fazem parte de um ecossistema crescente de tecnologias disponíveis por meio da [Content Authenticity Initiative \(CAI\)](#), da qual a Adobe é membro fundador. Além de ajudar a garantir que criadores sejam reconhecidos por seu trabalho, também indica quando a IA generativa é usada em uma imagem para que outros usuários ou consumidores estejam informados sobre como ela foi feita.

[Saiba mais sobre o princípio de transparência da Adobe.](#)

Entre em contato com um representante para saber mais sobre o compromisso da Adobe em desenvolver tecnologia de IA de modo responsável à medida que continuamos a inovar, aprender e crescer.