

The Adobe logo is displayed in white text on a red background.

Datasheet

# Adobe Firefly: Ethics and Bias



Implemented the right way, generative artificial intelligence (AI) brings new possibilities and power to existing workflows and amplifies creativity. However, as with most boundary-pushing technologies, the promise of generative AI comes with challenges and risks that need to be addressed—from copyright infringement to harmful impacts and bias. Building on our decade-plus legacy of AI innovation across our cloud technologies, we've developed and deployed Firefly generative AI in alignment with our [AI Ethics Principles](#) of accountability, responsibility, and transparency.

**Adobe is committed to developing creative generative AI thoughtfully, so you can have confidence in the content you're generating.**

## Firefly is Commercially Safe

Firefly's proprietary generative AI models are safe for commercial use, meaning customers can confidently publish outputs knowing Adobe has responsibly addressed licensing and copyright issues. We trained Firefly generative AI models on a vast number of images, audio, and video from licensed content, such as Adobe Stock, and public domain content where copyright has expired.



## IP indemnification

Some customers will have the opportunity to purchase an entitlement that comes with contractual IP indemnification in the unlikely event of a claim involving a Firefly output. This means that Adobe will assume responsibility for legal defense and monetary damages for claims that allege a Firefly output directly infringes or violates any third party's patent, copyright, trademark, publicity rights, or privacy rights. Adobe's liability for Firefly output claims will match the same limitations on technology-based IP claims under customers' existing contracts with Adobe. The Adobe indemnity will not cover claims that are based on or arise from the customer's action or content, including any modification of the Firefly output, the context in which the Firefly output is used, content provided for custom training, or any use of a Firefly output that violates the customer agreement.

[Learn more about Firefly's commercial safety and indemnification](#)







## Safe for business

Prior to AI features being released, our testing teams follow a detailed process for mitigating harmful and adverse outputs, prioritizing language that needs to be classified and filtered, creating prompt block-and-deny lists, and reducing instances of Not Safe for Work (NSFW) content.

[Learn more about Firefly's content moderation](#)

## Accountability

At Adobe, we take ownership over the outcomes of our AI-assisted tools and have standardized processes in place to uphold our AI ethics commitments, including:

-  **Establishing** governance procedures to track training data and AI models
-  **Educating** engineers and product managers on AI ethics issues
-  **Creating** an AI Ethics Review Board to oversee the promulgation of AI development and offer a sounding board for AI ethics concerns
-  **Requiring** an AI Impact Assessment during our services development process to help ensure an AI ethics review occurs before we deploy new AI technologies
-  **Developing** processes to ensure feedback on and remediation of any negative AI impacts after deployment

With processes dedicated to receiving and responding to concerns about our AI both internally and externally, we ensure that corrective action is taken when appropriate. In addition to our internal AI Ethics review board, Firefly has a built-in feedback mechanism so that users can report if a feature produces a result that is perceived as, for example, biased or inaccurate. This feedback loop with our user community is an important way to help ensure our tools minimize harm and remain accountable.

[Learn more about Adobe's commitment to accountability](#)

## Responsibility

At Adobe, we place a high value on taking responsibility for the impact of our company and the innovation we deliver to the world. We believe that placing thoughtful guardrails around our AI development and use will help realize the full potential of AI to benefit society while mitigating potentially harmful impacts. Our approach to responsible AI development encompasses:



**Designing** an AI system thoughtfully



**Evaluating** how it interacts with end users



**Exercising** due diligence to mitigate unwanted harmful bias



**Assessing** the human impact of AI technology

## Mitigating harm and bias

AI is only as good as the data on which it is trained, and we recognize that all data has bias. Responsible development of AI requires anticipating potential biases and resulting harms, taking preemptive steps to mitigate such concerns, measuring and documenting the prevalence of harm, and establishing systems to monitor and respond to unanticipated harmful outcomes. Even though we design for inclusiveness, we continuously assess the human impact of our AI systems for potentially unfair, discriminatory, or inaccurate results that could propagate harmful biases or stereotypes related to human attributes (e.g., race, gender, income, religion, disability, etc.). To help ensure our responsibility and ethics commitments are being met, we've also implemented dedicated resources and standardized processes to continuously assess the impact of Firefly generative AI, including:

- AI Ethics Committee – Works with our development teams to uphold our AI Ethics Principles at every stage of development and to proactively mitigate potentially harmful impact or bias
- AI Ethics Review Board – Rigorously reviews all Adobe AI-powered features and products before public release

## Transparency

We also believe that transparency builds trust, and we take the trust that our customers, partners, and industry peers put in our products seriously. We are committed to extending that transparency to AI generated outputs, applying Content Credentials to all images generated by Adobe Firefly. Part of a growing ecosystem of technologies available through the [Content Authenticity Initiative \(CAI\)](#), of which Adobe is a founding member, Content Credentials are a new kind of tamper-evident metadata that helps build credibility and trust while preventing the spread of online misinformation. This not only helps ensure creators get recognition for their work but indicates when generative AI is used in an image so other users or consumers are informed how it was made.

[Learn more about Adobe's transparency principle.](#)

Please reach out to a rep to learn how Adobe is committed to the responsible development of AI technology as we continue to innovate, learn and grow.