

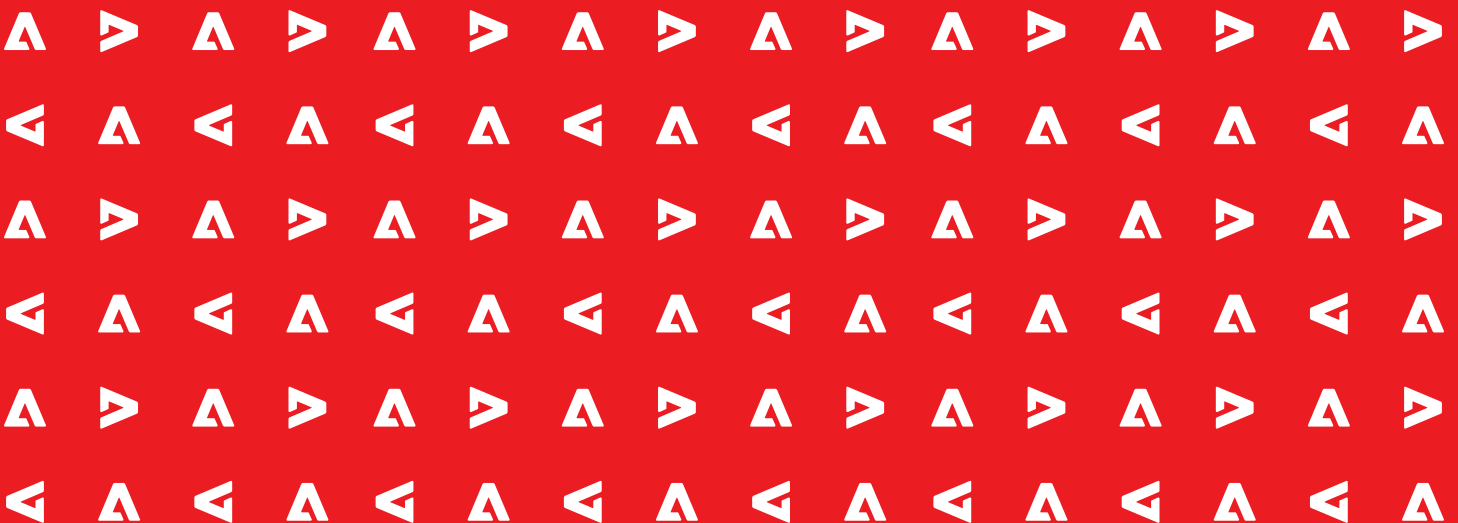


SOLUTION BRIEF

# GenAI Built for Business

Adobe's Commitment to Building AI Responsibly

April 2024



# Overview

Generative AI (GenAI) is a transformational technology that is fundamentally changing the way we all think about creativity. From individual creators to global enterprises, the opportunities for innovation using GenAI are immense. Implemented the right way, GenAI brings precision, power, speed, and ease to existing workflows — allowing people to focus on more strategic or creative parts of their work.

However, as with most boundary-pushing technologies, the promise of Generative AI comes with challenges and risks that need to be addressed. AI systems are based on data, and that data can mirror existing biases, include unsafe content, or potentially implicate other individuals' or corporations' intellectual property. Without appropriate guardrails, AI systems can unintentionally discriminate, disparage, offend, or otherwise cause harm.

Building and using GenAI systems responsibly requires a commitment to making the systems accountable, responsible, and transparent as well as to complying with emerging AI regulations and laws around the globe. Consumers and users of GenAI have a similar responsibility to use the technology in ways that minimize bias and harm.

Adobe is committed to developing and deploying AI systems thoughtfully, building responsibly on our foundation of AI Ethics principles.

## What is Generative AI?

Artificial intelligence (AI) uses a machine learning model, large datasets, and pattern recognition to be able to produce a specific type of result, such as personalized recommendations, voice recognition, language translations, and much more. Generative AI is a category of artificial intelligence that creates something new, such as text, images, videos, and digital experiences. For example, in the case of image generation, the AI model is trained on hundreds of millions of images so it can learn patterns and produce new images based on its dataset.

## Adobe's Approach to Responsible AI

At Adobe, Generative AI is the next step in the decade-plus legacy of AI innovation across our cloud technologies. What began with Adobe Sensei has now evolved into the powerful GenAI capabilities built into existing Adobe product offerings that amplify creativity and intelligence without replacing human imagination.

# Our AI Ethics Principles

All Adobe' AI features, whether using Adobe's own proprietary models or third-party partners' models, go through Adobe's AI Ethics governance process and are developed and deployed in alignment with our [AI Ethics Principles](#) of accountability, responsibility, and transparency:

- **Accountability** means that we take ownership over the outcomes of our AI-assisted tools. We have processes and resources dedicated to receiving and responding to concerns about our AI and taking corrective action as appropriate.
- **Responsibility** means that we approach designing and maintaining our AI technology with thoughtful evaluation and careful consideration of the impact of its deployment.
- **Transparency** means that we are open about and explain our use of AI to our customers, so they have a clear understanding of our AI systems and their application.

Guided by these three principles, Adobe has created standardized processes for our AI features and solutions, including our first-party models, from design and development through deployment. These processes include continuous data collection, training, and testing as well as a defined community feedback process.

## Adobe AI Ethics Governance

Adobe has dedicated resources throughout the company to meet our AI Ethics commitments.

- **AI Ethics Committee** – Works with our development teams to understand and incorporate our AI Ethics Principles
- **AI Ethics Review Board** – Reviews Adobe AI-powered features and products before release
- **AI@Adobe Working Group** – Helps navigate the exploration and adoption of generative AI internally within Adobe

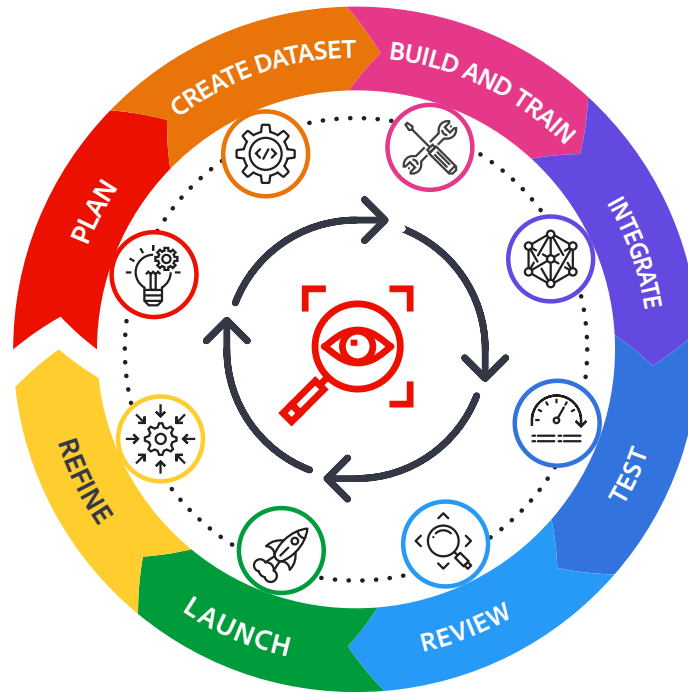
# How Adobe Builds Ethical AI Solutions

## Phase 1: Plan

At the start of development of a new AI-powered feature or technology begins, the product team submits an AI Ethics Assessment. The multi-part impact assessment asks questions that identify both the business problem the feature is intended to address, as well as pinpoint the potential for unintended outputs or perpetuate harmful biases and stereotypes. This assessment allows our AI Ethics Review Board to focus its efforts on features and products with the highest potential impact, without slowing down the pace of innovation.

If the Board determines that the answers to the initial assessment present a low risk, the product team may move forward with development. If, however, the Board determines that the AI feature points to a high risk, the product team works with the AI Ethics team throughout development to mitigate that risk. For first-party models, this may require adding data to the dataset, building additional filters, changing the algorithm, or retraining the model. For third-party models, we may build additional processes and filters in the input or output layers to help mitigate harm and bias. This process repeats until the AI feature or product meets Adobe's AI Ethics principles.

## AI Development Lifecycle



## Phase 2: Create Dataset

After the planning phase, the product team turns to determining the dataset required to solve their defined business problem. The provenance of the data Adobe inputs to train our models is important to adhering to our AI Ethics principles.

Adobe currently offers AI solutions that employ proprietary models designed to be safe for commercial use. Customers can use those commercially safe solutions to confidently publish outputs knowing Adobe has responsibly addressed licensing and copyright issues. For these models, we create image datasets based on the vast number of images, audio, and video in Adobe Stock along with licensed work and public domain content where copyright has expired.

Why does this matter? Because if an individual or a company uses an image that has been generated by an AI model trained on copyrighted or otherwise legally protected content, that individual or company can be subject to reputational damage and expensive lawsuits for the use of that protected content.

Adobe also uses third-party models in some of its products. While we have layered additional protection around those models, we provide options for less restrictive creative ideation to help meet specific customer requirements. We are also transparent about any use of third-party models so customers can make informed decisions.

### **What if the dataset isn't quite right?**

During the build and train phase of AI development, training is not a "one and done" process. The model typically must be "tuned," that is, trained on either a larger, more diverse dataset or on a completely different dataset than originally expected. For example, in the initial training of Adobe Firefly, the model produced hands with more than five fingers. This resulted in the product team realizing that they needed more images of hands in the dataset to improve the model.

For image models, such as Firefly, the product teams enlist the help of the Adobe Stock team to acquire additional images to expand the dataset. Adobe Stock only purchases or acquires images for which Adobe has monetarily compensated the creator, secured the consent of person photographed, or confirmed that the copyright license has expired.

## **Phase 3: Build and Train**

At this point, an iterative process begins. The product team trains the AI model on the initial dataset and evaluates the output. If the output is lacking or in any way does not meet the team's expectations, they add to the dataset. This process continues until the model produces the expected or desired output. For third party models, the process of integration follows the same iterative process.

## Phase 4: Create or Integrate into Application

A model itself doesn't have much value without an application (or surface) through which a user can access it. Further, a robust application typically involves multiple models to achieve the desired results, each of which have a specific purpose or function, not just a single model.

## Phase 5: Test

Adobe teams conduct testing to reduce the potential for biased and harmful outcomes in our GenAI products. Using a detailed process for mitigating offensive and adverse outputs, the teams focus on testing for unintended consequences, intentional abuse of the system, harmful content generation, and bias and stereotype amplification. These testing teams are instrumental in 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.

Because they are integrated into our existing product offerings, Adobe Security tests the GenAI features at the same time as they test the solution into which the feature is integrated. Our rigorous security testing process includes internal and outsourced penetration tests as well as extensive testing from an adversarial perspective. In addition, our security team works closely with industry researchers and ethical hackers to increase the amount of testing our AI features undergo. Documented product vulnerabilities are then sent to the product team for mitigation and the final results are published in product-specific Security Testing Reports.

## Phase 6: Review

Before AI-powered features are released to beta testing, our AI Ethics Review Board, which includes employees with diverse professional, gender, and ethnic backgrounds, may conduct a final review to help ensure the product team has worked to minimize bias and harm. The board's diversity helps to identify potential issues that a non-diverse team might not see—before launch and deployment.

## Phase 7: Launch and Use

Typically, Adobe AI-powered features and technologies are first released to a private enterprise beta program to gain important customer feedback, which is then integrated into the product before public release. These customers are valuable participants in our development lifecycle. After the enterprise private beta, the feature or product is released in a public beta testing scenario and finally launched as a standard feature of the product.

Once in customers' hands, Adobe GenAI powered by Firefly creates images that are safe for commercial use, because the Adobe Stock images on which it is trained are vetted and humanly verified to be copyright compliant. But more importantly, Adobe automatically applies Content Credentials to all content generated with Firefly, enabling creators to receive appropriate recognition for their work and consumers to verify the authenticity of the content.

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 as well as prevents the spread of online misinformation and disinformation.

With Content Credentials, creators can add information about themselves and their creative process directly to their content at export or download, so they can receive appropriate recognition for their work—and consumers can know who created the image and when and how it was created.

## Phase 8: Feedback and Refine

Because Adobe is committed to remediating negative AI impacts that emerge after deployment, we welcome and encourage feedback on our AI-powered features and technologies in many ways. Firefly has a built-in feedback mechanism so that users can report if a feature produces a result they perceive as, for example, biased or inaccurate. In our view, this feedback loop with our user community is one important way to help ensure our tools minimize harm and uphold Adobe's AI Ethics principles.

# Transparency Builds Trust

At Adobe, our AI Ethics principles guide every step of the AI development process, even before actual coding begins. Strict adherence to our AI governance process means that every product team approaches the design and development of AI with careful, thoughtful, and through consideration of the impact of its deployment.

We design for inclusiveness, testing and remediating potentially discriminatory or inaccurate results that might perpetuate harmful biases and stereotypes during design as well as once in customers' hands. Dedicated resources respond to customer concerns about Adobe AI and take corrective action as appropriate.

Finally, Adobe believes that transparency builds trust, and we take the trust our customers, partners, and industry peers put in our products seriously. We are transparent about our data collection, training, and testing practices, including how we test for and resolve issues related to harmful bias. And we are committed to extending that transparency to customer outputs, applying Content Credentials to all images generated by Adobe Firefly, making the provenance of the image easily discoverable so that users know that they can trust the authenticity of the content.

