stability.ai

Implications of AI for the Creative Process

Stability AI develops AI technology to help unlock humanity's potential. Our goal is to make foundational AI technology accessible to all, including through open research and development. AI will drive a wave of creativity, innovation, and productivity, and we are working to put these tools in the hands of workers and firms across the United States.

Background. Stability AI develops a variety of generative AI models. These models are software programs that can produce new content such as text, images, video, or audio. We are committed to releasing these models openly to promote access for all. Open releases allow developers to freely use, integrate, or adapt the underlying model to build new and innovative applications. Users can then access these models through a hosted application.

- **Stable Diffusion.** In 2022, Stability AI collaborated with research partners to release an open model known as Stable Diffusion. Stable Diffusion is a type of "latent diffusion" model that takes a text prompt from a user and "translates" that prompt into a new image, subject to our ethical use license.¹ By some measures, developer interest in Stable Diffusion has grown faster than any open software project in recent history.
- StableLM. In 2023, Stability AI launched the first in a suite of open language models known as StableLM. These models take a text prompt from a user and produce new text or software code. StableLM demonstrates how small AI models can deliver useful performance with appropriate training: StableLM has delivered surprisingly high performance in conversational and coding tasks, even though the first StableLM release was limited to three billion and seven billion parameter models – significantly smaller than models like GPT-3 at 175 billion parameters.



Above left: Image generated from the prompt "photograph of an astronaut riding a pink horse". Above right: Text generated with a fine-tuned version of StableLM.

¹ The <u>Open Responsible AI License</u> prohibits unlawful, exploitative, or misleading use of Stable Diffusion.

stability.ai

Why it matters. Al is a tool, not a substitute for creators. Instead, Al can help to simplify the creative process for artists, writers, and developers. Al can help existing creators boost their productivity, stimulate new ideas, and experiment with new concepts. In addition, Al can lower barriers to entry for people who do not have the resources or training to realize their creative potential. As with other assistive technologies – from paintbrushes to typewriters to Photoshop – the user ultimately determines the content and usage of any generated material.

How it works. These models analyze existing content to learn the relationship between words, ideas, and visual or textual features. With that acquired understanding, and creative direction from the user, these models can help to generate new content.

For example, an image model can review pairs of text captions and images to understand the appearance of "fur" on a dog, "ripples" on water, moods like "bleak", or styles like "Renaissance". Likewise, a language model can review existing text to learn the typical arrangement of words in a "blog", "email", or "speech"; understand the features that contribute to tones like "funny" or "formal"; and learn to distinguish, summarize, expand, or combine different concepts – from "banana" to "Wall Street" to "print()".

It is important to properly characterize the training process. These models are not a collage machine or search engine for existing content. They use knowledge learned from analyzing content to help a user generate new works. They do not use the original works themselves, and the original works are not stored in the AI model.



These models do not "stitch together" original works. Above: Image models notoriously struggle with hands. In this case, the model produces a nine-fingered hand because it hasn't understood that a hand typically has five fingers, and it isn't searching a database of existing images with hands. Instead, it has learned that a "hand" means a flesh-coloured artefact accompanied by some number of appendages.

stability.ai

Working with creators. Training these models is an acceptable and transformative use of content that is protected by fair use. However, we acknowledge the depth of feeling on these issues among creators, and we support efforts to improve creators' control over their content. We are working towards that goal through best practices, new standards, and dataset techniques.

- **Best practices in training.** Datasets used for training Stable Diffusion respect protocols like robots.txt that indicate whether a website consents to automated data collection.² In addition, Stability AI has solicited opt-out requests from creators, and will honor opt-out requests in upcomingreleases of Stable Diffusion. Going forward, we are exploring new standards for opt-outs, so that opt-out metadata follows the content wherever it goes.
- Authenticity standards. We are implementing content authenticity standards for our hosted AI applications so that users and platforms can identify AI-assisted content.³ By distinguishing AI content, these standards can help ensure that users apply an appropriate standard of care when interacting with AI content; help to limit the spread of disinformation; and help to protect human artists from unfair mimicry or passing off.
- **Quality of datasets.** We are working to improve the quality of datasets. For example, improving diversity and reducing duplication in training data can help to mitigate the risk of "overfitting", which occurs when a model erroneously overrepresents elements of a particular image from the dataset (e.g., if a model has only seen sunsets, it might learn that the sky is always orange). Additionally, datasets that are more representative of diverse culture, language, demographics, and values can help to mitigate the risk of bias.



Above: An example of content authenticity metadata indicating an image was generated with an AI tool.

We believe the community will continue to value human-generated content in a post-AI creative economy. Photoshop didn't destroy photography, and autocorrect didn't replace literature. We carry a complex digital camera in our pockets, yet we continue to value painting. We have machines that can run faster than athletes, but we continue to place a premium on sport. The same will be true of AI tools, and we welcome an ongoing dialogue with the creative community about the fair deployment of these technologies.

² Stable Diffusion is trained on a subset of the open LAION-5B dataset.

³ C2PA standards in partnership with the Content Authenticity Initiative (CAI), available here.