

# Stable Diffusion

**Stable Diffusion** is a deep learning, text-to-image model released in 2022. It is primarily used to generate detailed images conditioned on text descriptions, though it can also be applied to other tasks such as inpainting, outpainting, and generating image-to-image translations guided by a text prompt.<sup>[3]</sup> It was developed by researchers from the CompVis Group at Ludwig Maximilian University of Munich and Runway with a compute donation by Stability AI and training data from non-profit organizations.<sup>[4][5][6][7]</sup>

Stable Diffusion is a latent diffusion model, a kind of deep generative artificial neural network. Its code and model weights have been released publicly,<sup>[8]</sup> and it can run on most consumer hardware equipped with a modest GPU with at least 8 GB VRAM. This marked a departure from previous proprietary text-to-image models such as DALL-E and Midjourney which were accessible only via cloud services.<sup>[9][10]</sup>

## Development

The development of Stable Diffusion was funded and shaped by the start-up company Stability AI.<sup>[11][10][12][13]</sup> The technical license for the model was released by the CompVis group at Ludwig Maximilian University of Munich.<sup>[10]</sup> Development was led by Patrick Esser of Runway and Robin Rombach of CompVis, who were among the researchers who had earlier invented the latent diffusion model architecture used by Stable Diffusion.<sup>[7]</sup> Stability AI also credited EleutherAI and LAION (a German nonprofit which assembled the dataset on which Stable Diffusion was trained) as supporters of the project.<sup>[7]</sup>

In October 2022, Stability AI raised US\$101 million in a round led by Lightspeed Venture Partners and Coatue Management.<sup>[14]</sup>

## Technology

### Architecture

Stable Diffusion uses a kind of diffusion model (DM), called a latent diffusion model (LDM) developed by the CompVis group at LMU Munich.<sup>[15][8]</sup> Introduced in 2015, diffusion

### Stable Diffusion



An image generated by Stable Diffusion based on the text prompt "a photograph of an astronaut riding a horse"

<b><u>Original author(s)</u></b>	Runway, CompVis, and Stability AI
<b><u>Developer(s)</u></b>	Stability AI
<b><u>Initial release</u></b>	August 22, 2022
<b><u>Stable release</u></b>	2.1 (model) <sup>[1]</sup> / December 7, 2022
<b><u>Repository</u></b>	<u>github.com</u> / <u>Stability-AI</u> / <u>stablediffusion</u> ( <u>https://github.com/Stability-AI/stablediffusion</u> )
<b><u>Written in</u></b>	<u>Python</u> <sup>[2]</sup>
<b><u>Operating system</u></b>	Any that support <u>CUDA kernels</u>
<b><u>Type</u></b>	<u>Text-to-image model</u>
<b><u>License</u></b>	<u>Creative ML</u> <u>OpenRAIL-M</u>
<b><u>Website</u></b>	<u>ommer-lab.com</u> / <u>research/latent-diffusion-models/</u>

models are trained with the objective of removing successive applications of Gaussian noise on training images, which can be thought of as a sequence of denoising autoencoders. Stable Diffusion consists of 3 parts: the variational autoencoder (VAE), U-Net, and an optional text encoder.<sup>[16]</sup> The VAE encoder compresses the image from pixel space to a smaller dimensional latent space, capturing a more fundamental semantic meaning of the image.<sup>[15]</sup> Gaussian noise is iteratively applied to the compressed latent representation during forward diffusion.<sup>[16]</sup> The U-Net block, composed of a ResNet backbone, denoises the output from forward diffusion backwards to obtain a latent representation. Finally, the VAE decoder generates the final image by converting the representation back into pixel space.<sup>[16]</sup>

(<https://ommer-la-b.com/research/latent-diffusion-models/>)

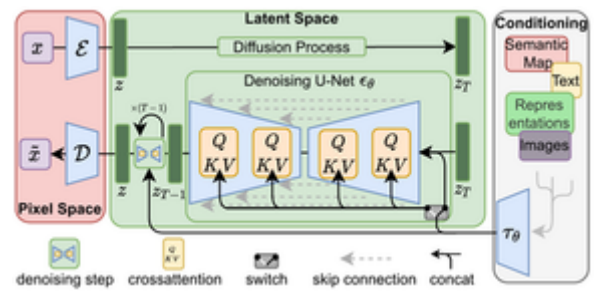


Diagram of the latent diffusion architecture used by Stable Diffusion

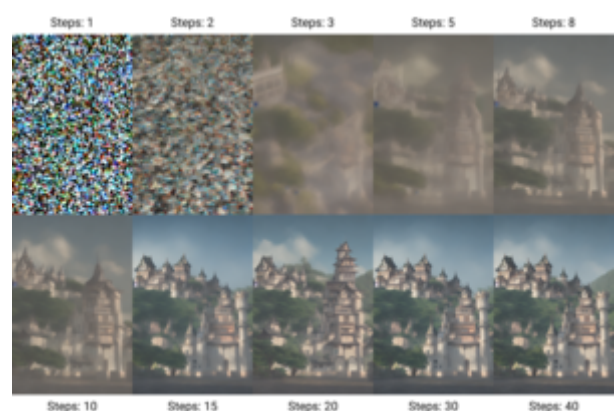
The denoising step can be flexibly conditioned on a string of text, an image, or another modality. The encoded conditioning data is exposed to denoising U-Nets via a cross-attention mechanism.<sup>[16]</sup> For conditioning on text, the fixed, pretrained CLIP ViT-L/14 text encoder is used to transform text prompts to an embedding space.<sup>[8]</sup> Researchers point to increased computational efficiency for training and generation as an advantage of LDMs.<sup>[7][15]</sup>

With 860 millions of parameters in the U-Net and 123 millions in the text encoder, Stable Diffusion is considered relatively lightweight by 2022 standards, and unlike other diffusion models, it can run on consumer GPUs.<sup>[17]</sup>

## Training data

Stable Diffusion was trained on pairs of images and captions taken from LAION-5B, a publicly available dataset derived from Common Crawl data scraped from the web, where 5 billion image-text pairs were classified based on language and filtered into separate datasets by resolution, a predicted likelihood of containing a watermark, and predicted "aesthetic" score (e.g. subjective visual quality).<sup>[18]</sup> The dataset was created by LAION, a German non-profit which receives funding from Stability AI.<sup>[18][19]</sup> The Stable Diffusion model was trained on three subsets of LAION-5B: laion2B-en, laion-high-resolution, and laion-aesthetics v2 5+.<sup>[18]</sup> A third-party analysis of the model's training data identified that out of a smaller subset of 12 million images taken from the original wider dataset used, approximately 47% of the sample size of images came from 100 different domains, with Pinterest taking up 8.5% of the subset, followed by websites such as WordPress, Blogspot, Flickr, DeviantArt and Wikimedia Commons.<sup>[20][18]</sup>

## Training procedures



The denoising process used by Stable Diffusion. The model generates images by iteratively denoising random noise until a configured number of steps have been reached, guided by the CLIP text encoder pretrained on concepts along with the attention mechanism, resulting in the desired image depicting a representation of the trained concept.

The model was initially trained on the laion2B-en and laion-high-resolution subsets, with the last few rounds of training done on LAION-Aesthetics v2 5+, a subset of 600 million captioned images which the LAION-Aesthetics Predictor V2 predicted that humans would, on average, give a score of at least 5 out of 10 when asked to rate how much they liked them.<sup>[21][18][22]</sup> The LAION-Aesthetics v2 5+ subset also excluded low-resolution images and images which LAION-5B-WatermarkDetection identified as carrying a watermark with greater than 80% probability.<sup>[18]</sup> Final rounds of training additionally dropped 10% of text conditioning to improve Classifier-Free Diffusion Guidance.<sup>[23]</sup>

The model was trained using 256 Nvidia A100 GPUs on Amazon Web Services for a total of 150,000 GPU-hours, at a cost of \$600,000.<sup>[24][25][26]</sup>

## Limitations

Stable Diffusion has issues with degradation and inaccuracies in certain scenarios. Initial releases of the model were trained on a dataset that consists of 512×512 resolution images, meaning that the quality of generated images noticeably degrades when user specifications deviate from its "expected" 512×512 resolution;<sup>[27]</sup> the version 2.0 update of the Stable Diffusion model later introduced the ability to natively generate images at 768×768 resolution.<sup>[28]</sup> Another challenge is in generating human limbs due to poor data quality of limbs in the LAION database.<sup>[29]</sup> The model is insufficiently trained to understand human limbs and faces due to the lack of representative features in the database, and prompting the model to generate images of such type can confound the model.<sup>[30]</sup>

Accessibility for individual developers can also be a problem. In order to customize the model for new use cases that are not included in the dataset, such as generating anime characters ("waifu diffusion"),<sup>[31]</sup> new data and further training are required. Fine-tuned adaptations of Stable Diffusion created through additional retraining have been used for a variety of different use-cases, from medical imaging<sup>[32]</sup> to algorithmically generated music.<sup>[33]</sup> However, this fine-tuning process is sensitive to the quality of new data; low resolution images or different resolutions from the original data can not only fail to learn the new task but degrade the overall performance of the model. Even when the model is additionally trained on high quality images, it is difficult for individuals to run models in consumer electronics. For example, the training process for waifu-diffusion requires a minimum 30 GB of VRAM,<sup>[34]</sup> which exceeds the usual resource provided in such consumer GPUs as Nvidia's GeForce 30 series, which has only about 12 GB.<sup>[35]</sup>

The creators of Stable Diffusion acknowledge the potential for algorithmic bias, as the model was primarily trained on images with English descriptions.<sup>[25]</sup> As a result, generated images reinforce social biases and are from a western perspective, as the creators note that the model lacks data from other communities and cultures. The model gives more accurate results for prompts that are written in English in comparison to those written in other languages, with western or white cultures often being the default representation.<sup>[25]</sup>

## End-user fine-tuning

To address the limitations of the model's initial training, end-users may opt to implement additional training to fine-tune generation outputs to match more specific use-cases. There are three methods in which user-accessible fine-tuning can be applied to a Stable Diffusion model checkpoint:

- An "embedding" can be trained from a collection of user-provided images, and allows the model to generate visually similar images whenever the name of the embedding is used within a generation prompt.<sup>[36]</sup> Embeddings are based on the "textual inversion" concept developed by researchers from Tel Aviv University in 2022 with support from Nvidia, where vector representations for specific tokens used by the model's text encoder are linked to new

pseudo-words. Embeddings can be used to reduce biases within the original model, or mimic visual styles.<sup>[37]</sup>

- A "hypernetwork" is a small pretrained neural network that is applied to various points within a larger neural network, and refers to the technique created by NovelAI developer Kurumuz in 2021, originally intended for text-generation transformer models. Hypernetworks steer results towards a particular direction, allowing Stable Diffusion-based models to imitate the art style of specific artists, even if the artist is not recognised by the original model; they process the image by finding key areas of importance such as hair and eyes, and then patch these areas in secondary latent space.<sup>[38]</sup>
- DreamBooth is a deep learning generation model developed by researchers from Google Research and Boston University in 2022 which can fine-tune the model to generate precise, personalised outputs that depict a specific subject, following training via a set of images which depict the subject.<sup>[39]</sup>

## Capabilities

---

The Stable Diffusion model supports the ability to generate new images from scratch through the use of a text prompt describing elements to be included or omitted from the output.<sup>[8]</sup> Existing images can be re-drawn by the model to incorporate new elements described by a text prompt (a process known as "guided image synthesis"<sup>[40]</sup>) through its diffusion-denoising mechanism.<sup>[8]</sup> In addition, the model also allows the use of prompts to partially alter existing images via inpainting and outpainting, when used with an appropriate user interface that supports such features, of which numerous different open source implementations exist.<sup>[41]</sup>

Stable Diffusion is recommended to be run with 10 GB or more VRAM, however users with less VRAM may opt to load the weights in float16 precision instead of the default float32 to tradeoff model performance with lower VRAM usage.<sup>[27]</sup>

## Text to image generation

The text to image sampling script within Stable Diffusion, known as "txt2img", consumes a text prompt in addition to assorted option parameters covering sampling types, output image dimensions, and seed values. The script outputs an image file based on the model's interpretation of the prompt.<sup>[8]</sup> Generated images are tagged with an invisible digital watermark to allow users to identify an image as generated by Stable Diffusion,<sup>[8]</sup> although this watermark loses its efficacy if the image is resized or rotated.<sup>[42]</sup>

Each txt2img generation will involve a specific seed value which affects the output image. Users may opt to randomize the seed in order to explore



Demonstration of the effect of negative prompts on image generation

- **Top:** no negative prompt
- **Centre:** "green trees"
- **Bottom:** "round stones, round rocks"

different generated outputs, or use the same seed to obtain the same image output as a previously generated image.<sup>[27]</sup> Users are also able to adjust the number of inference steps for the sampler; a higher value takes a longer duration of time, however a smaller value may result in visual defects.<sup>[27]</sup> Another configurable option, the classifier-free guidance scale value, allows the user to adjust how closely the output image adheres to the prompt.<sup>[23]</sup> More experimentative use cases may opt for a lower scale value, while use cases aiming for more specific outputs may use a higher value.<sup>[27]</sup>

Additional text2img features are provided by front-end implementations of Stable Diffusion, which allow users to modify the weight given to specific parts of the text prompt. Emphasis markers allow users to add or reduce emphasis to keywords by enclosing them with brackets.<sup>[43]</sup> An alternative method of adjusting weight to parts of the prompt are "negative prompts". Negative prompts are a feature included in some front-end implementations, including Stability AI's own DreamStudio cloud service, and allow the user to specify prompts which the model should avoid during image generation. The specified prompts may be undesirable image features that would otherwise be present within image outputs due to the positive prompts provided by the user, or due to how the model was originally trained, with mangled human hands being a common example.<sup>[41][1]</sup>

## Image modification

Stable Diffusion also includes another sampling script, "img2img", which consumes a text prompt, path to an existing image, and strength value between 0.0 and 1.0. The script outputs a new image based on the original image that also features elements provided within the text prompt. The strength value denotes the amount of noise added to the output image. A higher strength value produces more variation within the image but may produce an image that is not semantically consistent with the prompt provided.<sup>[8]</sup>

The ability of img2img to add noise to the original image makes it potentially useful for data anonymization and data augmentation, in which the visual features of image data are changed and anonymized.<sup>[44]</sup> The same process may also be useful for image upscaling, in which the resolution of an image is increased, with more detail potentially being added to the image.<sup>[44]</sup> Additionally, Stable Diffusion has been experimented with as a tool for image compression. Compared to JPEG and WebP, the recent methods used for image compression in Stable Diffusion face limitations in preserving small text and faces.<sup>[45]</sup>

Additional use-cases for image modification via img2img are offered by numerous front-end implementations of the Stable Diffusion model. Inpainting involves selectively modifying a portion of an existing image delineated by a user-provided layer mask, which fills the masked space with newly generated content based on the provided prompt.<sup>[41]</sup> A dedicated model specifically fine-tuned for inpainting use-cases was created by Stability AI alongside the release of Stable Diffusion 2.0.<sup>[28]</sup> Conversely, outpainting extends an image beyond its original dimensions, filling the previously empty space with content generated based on the provided prompt.<sup>[41]</sup>

A depth-guided model, named "depth2img", was introduced with the release of Stable Diffusion 2.0 on November 24, 2022; this model infers the depth of the provided input image, and generates a new output image based on both the text prompt and the depth information, which allows the coherence and depth of the original input image to be maintained in the generated output.<sup>[28]</sup>

## ControlNet

ControlNet<sup>[46]</sup> is a neural network architecture designed to manage diffusion models by incorporating additional conditions. It duplicates the weights of neural network blocks into a "locked" copy and a "trainable" copy. The "trainable" copy learns the desired condition, while the "locked" copy preserves the

original model. This approach ensures that training with small datasets of image pairs does not compromise the integrity of production-ready diffusion models. The "zero convolution" is a 1×1 convolution with both weight and bias initialized to zero. Before training, all zero convolutions produce zero output, preventing any distortion caused by ControlNet. No layer is trained from scratch; the process is still fine-tuning, keeping the original model secure. This method enables training on small-scale or even personal devices.

## Usage and controversy

---

Stable Diffusion claims no rights on generated images and freely gives users the rights of usage to any generated images from the model provided that the image content is not illegal or harmful to individuals. The freedom provided to users over image usage has caused controversy over the ethics of ownership, as Stable Diffusion and other generative models are trained from copyrighted images without the owner's consent.<sup>[47]</sup>

As visual styles and compositions are not subject to copyright, it is often interpreted that users of Stable Diffusion who generate images of artworks should not be considered to be infringing upon the copyright of visually similar works.<sup>[48]</sup> However, individuals depicted in generated images may be protected by personality rights if their likeness is used,<sup>[48]</sup> and intellectual property such as recognizable brand logos still remain protected by copyright. Nonetheless, visual artists have expressed concern that widespread usage of image synthesis software such as Stable Diffusion may eventually lead to human artists, along with photographers, models, cinematographers, and actors, gradually losing commercial viability against AI-based competitors.<sup>[13]</sup>

Stable Diffusion is notably more permissive in the types of content users may generate, such as violent or sexually explicit imagery, in comparison to other commercial products based on generative AI.<sup>[49]</sup> Addressing the concerns that the model may be used for abusive purposes, CEO of Stability AI, Emad Mostaque, argues that "[it is] peoples' responsibility as to whether they are ethical, moral, and legal in how they operate this technology",<sup>[10]</sup> and that putting the capabilities of Stable Diffusion into the hands of the public would result in the technology providing a net benefit, in spite of the potential negative consequences.<sup>[10]</sup> In addition, Mostaque argues that the intention behind the open availability of Stable Diffusion is to end corporate control and dominance over such technologies, who have previously only developed closed AI systems for image synthesis.<sup>[10][49]</sup> This is reflected by the fact that any restrictions Stability AI places on the content that users may generate can easily be bypassed due to the availability of the source code.<sup>[47]</sup>

## Litigation

---

In January of 2023, three artists: Sarah Andersen, Kelly McKernan, and Karla Ortiz filed a copyright infringement lawsuit against Stability AI, Midjourney, and DeviantArt, claiming that these companies have infringed the rights of millions of artists by training AI tools on five billion images scraped from the web without the consent of the original artists.<sup>[50]</sup> The same month, Stability AI was also sued by Getty Images for using its images in the training data.<sup>[12]</sup>

## License

---

Unlike models like DALL-E, Stable Diffusion makes its source code available,<sup>[51][8]</sup> along with the model (pretrained weights). It applies the Creative ML OpenRAIL-M license, a form of Responsible AI License (RAIL), to the model (M).<sup>[52]</sup> The licence prohibits certain use cases, including crime, libel, harassment, doxing, "exploiting ... minors", giving medical advice, automatically creating legal obligations, producing



legal evidence, and "discriminating against or harming individuals or groups based on ... social behavior or ... personal or personality characteristics ... [or] legally protected characteristics or categories".<sup>[53][54]</sup> The user owns the rights to their generated output images, and is free to use them commercially.<sup>[55]</sup>

## See also

---

- [15.ai](#)
- [Artificial intelligence art](#)
- [Craiyon](#)
- [Imagen \(Google Brain\)](#)
- [Synthography](#)
- [Hugging Face](#)

## References

---

1. "Stable Diffusion v2.1 and DreamStudio Updates 7-Dec 22" (<https://stability.ai/blog/stablediffusion2-1-release7-dec-2022>). *stability.ai*. Archived (<https://web.archive.org/web/20221210062732/https://stability.ai/blog/stablediffusion2-1-release7-dec-2022>) from the original on December 10, 2022.
2. Ryan O'Connor (August 23, 2022). "How to Run Stable Diffusion Locally to Generate Images" (<https://www.assemblyai.com/blog/how-to-run-stable-diffusion-locally-to-generate-images/>). Retrieved May 4, 2023.
3. "Diffuse The Rest - a Hugging Face Space by huggingface" (<https://huggingface.co/spaces/huggingface/diffuse-the-rest>). *huggingface.co*. Archived (<https://web.archive.org/web/20220905141431/https://huggingface.co/spaces/huggingface/diffuse-the-rest>) from the original on 2022-09-05. Retrieved 2022-09-05.
4. "Leaked deck raises questions over Stability AI's Series A pitch to investors" (<https://sifted.eu/articles/stability-ai-fundraise-leak>). *sifted.eu*. Retrieved 2023-06-20.
5. "Revolutionizing image generation by AI: Turning text into images" (<https://www.lmu.de/en/newsroom/news-overview/news/revolutionizing-image-generation-by-ai-turning-text-into-images.html>). *www.lmu.de*. Retrieved 2023-06-21.
6. Mostaque, Emad (November 2, 2022). "Stable Diffusion came from the Machine Vision & Learning research group (CompVis) @LMU\_Muenchen" (<https://twitter.com/EMostaque/status/1587844074064822274?lang=en>). *Twitter*. Retrieved 2023-06-22.
7. "Stable Diffusion Launch Announcement" (<https://stability.ai/blog/stable-diffusion-announcement>). *Stability.Ai*. Archived (<https://web.archive.org/web/20220905105009/https://stability.ai/blog/stable-diffusion-announcement>) from the original on 2022-09-05. Retrieved 2022-09-06.
8. "Stable Diffusion Repository on GitHub" (<https://github.com/CompVis/stable-diffusion>). CompVis - Machine Vision and Learning Research Group, LMU Munich. 17 September 2022. Retrieved 17 September 2022.
9. "The new killer app: Creating AI art will absolutely crush your PC" (<https://www.pcworld.com/article/916785/creating-ai-art-local-pc-stable-diffusion.html>). *PCWorld*. Archived (<https://web.archive.org/web/20220831065139/https://www.pcworld.com/article/916785/creating-ai-art-local-pc-stable-diffusion.html>) from the original on 2022-08-31. Retrieved 2022-08-31.
10. Vincent, James (15 September 2022). "Anyone can use this AI art generator — that's the risk" (<https://www.theverge.com/2022/9/15/23340673/ai-image-generation-stable-diffusion-explained-ethics-copyright-data>). *The Verge*.

11. "The AI Founder Taking Credit For Stable Diffusion's Success Has A History Of Exaggeration" (<https://www.forbes.com/sites/kenrickcai/2023/06/04/stable-diffusion-emad-mostaque-stability-ai-exaggeration/?sh=347a8fcb75c5>). *www.forbes.com*. Retrieved 2023-06-20.
12. Korn, Jennifer (2023-01-17). "Getty Images suing the makers of popular AI art tool for allegedly stealing photos" (<https://www.cnn.com/2023/01/17/tech/getty-images-stability-ai-lawsuit/index.html>). *CNN*. Retrieved 2023-01-22.
13. Heikkilä, Melissa (16 September 2022). "This artist is dominating AI-generated art. And he's not happy about it" (<https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/>). *MIT Technology Review*.
14. Wiggers, Kyle (17 October 2022). "Stability AI, the startup behind Stable Diffusion, raises \$101M" (<https://techcrunch.com/2022/10/17/stability-ai-the-startup-behind-stable-diffusion-raises-101m/>). *Techcrunch*. Retrieved 2022-10-17.
15. Rombach; Blattmann; Lorenz; Esser; Ommer (June 2022). *High-Resolution Image Synthesis with Latent Diffusion Models* ([https://openaccess.thecvf.com/content/CVPR2022/papers/Rombach\\_High-Resolution\\_Image\\_Synthesis\\_With\\_Latent\\_Diffusion\\_Models\\_CVPR\\_2022\\_paper.pdf](https://openaccess.thecvf.com/content/CVPR2022/papers/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.pdf)) (PDF). International Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA. pp. 10684–10695. [arXiv:2112.10752](https://arxiv.org/abs/2112.10752) (<https://arxiv.org/abs/2112.10752>).
16. Alammar, Jay. "The Illustrated Stable Diffusion" (<https://jalammar.github.io/illustrated-stable-diffusion/>). *jalammar.github.io*. Retrieved 2022-10-31.
17. [https://huggingface.co/docs/diffusers/v0.5.1/en/api/pipelines/stable\\_diffusion](https://huggingface.co/docs/diffusers/v0.5.1/en/api/pipelines/stable_diffusion)
18. Baio, Andy (2022-08-30). "Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator" (<https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/>). *Waxy.org*. Retrieved 2022-11-02.
19. "This artist is dominating AI-generated art. And he's not happy about it" (<https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/>). *MIT Technology Review*. Retrieved 2022-11-02.
20. Ivanovs, Alex (2022-09-08). "Stable Diffusion: Tutorials, Resources, and Tools" (<https://stackdiary.com/stable-diffusion-resources/>). *Stack Diary*. Retrieved 2022-11-02.
21. Schuhmann, Christoph (2022-11-02), *CLIP+MLP Aesthetic Score Predictor* (<https://github.com/christophschuhmann/improved-aesthetic-predictor>), retrieved 2022-11-02
22. "LAION-Aesthetics | LAION" (<https://laion.ai/blog/laion-aesthetics>). *laion.ai*. Archived (<https://web.archive.org/web/20220826121216/https://laion.ai/blog/laion-aesthetics/>) from the original on 2022-08-26. Retrieved 2022-09-02.
23. Ho, Jonathan; Salimans, Tim (2022-07-25). "Classifier-Free Diffusion Guidance". [arXiv:2207.12598](https://arxiv.org/abs/2207.12598) (<https://arxiv.org/abs/2207.12598>) [cs.LG (<https://arxiv.org/archive/cs.LG>)].
24. Mostaque, Emad (August 28, 2022). "Cost of construction" (<https://twitter.com/emostaque/status/1563870674111832066>). *Twitter*. Archived (<https://web.archive.org/web/20220906155426/https://twitter.com/EMostaque/status/1563870674111832066>) from the original on 2022-09-06. Retrieved 2022-09-06.
25. "CompVis/stable-diffusion-v1-4 · Hugging Face" (<https://huggingface.co/CompVis/stable-diffusion-v1-4>). *huggingface.co*. Retrieved 2022-11-02.
26. Wiggers, Kyle (2022-08-12). "A startup wants to democratize the tech behind DALL-E 2, consequences be damned" (<https://techcrunch.com/2022/08/12/a-startup-wants-to-democratize-the-tech-behind-dall-e-2-consequences-be-damned/>). *TechCrunch*. Retrieved 2022-11-02.
27. "Stable Diffusion with 🖌️ Diffusers" ([https://huggingface.co/blog/stable\\_diffusion](https://huggingface.co/blog/stable_diffusion)). *huggingface.co*. Retrieved 2022-10-31.



28. "Stable Diffusion 2.0 Release" (<https://stability.ai/blog/stable-diffusion-v2-release>). *stability.ai*. Archived (<https://web.archive.org/web/20221210062729/https://stability.ai/blog/stable-diffusion-v2-release>) from the original on December 10, 2022.
29. "LAION" (<https://laion.ai/>). *laion.ai*. Retrieved 2022-10-31.
30. "Generating images with Stable Diffusion" (<https://blog.paperspace.com/generating-images-with-stable-diffusion/>). *Paperspace Blog*. 2022-08-24. Retrieved 2022-10-31.
31. "hakurei/waifu-diffusion · Hugging Face" (<https://huggingface.co/hakurei/waifu-diffusion>). *huggingface.co*. Retrieved 2022-10-31.
32. Chambon, Pierre; Bluethgen, Christian; Langlotz, Curtis P.; Chaudhari, Akshay (2022-10-09). "Adapting Pretrained Vision-Language Foundational Models to Medical Imaging Domains". *arXiv:2210.04133* (<https://arxiv.org/abs/2210.04133>) [cs.CV (<https://arxiv.org/archive/cs>)].
33. Seth Forsgren; Hayk Martiros. "Riffusion - Stable diffusion for real-time music generation" (<https://www.riffusion.com/about>). *Riffusion*. Archived (<https://web.archive.org/web/20221216092717/https://www.riffusion.com/about>) from the original on December 16, 2022.
34. Mercurio, Anthony (2022-10-31), *Waifu Diffusion* (<https://github.com/harubaru/waifu-diffusion/blob/6bf942eb6368ebf6bcbbb24b6ba8197bda6582a0/docs/en/training/README.md>), retrieved 2022-10-31
35. Smith, Ryan. "NVIDIA Quietly Launches GeForce RTX 3080 12GB: More VRAM, More Power, More Money" (<https://www.anandtech.com/show/17204/nvidia-quietly-launches-geforce-rtx-3080-12gb-more-vram-more-power-more-money>). *www.anandtech.com*. Retrieved 2022-10-31.
36. Dave James (October 28, 2022). "I thrashed the RTX 4090 for 8 hours straight training Stable Diffusion to paint like my uncle Hermann" (<https://www.pcgamer.com/nvidia-rtx-4090-stable-diffusion-training-aharon-kahana/>). *PC Gamer*. Archived (<https://web.archive.org/web/20221109154310/https://www.pcgamer.com/nvidia-rtx-4090-stable-diffusion-training-aharon-kahana/>) from the original on November 9, 2022.
37. Gal, Rinon; Alaluf, Yuval; Atzmon, Yuval; Patashnik, Or; Bermano, Amit H.; Chechik, Gal; Cohen-Or, Daniel (2022-08-02). "An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion". *arXiv:2208.01618* (<https://arxiv.org/abs/2208.01618>) [cs.CV (<https://arxiv.org/archive/cs>)].
38. "NovelAI Improvements on Stable Diffusion" (<https://blog.novelai.net/novelai-improvements-on-stable-diffusion-e10d38db82ac>). *NovelAI*. October 11, 2022. Archived (<https://archive.vn/x9zHS>) from the original on October 27, 2022.
39. Yuki Yamashita (September 1, 2022). "愛犬の合成画像を生成できるAI 文章で指示するだけでコスプレ 米Googleが開発" (<https://www.itmedia.co.jp/news/articles/2209/01/news041.html>). *ITmedia Inc.* (in Japanese). Archived (<https://web.archive.org/web/20220831232021/https://www.itmedia.co.jp/news/articles/2209/01/news041.html>) from the original on August 31, 2022.
40. Meng, Chenlin; He, Yutong; Song, Yang; Song, Jiaming; Wu, Jiajun; Zhu, Jun-Yan; Ermon, Stefano (August 2, 2021). "SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations". *arXiv:2108.01073* (<https://arxiv.org/abs/2108.01073>) [cs.CV (<https://arxiv.org/archive/cs>)].
41. "Stable Diffusion web UI" (<https://github.com/AUTOMATIC1111/stable-diffusion-webui-feature-showcase>). *GitHub*. 10 November 2022.
42. *invisible-watermark* (<https://github.com/ShieldMnt/invisible-watermark/blob/9802ce3e0c3a5ec43b41d503f156717f0c739584/README.md>), Shield Mountain, 2022-11-02, retrieved 2022-11-02
43. "stable-diffusion-tools/emphasis at master · JohannesGaessler/stable-diffusion-tools" (<https://github.com/JohannesGaessler/stable-diffusion-tools>). *GitHub*. Retrieved 2022-11-02.

44. Luzi, Lorenzo; Siahkoohi, Ali; Mayer, Paul M.; Casco-Rodriguez, Josue; Baraniuk, Richard (2022-10-21). "Boomerang: Local sampling on image manifolds using diffusion models". [arXiv:2210.12100](https://arxiv.org/abs/2210.12100) (<https://arxiv.org/abs/2210.12100>) [cs.CV (<https://arxiv.org/archive/cs>)].
45. Bühlmann, Matthias (2022-09-28). "Stable Diffusion Based Image Compression" (<https://pub.towardsai.net/stable-diffusion-based-image-compresssion-6f1f0a399202>). *Medium*. Retrieved 2022-11-02.
46. Zhang, Lvmin (10 February 2023). "Adding Conditional Control to Text-to-Image Diffusion Models". [arXiv:2302.05543](https://arxiv.org/abs/2302.05543) (<https://arxiv.org/abs/2302.05543>) [cs.CV (<https://arxiv.org/archive/cs>)].
47. Cai, Kenrick. "Startup Behind AI Image Generator Stable Diffusion Is In Talks To Raise At A Valuation Up To \$1 Billion" (<https://www.forbes.com/sites/kenrickcai/2022/09/07/stability-ai-funding-round-1-billion-valuation-stable-diffusion-text-to-image/>). *Forbes*. Retrieved 2022-10-31.
48. "高性能画像生成AI「Stable Diffusion」無料リリース。「kawaii」までも理解し創造する画像生成AI" (<https://automaton-media.com/articles/newsjp/20220824-216074/>). *Automaton Media* (in Japanese). August 24, 2022.
49. Ryo Shimizu (August 26, 2022). "Midjourneyを超えた？無料の作画AI「#StableDiffusion」が「AIを民主化した」と断言できる理由" (<https://www.businessinsider.jp/post-258369>). *Business Insider Japan* (in Japanese).
50. James Vincent "AI art tools Stable Diffusion and Midjourney targeted with copyright lawsuit" *The Verge*, 16 January, 2023. (<https://www.theverge.com/2023/1/16/23557098/generative-ai-art-copyright-legal-lawsuit-stable-diffusion-midjourney-deviantart>)
51. "Stable Diffusion Public Release" (<https://stability.ai/blog/stable-diffusion-public-release>). *Stability.Ai*. Archived (<https://web.archive.org/web/20220830210535/https://stability.ai/blog/stable-diffusion-public-release>) from the original on 2022-08-30. Retrieved 2022-08-31.
52. "From RAIL to Open RAIL: Topologies of RAIL Licenses" (<https://www.licenses.ai/blog/2022/8/18/naming-convention-of-responsible-ai-licenses>). *Responsible AI Licenses (RAIL)*. 18 August 2022. Retrieved 2023-02-20.
53. "Ready or not, mass video deepfakes are coming" (<https://www.washingtonpost.com/technology/2022/08/30/deep-fake-video-on-agg/>). *The Washington Post*. 2022-08-30. Archived (<https://web.archive.org/web/20220831115010/https://www.washingtonpost.com/technology/2022/08/30/deep-fake-video-on-agg/>) from the original on 2022-08-31. Retrieved 2022-08-31.
54. "License - a Hugging Face Space by CompVis" (<https://huggingface.co/spaces/CompVis/stable-diffusion-license>). *huggingface.co*. Archived (<https://web.archive.org/web/20220904215616/https://huggingface.co/spaces/CompVis/stable-diffusion-license>) from the original on 2022-09-04. Retrieved 2022-09-05.
55. Katsuo Ishida (August 26, 2022). "言葉で指示した画像を凄いAIが描き出す「Stable Diffusion」～画像は商用利用も可能" (<https://forest.watch.impress.co.jp/docs/review/1434893.html>). *Impress Corporation* (in Japanese).

---

## External links

- [Stable Diffusion Demo](https://huggingface.co/spaces/stabilityai/stable-diffusion) (<https://huggingface.co/spaces/stabilityai/stable-diffusion>)
  - [Interactive Explanation of Stable Diffusion](https://poloclub.github.io/diffusion-explainer/) (<https://poloclub.github.io/diffusion-explainer/>)
- 

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Stable\\_Diffusion&oldid=1163256641](https://en.wikipedia.org/w/index.php?title=Stable_Diffusion&oldid=1163256641)"

■