

A decorative graphic on the left side of the slide consisting of overlapping geometric shapes. It includes a blue parallelogram, a light green parallelogram, and a dark grey parallelogram, all with sharp, angular edges.

AI Art & Data Ethics

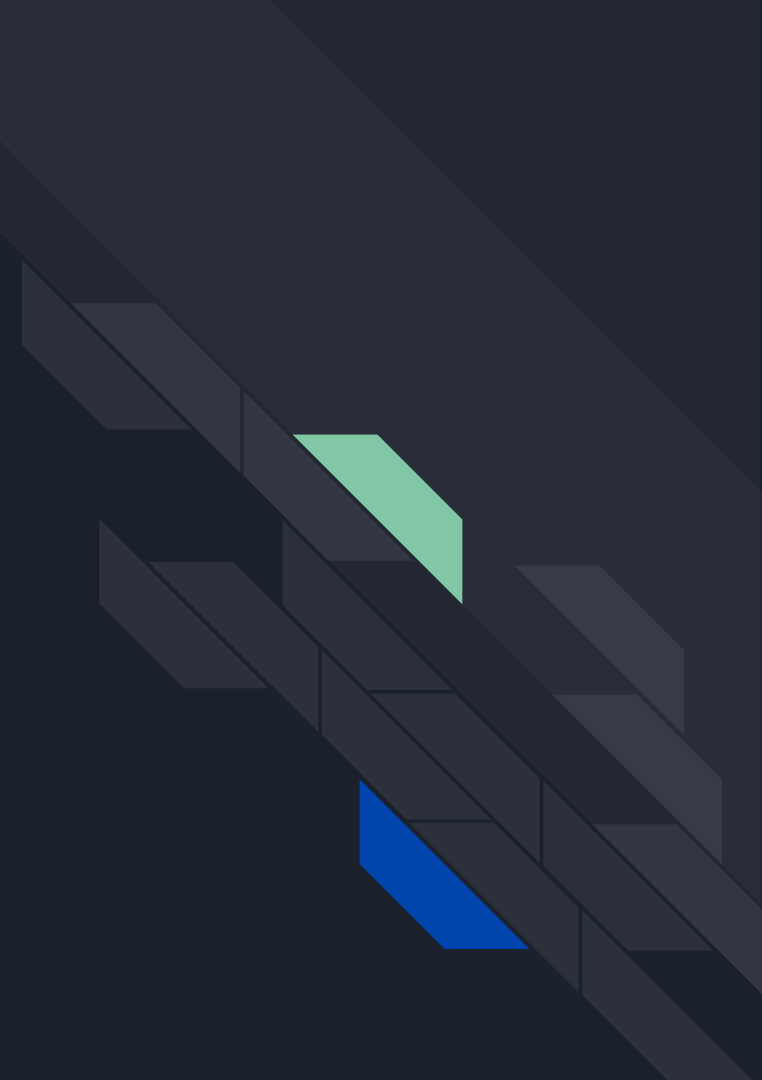
Pieper Smith and Ryan Lunas



Introduction

- ★ What Is AI Image Generation?
- ★ AI Image Generation Background
 - Texture Synthesis, Convolutional Neural Networks, Generative Adversarial Networks, Variational Autoencoders, NLP Transformers and LLMs, Diffusion Models, Contrastive Language-Image Pre-Training (CLIP), JFT-300M and LAION-5B
- ★ AI Image Generation Companies
- ★ AI Data Ethics: Copyright Lawsuits
- ★ Nightshade and Data Poisoning
- ★ A Few Pros and Cons of AI Image Generation

What Is AI Image Generation and How Does It Work?





What Is an AI Image Generator?

- ★ **Informally** - They are computer programs that take a **prompt** and generate an image that attempts to represent that **prompt**.
- ★ **Prompt** - Data given to the model that can be mapped to the output space such as text (text-to-image) or an image (image-to-image).
- ★ **More Formally** - They are computerized **generative models** trained by **machine learning** techniques to synthesize images that match a certain **prompt**.
- ★ **Generative models** are statistical models of the joint probability distribution between observable variables (in the feature space) and target variables (in the output space). (Feller, 1957)
- ★ **Machine learning** - The field of study concerned with enabling machines (computers) to “learn” from inputs and to imitate intelligence.

AI Image Generation – Origins

Texture Synthesis - Early through late 2000s, computer vision research to generate new images based on a source texture.

An example of an early image-to-image model.

Image Quilting

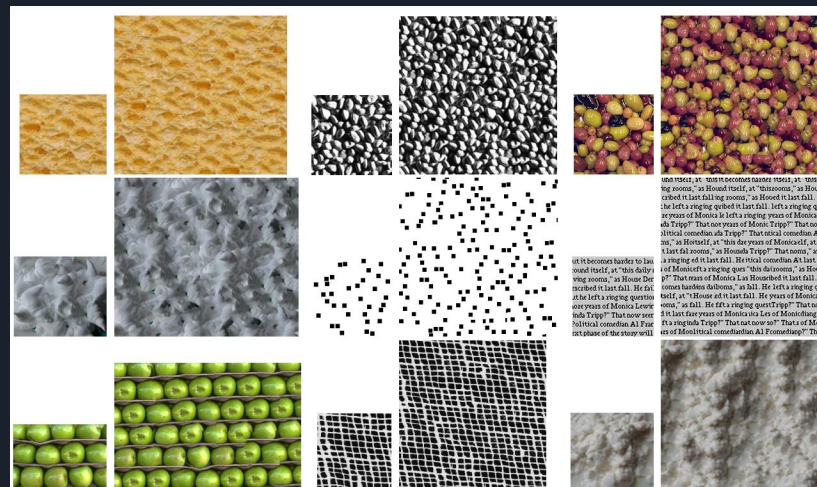
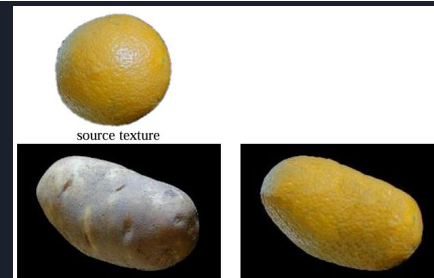



Figure 3: Image quilting synthesis results for a wide range of textures. The resulting texture (right side of each pair) is synthesized at twice the size of the original (left).



Texture Transfer

(Jiang et al., 2023)
(Efros & Freeman, 2001)



AI Image Generation – Recent Techniques

Convolutional Neural Networks (CNN) (2012) - They break up images into small pieces and apply techniques that bring out the features of each piece. The processed pieces are then passed to a neural network which will associate them with a single output which represents the most probable item.

Generative Adversarial Networks (GAN) (2014) - One of the first papers about these has a really good explanation: "The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency. Competition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles."
(Goodfellow et al., 2014)

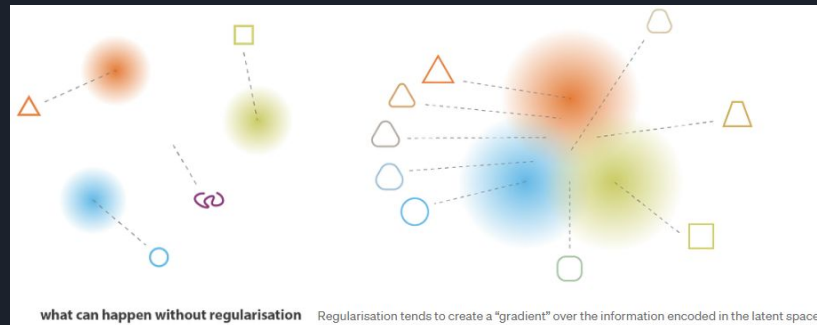
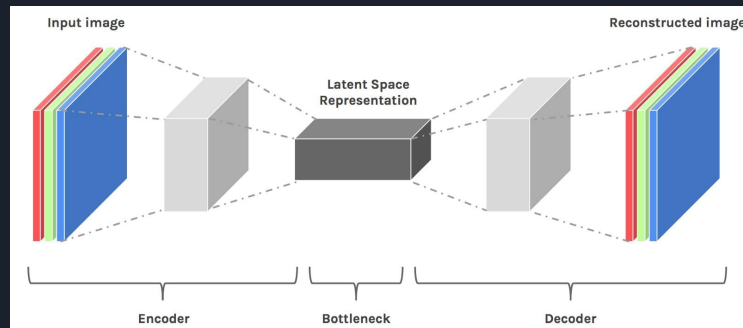
Recent Techniques Continued

Latent Space - A compressed, more organized space where data points with similar characteristics are located closer to each other.

Variational Autoencoders (VAEs) (2013) - A machine learning architecture that uses statistical methods to improve the inference capabilities of CNNs. An encoder processes inputs and embeds them in the **latent space**, the decoder uses this to produce a similar, but unique item.

(Asperti, 2023)
(Kingma & Welling, 2013)
(Rocca & Rocca, 2019)
(Despois, 2017)

A CNN



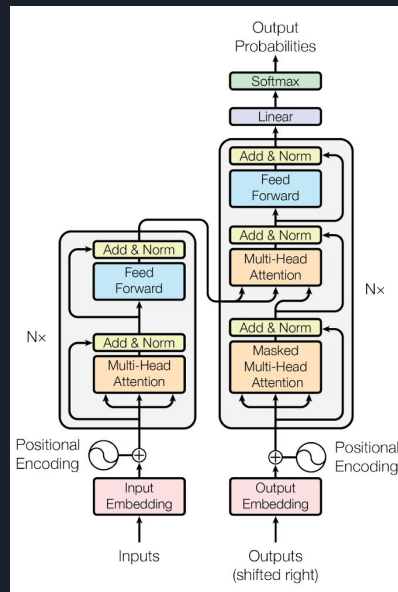
Regularisation

Recent Techniques Continued

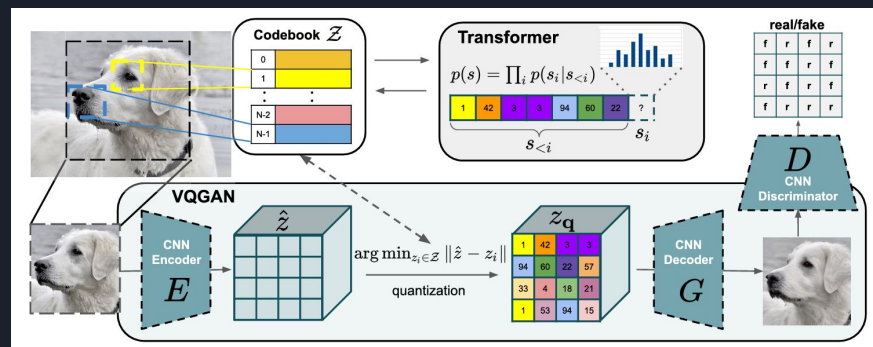
NLP Transformers and LLMs

(2017 - 2021) - NLP (Natural Language Processing) algorithms used to enhance or perform image synthesis. This led to the creation of NLP transformers, which use the transformer architecture in conjunction with NLP algorithms to model that have long-range interactions DALL-E is an image generator built using transformers and the GPT-3 LLM (Large Language Model).

(Vaswani et al., 2017)
(Esser et al., 2020)
(Ramesh et al., 2021)



Transformer Architecture

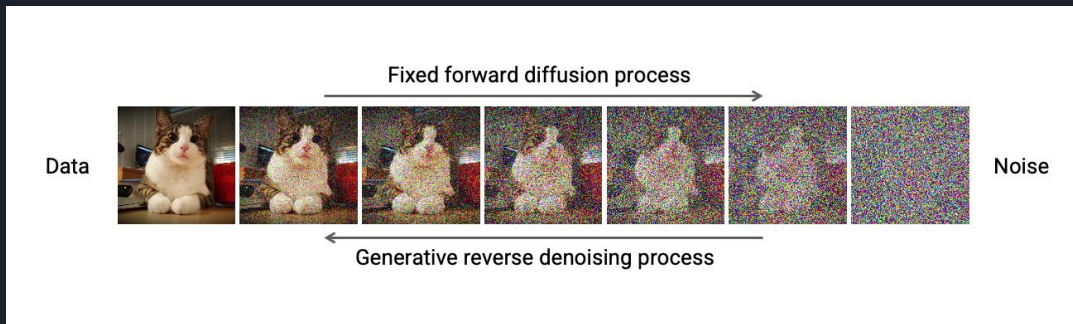


NLP Transformer

Recent Techniques Continued

Diffusion Models (2021) - Inspired by thermodynamic diffusion in physics. The models are trained using images that have partial noise applied to them. The model “learns” to undo this noise to make coherent images. Images are generated through repeated denoising of an image that is originally 100% noise. These models were used alongside NLP transformers and a set of contrastive pretrained autoencoders known as CLIP to create Stable Diffusion. (Rombach et al., 2021)

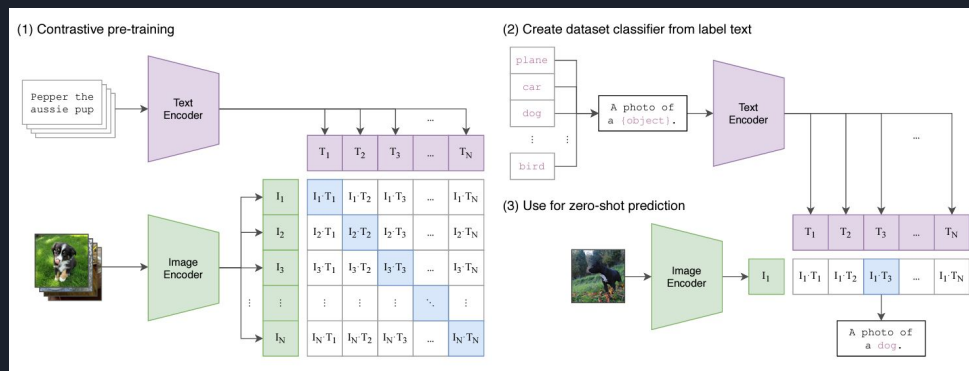
Diffusion Model Denoising



Recent Techniques Continued

Contrastive Language-Image Pre-Training (CLIP) (2021) -

Provides a neural network architecture designed to map large datasets of image-text pairs to allow it to associate text prompts with images. This essentially makes it so that the text of "A photo of an apple" and an actual photo of an apple are seen as similar to the network. Used by DALL-E 2 and 3 to perform image generation.

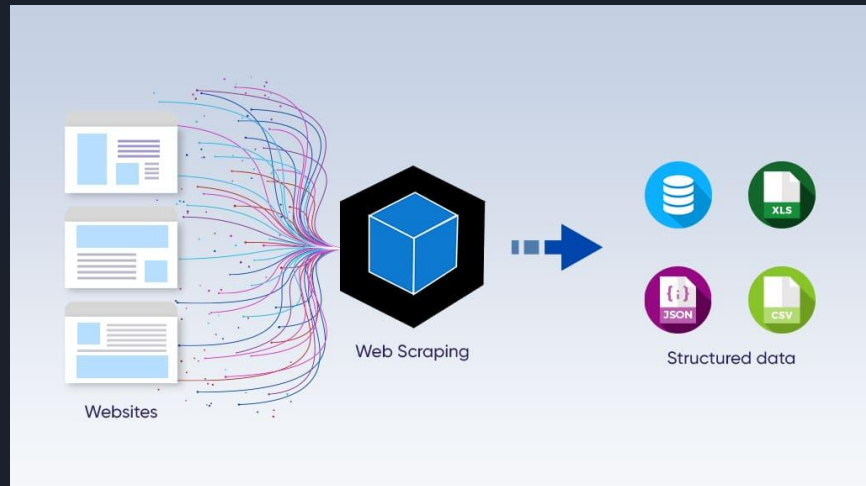


CLIP - Joint Training of Image and Text Encoders

Recent Techniques Continued

The current state of the art models are contrastively pretrained diffusion models trained on image-text pair datasets.

JFT-300M and LAION-5B (2022) - Image-text datasets used to train the models, they gather their images through web scraping and contain millions to billions of image-text pairs. The LAION-5B dataset is currently unavailable (since December 2023) after illegal images were found within the dataset.



AI Image Companies

All of the most recent models in use by these companies are contrastively pretrained diffusion models.

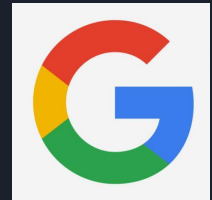
OpenAI - The makers of ChatGPT and DALL-E. They are responsible for the GPT series of LLMs as well as CLIP. Their most current image generator model is DALL-E 3.

Stability AI - The makers of Stable Diffusion and Dream Studio. Stable Diffusion is open-source and is trained on publicly available datasets.

Midjourney - The makers of the Midjourney service. Users connect to a discord server and ask a bot to generate their images from prompts.

Google - The makers of Bard and Gemini AI. Their most recent model is Imagen 2 which was only released for public use very recently.

And Many More...



Data Ethics





AI Copyright Lawsuits

- ★ These image generating models are built off of billions of images scraped from the web
- ★ In addition to inappropriate or illegal content, some of these images were copyrighted
- ★ **Andersen v. Stability AI Ltd.** - A lawsuit from October of 2023 against Stability AI, Midjourney, and DeviantArt accusing them of copyright infringement.
- ★ The plaintiffs of the lawsuit claimed that Stable Diffusion was trained on their artworks, which allows users to create images in the style of specific artists (*Andersen v. Stability AI Ltd.*, 2023)



AI Copyright Lawsuits Continued

- ★ Plaintiffs argue that being able to copy an artists style through AI image generators harms their livelihood
 - “Until now, when a purchaser seeks a new image “in the style” of a given artist, they must pay to commission or license an original image from that artist. Now, those purchasers can use the artist’s works contained in Stable Diffusion along with the artist’s name to generate new works in the artist’s style without compensating the artist” (*Andersen v. Stability AI Ltd.*, 2023)
- ★ Images created via image generators can’t be copyrighted, due to the lack of human authorship (Brittain), and so are public domain.
- ★ Public domain images can be used for commercial purposes by anyone. Not only could anyone create works in the style of an artist, and take away money from them that way, they could also sell these AI created artworks and make a profit from it
 - “The harm to artists is not hypothetical—works generated by AI Image Products “in the style” of a particular artist are already sold on the internet, siphoning commissions from the artists themselves” (*Andersen v. Stability AI Ltd.*, 2023)



AI Copyright - Dismissal

- ★ The lawsuit claimed that “Defendants, by and through the use of their AI Image Products, benefit commercially and profit richly from the use of copyrighted images.”
- ★ This case was dismissed, but some of the key claims were allowed to move forward
 - The claims not allowed to move forward were “copyright infringement, right of publicity, unfair competition and breach of contract claims against DeviantArt and Midjourney” (Cho, 2024), and the original allegations were said to be “defective in numerous respects” (*Andersen v. Stability AI Ltd.*, 2023)

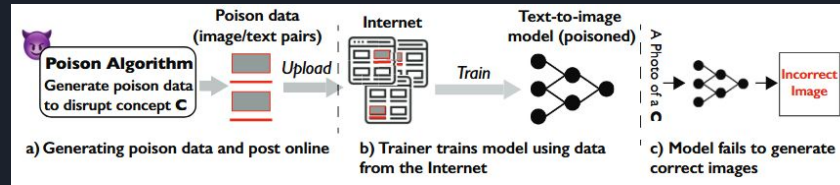


AI Copyright - Recent Developments

- ★ Feb 8th case denied DeviantArt's motion to strike the case for good
- ★ The case is currently about the use of the plaintiff's names or styles when the companies commercially promote their products (Cho, 2024)
- ★ DeviantArt's motion was denied because having ruling on this would be in the public interest of the artists of California
 - Specifically Californian artists because this motion was based on California's anti-SLAPP laws
 - Also, the plaintiffs have claimed, but have not been able to prove that their names/art styles were used to advertise Midjourney products. If their claims are true, they will not fall under the anti-SLAPP laws (Cho, 2024)

Recent Developments – Nightshade

- ★ Another recent development: the release of Nightshade
- ★ **Nightshade** - changes what AI image generators see whenever they look at art (e.g. a handbag instead of a cow)



- ★ Some companies such as Stable Diffusion have no 'opt-out' for artists
- ★ Some such as Dalle 3 do, and let artists opt-out from their work being used to train Dalle's future image generation models
 - Meant to "help deter model trainers who disregard copyrights, opt-out lists, and do-not-scrape/robots.txt directives" (*What is Nightshade?*, 2024)
- ★ Publicly released in late January, it got over 250,000 downloads in 5 days, showing that many artists are concerned about their art work being used as training data without their consent

Nightshade and Data Poisoning

- ★ While the original purpose of Nightshade was to protect artists, some worry it may be used maliciously instead



A Few Pros and Cons of AI Image Generators



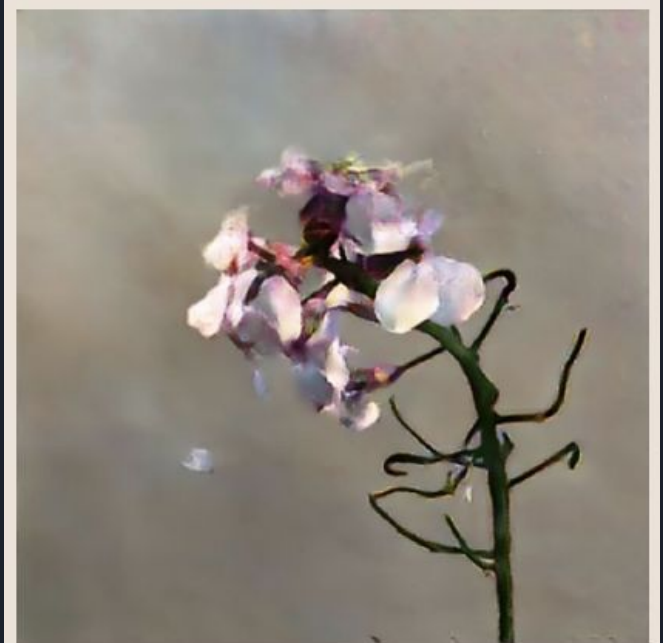


Cons

- ★ Taking money from artists
- ★ Some say the AI art is missing human emotion, and isn't 'real art', or creative
 - "However, though models have the ability to generate surprising outputs, they entirely lack the capacity to anchor these outputs in the world. Unlike the kind of creativity valued in humans both in and beyond the arts, ML has little scope for contextualisation" (Ploin et al., 2022)
- ★ Once people find out art was AI generated, they see it as less creative
 - "...people devalue art labeled as AI-made across a variety of dimensions, even when they report it is indistinguishable from human-made art, and even when they believe it was produced collaboratively with a human" (Horton et al., 2023)

Pros

- ★ The same study said that AI art could potentially help people appreciate human creativity more
- ★ AI image generation can help non-artists visualize their ideas, and provide inspiration to artists
- ★ Some artists are using smaller, self built image generation models trained off their own work to create art (Ploin et al., 2022)



Learning Nature (b38,4106,16)
(2018) by David Young



Conclusion

- ★ **Image generation models** - models that create images based on written prompts. They have come a long way in just over two decades
- ★ However, some large companies have issues with **data ethics**, and are **using copyrighted materials as training data**
- ★ Many artists are using **Nightshade** to discourage companies from using their art as training data.
- ★ There are a lot of pros and cons to AI image generators, and overall, **they can't be called inherently good or bad**. This presentation focused a lot on the negatives, due to our focus on data ethics, but AI image generation isn't all bad, and it's not our intention to say AI art is good or bad, or judge whether or not it's art at all!
- ★ Those are things *you* can hopefully start to decide for yourself, after our presentation!



Sources

- Andersen v. *Stability AI Ltd.*, 3:23-cv-00201-WHO (N.D. Cal. Jan. 13, 2023).
<https://findfx.thomsonreuters.com/gfx/legaldocs/myvmogidxvr/IP%20AI%20COPYRIGHT%20complaint.pdf>
- Andersen v. *Stability AI Ltd.*, 23-cv-00201-WHO (N.D. Cal. Oct. 30, 2023).
<https://casetext.com/case/andersen-v-stability-ai-ltd/case-details>.
- Asperti, A. (2023). Generative models and their latent space. *The Academic*.
<https://theacademic.com/generative-models-and-their-latent-space/>
- Brittain, B. (2024, January 23). *Computer scientist makes case for AI-generated copyrights in US appeal*. Reuters.
<https://www.reuters.com/legal/litigation/computer-scientist-makes-case-ai-generated-copyrights-us-appeal-2024-01-23/>
- Cho, W. (2024, February 9). *AI Companies Take Hit as Judge Says Artists Have “Public Interest” In Pursuing Lawsuits*. The Hollywood Reporter.
<https://www.hollywoodreporter.com/business/business-news/artist-lawsuit-ai-midjourney-art-1235821096/>
- DALL·E 3. Retrieved February 10, 2024, from
<https://web.archive.org/web/20240210101847/https://openai.com/dall-e-3>
- Despois, J. (2017). *Latent space visualization — Deep Learning bits #2*. Hackernoon.
<https://hackernoon.com/latent-space-visualization-deep-learning-bits-2-bd09a46920df>
- Efros, A.A., and Freeman, W.T. (2001). *Image quilting for texture synthesis and transfer*. Proceedings of the 28th annual conference on Computer graphics and interactive techniques (SIGGRAPH '01). Association for Computing Machinery, New York, NY, USA, 341–346. <https://doi.org/10.1145/383259.383296>
- Esser, P., Rombach, R., and Ommer, B.. (2020). *Taming Transformers for High-Resolution Image Synthesis*. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. <https://doi.org/10.48550/arXiv.2012.09841>
- Feller, W. (1957). “An introduction to probability theory and its applications”, Vol. 1, 3rd edition. pp. 217–218.



Sources Continued

- Franzen, C. (2024, January 23). *AI poisoning tool Nightshade received 250,000 downloads in 5 days: 'beyond anything we imagined'*. Venture Beat. <https://venturebeat.com/ai/ai-poisoning-tool-nightshade-received-250000-downloads-in-5-days-beyond-anyting-we-imagined/>
- Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, D., Bengio, Y. (2014). Generative Adversarial Networks. *arXiv e-prints*. <https://doi.org/10.48550/arXiv.1406.2661>
- Horton, C. B., Jr, White, M. W., & Iyengar, S. S. (2023). Bias against AI art can enhance perceptions of human creativity. *Scientific reports*, 13(1), 19001. <https://doi.org/10.1038/s41598-023-45202-3>
- Jiang, H.H., Brown, L., Cheng, J., Khan, M., Gupta, A., Workman, D., Hanna, A., Flowers, J., and Gebru, T. (2023). *AI Art and its Impact on Artists*. Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES '23). Association for Computing Machinery, New York, NY, USA, 363–374. <https://doi.org/10.1145/3600211.3604681>
- Kingma, D.P., and Welling, M. Submitted 2013. Revised 2022. *Auto-Encoding Variational Bayes*. <https://doi.org/10.48550/arXiv.1312.6114>
- Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012). *Image Net Classification with Deep Convolutional Neural Networks*. Advances in Neural Information Processing Systems, F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger (Eds.), Vol. 25. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
- Ploin, A., Eynon, R., Hjorth I. & Osborne, M.A. (2022). *AI and the Arts: How Machine Learning is Changing Artistic Work*. Report from the Creative Algorithmic Intelligence Research Project. Oxford Internet Institute, University of Oxford, UK. <https://www.oii.ox.ac.uk/news-events/reports/ai-the-arts/>
- Radford, A., Kim, J.W., Hallacy, C., Ramesh, D., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger G., and Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. *International Conference on Machine Learning*. <https://doi.org/10.48550/arXiv.2103.00020>



Sources Continued

- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., and Sutskever, I. (2021). Zero-Shot Text-to-Image Generation. *arXiv* <https://doi.org/10.48550/arXiv.2102.12092>
- Rocca, J., and Rocca, B. (2019). *Understanding Variational Autoencoders (VAEs)*. Towards Data Science. <https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73>
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2021). High-Resolution Image Synthesis with Latent Diffusion Models. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.48550/arXiv.2112.10752>
- Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, R., Katta, A., Mullis, C., Wortsman, M., Schramowski, P., Kundurthy, S., Crowson, K., Schmidt, L., Kaczmarczyk R., and Jitsev, J. (2022). LAION-5B: An open large-scale dataset for training next generation image-text models. *arXiv*. <https://doi.org/10.48550/arXiv.2210.08402>
- Shan, S., Ding, W., Passananti, J., Zheng, H., Zhao, B.Y. (2023). Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models. *arXiv*. <https://doi.org/10.48550/arXiv.2310.13828>
- Stable Diffusion Online*. Retrieved February 10, 2024, from <https://web.archive.org/web/20240210223457/https://stablediffusionweb.com/>
- Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. (2017). Attention is All you Need. *Neural Information Processing Systems*. <https://doi.org/10.48550/arXiv.1706.03762>
- What is Nightshade?* Retrieved February 10, 2024, from <https://web.archive.org/web/20240209154004/https://nightshade.cs.uchicago.edu/whatis.html>