

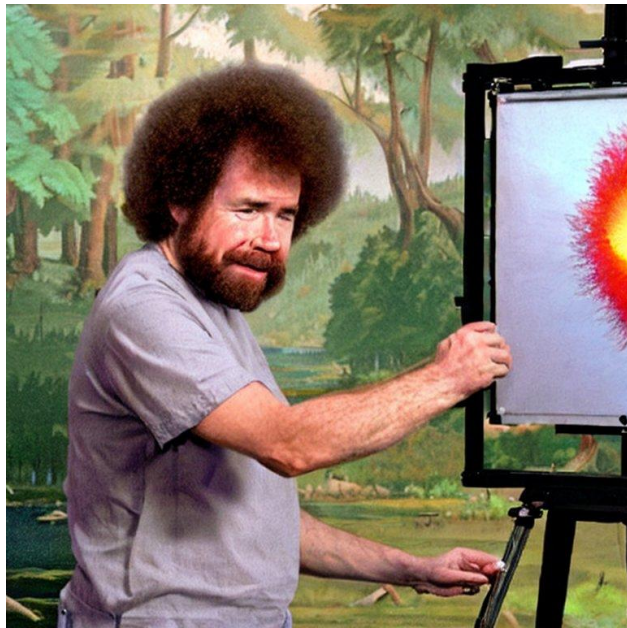
The Legality and Morality of AI in Creating Artwork

By Chase Griffith and Garrett Russell



Stable Diffusion AI

- Stable Diffusion AI is an open source text-to-image model.
- Of course, this isn't the only text-to-image AI model on the market.
 - OpenAI's DALL-E 2
 - Google's Imagen
- So, what sets Stable Diffusion apart from these two models and why is it so much more popular?
 - First and foremost, it is FREE and OPEN SOURCE.
 - Stable Diffusion uses the latent space of image rather than the pixel space, which the two models above use.





Small Tangent: Latent Space vs. Pixel Space

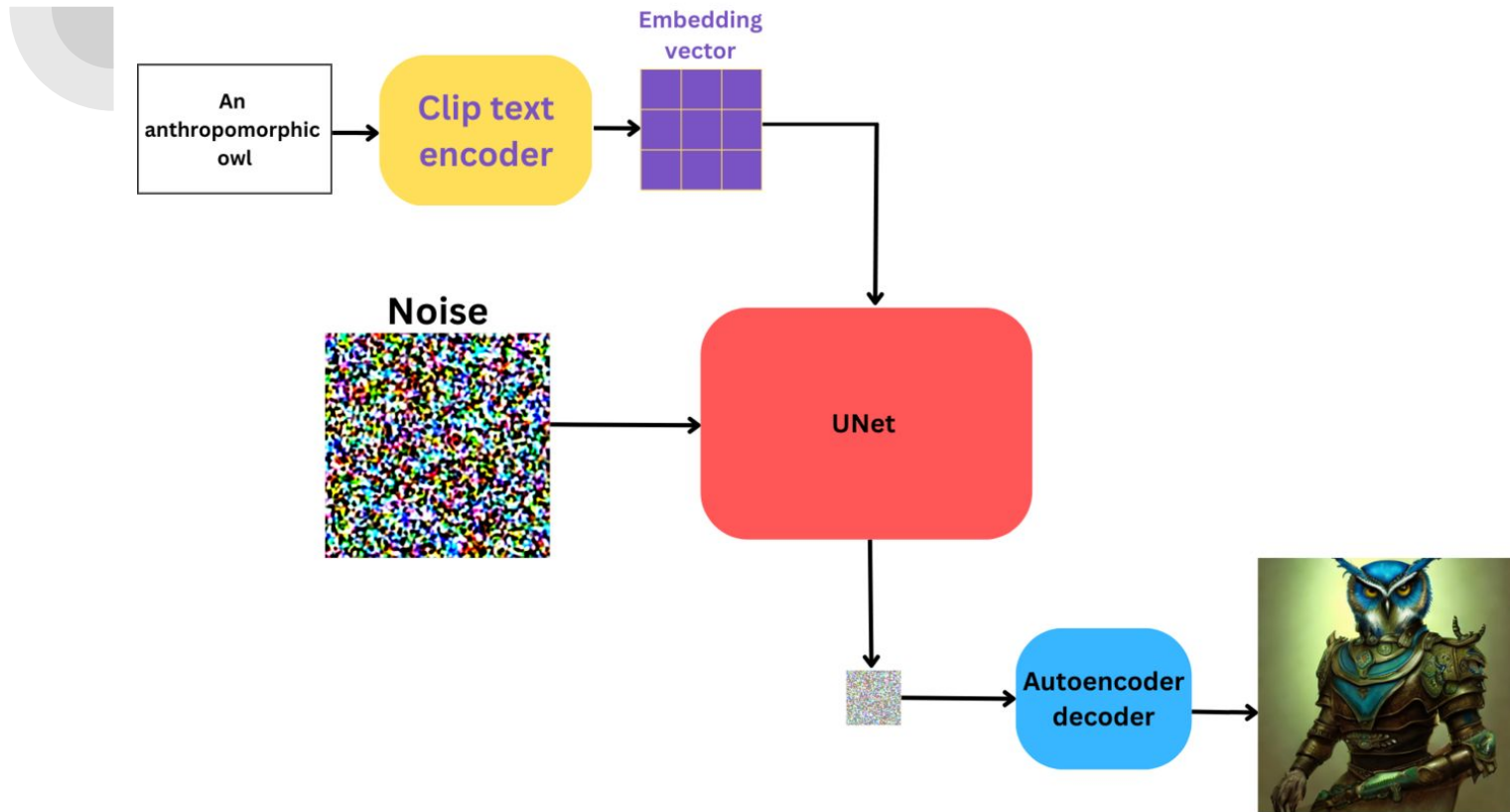
- Pixel space is how we imagine the vectors in high-dimensional Euclidean spaces.
- Changing between two images in pixel space involves moving the individual pixels to different spaces in a vector (top image).
- Changing images in latent space means that the model has learned the difference between the two images and can more efficiently move between the two images (bottom image).
- Latent space is something a ML model has to *learn*.





How Stable Diffusion Works

- The model is split into three sections:
 - A pre-trained text encoder
 - Takes the text and turns it into embeddings
 - A UNet noise predictor
 - Predicts noise
 - A variational autoencoder-decoder
 - The encoder figures out the latent space information based on the embeddings
 - The decoder predicts the image from latent space that was conditioned by the text prompt



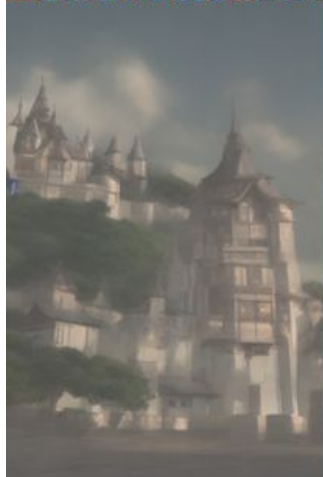
Steps: 1

Steps: 2

Steps: 3

Steps: 5

Steps: 8



Steps: 10

Steps: 15

Steps: 20

Steps: 30

Steps: 40

Possible Issues with Stable Diffusion

- It is possible to create images with stable diffusion that is in the art styles of famous artists.
 - *Using the original artists work as a reference*





Stable Diffusion's Dataset: LAION 5B

- LAION 5B is an open source data set that contains over 5 billion CLIP-filtered text and image pairs.
 - This is not the first dataset of this size, but it is the first to be openly available.
- All of the data from this dataset is scraped from publicly available images on the internet.

Backend url:

<https://knn5.laion>

Index:

laion_5B

french cat



[Clip retrieval](#) works by converting the text query to a CLIP embedding, then using that embedding to query a knn index of clip image embeddings

Display captions ☒

Display full

captions ☐

Display similarities



Safe mode ☒

Hide duplicate urls



Hide (near)

duplicate images ☒

Search over

image

Search with
multilingual clip



french cat



french cat



How to tell if your feline is french. He wears a b...



イケメン猫モデル
「トキ・ナンタケッ
ト」がかっこいい -
NAVER まとめ



Hilarious pics of funny cats! funnycatsgif.com



Hipster cat



網友挑戰「加幾筆畫
出最創意貓咪圖片」，
笑到岔氣之後我也手



cat in a suit Georgian
sells tomatoes



French Bread Cat Loaf
Metal Print



Issues With LAION 5B

- Unfortunately, not everything made public on the internet was put there legally or with consent from all parties.
- Example: one woman found her medical records were leaked onto the internet and were used in this dataset.
- Fun Fact! There's a website where you can look up whether or not a picture of you is in this [dataset!](#)



The Class Action against Stability AI, Midjourney, and DeviantArt

- Currently, there is a class action filed against the companies listed above for the following reasons
 - First, the companies above all use the LAION dataset, and as such use all of the stored artwork inside of it for their text-to-image based AI.
 - Second, creators of these images have expressed repeated complaints about their work being used without their permission. In some cases the models are used to replicate specific creators art styles and then sell the work.
 - Third, because the dataset stores a version of the original image inside of the dataset and every image returned is effectively a collection of artifacts of these original images.
 - Fourth, the AI models aren't inherently transformative as they are not creating any truly new image.



Google V. Authors Guild

- In 2003, Google began creating an online database of different books and literature
 - They acquired the literature by going to different libraries and checking out dozens of books at a time and scanning them for their database. They would then return those books and check out another set to begin the process again
 - The important bit is that they did this without consulting the authors of said literature to obtain permission to digitize their work.
 - This is a violation of copyright law on paper. However the court found that the digitization of the material and the implementation of a search system for their archives was transformative

What does that have to do with the Stability AI case?

- Stability AI's model or to be more specific, the LAION dataset, is doing a similar thing to what Google was doing with the libraries in 2003. They gathered images from different sources on the internet without a creator's permission or consent and placed them into the training and test dataset for their AI model.
- The difference between the two is that the dataset has not stopped gathering new information. Google stopped scanning literature after authors raised concerns, and allowed them to opt-in or out. LAION has not expressed the same courtesy.



The Class Action Potential Results

- The court might find that the work is transformative as it doesn't have to replicate the work of specific creators, and can create “new” work.
- The court might find that the dataset doesn't regurgitate copies from the dataset, but only uses them to train it and nothing else.
- The court might not find it to be transformative due to the fact that the prompt has to be based on some other work inside of the dataset, and as such the work is derivative.
- The court might also find issue with the images inside the dataset, and could order the LAION dataset to be pruned of all work made by non-consenting creators.



Our prediction

- We find it likely that like the Google V. Authors guild case, the court will find that the LAION dataset and Stable Diffusion model is guilty of copyright infringement, but instead of ruling it as a transformative work, it will rule that is purely derivative.
- The main key feature is that the AI model and the dataset aren't creating anything truly new using any meaningful heuristics by the user.
- In addition, there has been a court ruling separate from this class action: *works created by an AI model are not copyrightable in the U.S.*
- What this means is that AI images do not have human authorship and are not capable of being original works.



Why is this important?

- AI has been for many years, a tool used by computer and data scientists for research and development of new applications. AI models are now being given to the public for miscellaneous uses haphazardly.
- Stable Diffusion and Midjourney are examples of models given out without questioning any morality of using essentially stolen artwork to create a new product.
- Due to how these models were put out and marketed (ex. “Create a star wars version of yourself”), most people don’t think of these models as doing any real harm.



What could we wish we could do more of?

- We wish that we could've maybe taken a whack at potentially coding something very basic with a text-to-image model, but found that the effort and time available to us was best used to explain the issue as it relates to machine learning and AI as a field. Though ideally doing both would've been nice
- Also getting some anecdotes with some artists and computer scientists. Some were gotten but they were very much in a conversational setting and we don't have any transcripts for that, so it wouldn't work in this setting.