Pixels to Picasso: What's the Deal with Neural Art?

Neural art refers to visual art that is created at least partially by a neural network. While neural networks have gone in and out of favor among AI researchers over the last 80 years, neural art has recently surged in popularity due to advances that have enabled the easy generation of aesthetically pleasing and coherent visual outputs. We are at the precipice of paradigmatic change, both within the art community and in any domain that visual media touches. As we hurtle toward a strange new future for art and creativity, it is important for us to collectively shape it with a consciousness of the values being embedded in it. However, there is a lack of discourse in this area. This thesis is an attempt to introduce the concepts behind neural art and the concerns surrounding it, as a jumping-off point for increased understanding, discussion, and collaboration. Writing for a non-technical reader, I start with a primer on neural art. Some techniques are further developed in five personal works in the next chapter. This is followed by findings from a 43 participant survey regarding societal concerns about neural art, and a discussion of pertinent latest challenges. Based on the research, ten implications are discussed.

`https://github.com/oliviaseow/pixels2picasso`

By Olivia Seow, oliviaseow@gmail.com

May 11, 2022

Contents

This page intentionally left blank

Chapter 1

Introduction

We are neural nets. Anything we can do they can do.

— Geoffrey Hinton, WIRED Interview (May 2019)

The banner image above was created using a natural language prompt, "dendritic connections between neurons", on a generative neural network known as guided diffusion [25]. *Neural art*, defined in this work as visual art that is created at least partially by a neural network, started gaining momentum around seven years ago [9]. Research over the last year has been particularly significant, with exceptionally coherent and aesthetically pleasing visual outputs.

Neural networks are computational systems that are loosely modelled after the mammalian brain, which comprises a vast network of nerve cells, or neurons, each selectively firing in response to sensory stimuli [35]. This process contributes to our seemingly effortless capacity for pattern and image recognition, knowledge retention, and insightful cognition. Computational *deep learning* is analogously achieved by stacking layers of computational units that respond when presented with a specific pattern. During *training*, individual layers learn to activate so as to produce a favorable aggregate representation of the input, as evaluated by a *loss function*. Put differently, a neural network distills knowledge from input data using mathematical pattern derivation and matching. A simplified representation of a generative machine learning process can be found in Figure 1-1.

Neural networks are used in *machine learning* (ML), or non rule-based task per-

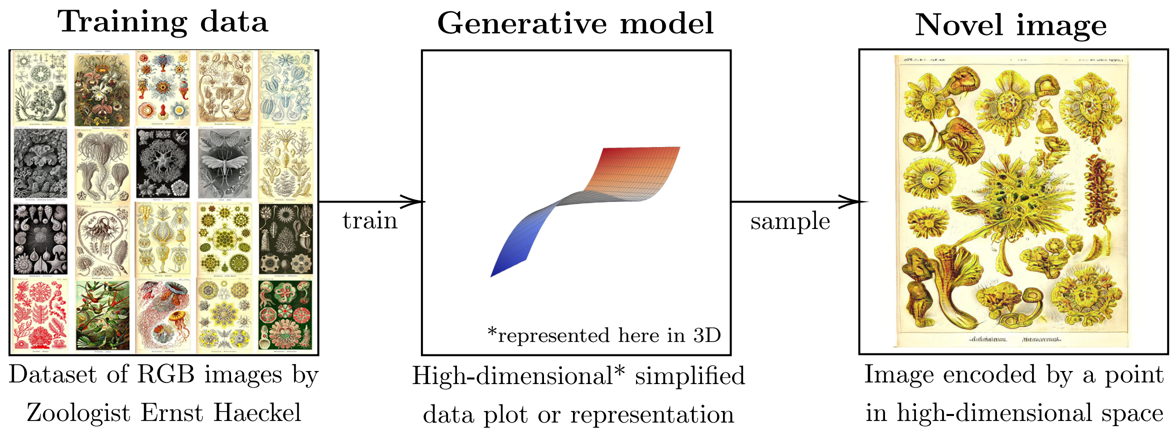


Figure 1-1: Simplified representation of a generative machine learning process.

formance by machines. This is under the broader umbrella of *artificial intelligence* (AI), where computers execute tasks commonly associated with intelligent beings. Given recent rapid advancements in the capabilities of generative neural networks, we are at the precipice of paradigmatic change, both within the art community and in any domain that visual media touches. Consider Kranzberg's first law, that technology is neither good nor bad; nor is it neutral [42]. AI research has been criticized for prioritizing shallow measures of progress without regard for the human context of these supposed advances [54]. As we hurtle toward a strange new future for art and creativity, it is important for us to collectively shape it with a consciousness of the values being embedded. This will create a livable world *by design*, versus scrambling to limit the harms of technologies that have proliferated [18]. However, preliminary observations in art and tech forums suggest that discourse on the values and impact of neural art research is limited. This may be attributable to the lack of diverse perspectives understanding and thus influencing AI research of all types, an issue that unfortunately affects the world with ever more immediate and detrimental implications [65]. To address this issue in neural art, we first need an increased understanding about the motivation, tools, and techniques for creating such art, and a framework for all to contribute to the discussion and thus research directions.

Past works that overview neural art techniques include Franceschelli and Musolesi's 'Creativity and Machine Learning: A Survey' [31] and Akten's 'Deep Visual Instruments: Realtime Continuous, Meaningful Human Control over Deep Neural Networks for Creative Expression' [20]. However, these are written for researchers and arguably inaccessible for a non-technical audience. Foster's 'Generative Deep Learning' [30] and Kogan's 'Machine Learning for Art' [49] are good for technical novices, but do not currently include latest developments in neural art. This is primarily limited to technical research papers or scattered ad hoc online.

Moreover, the works that cover technical concepts are largely detached from discussions about neural art's impact on society, such as those carried out in conference workshops like *Ethical Considerations in Creative Applications of Computer Vision* at CVPR, and *Computer Vision for Fashion, Art and Design* at ECCV and ICCV. Once again, these paid workshops are often inaccessible to those outside of research. A retrospective is offered in 'Ethics and Creativity in Computer Vision' by Rostamzadeh et al. [56], but the short brief covers only two themes, cultural appropriation and issues relating to training data.

Some topics that *are* heatedly debated are about philosophical definitions of art, aesthetics, and creativity, and whether computer art qualifies under those definitions. Works addressing these issues include 'The Artist in the Machine: the World of AI Powered Creativity' by Miller [47], 'Can Computers Create Art?' and 'Aesthetics of Neural Network Art' by Hertzmann [39, 38], 'Deep Else: A Critical Framework for AI Art' by Grba [37], and 'Work of Art in the Age of Mechanical Reproduction' by Benjamin [21]. In this paper, definitions are kept amorphous where specific interpretations are not foundational to the discussion.

This thesis introduces the concepts behind neural art and the concerns surrounding it, as a jumping-off point for increased understanding, discussion, and collaboration. Writing for a non-technical reader, I start with an overview of neural art techniques in the form of a question and answer primer. Next, five personal works related to neural art are described. These inform a survey conducted on societal concerns in neural art, which is discussed in the subsequent chapter. Pertinent latest challenges based on primary and secondary research in neural art are then presented, followed by ten implications for the future.

Chapter 2

Primer on Neural Art: Techniques & Resources

Without the aid of a computer, it would not possible to materialize quite so faithfully an image that previously existed only in the artist's mind. This may sound paradoxical, but the machine, which is thought to be cold and inhuman, can help to realize what is most subjective, unattainable, and profound in a human being.

— Vera Molnár, Interruptions (SIGGRAPH 1998)

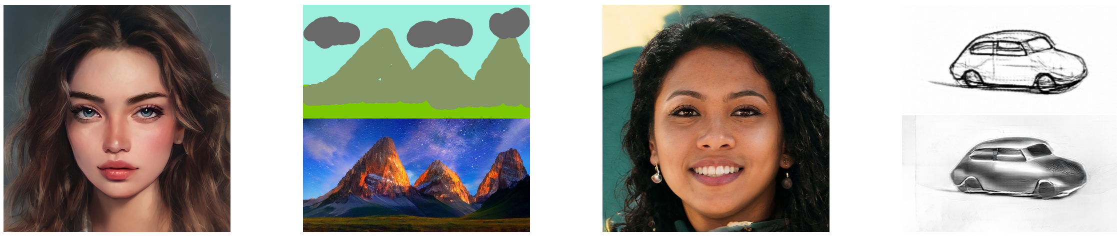
We begin with background, techniques, and resources about neural art. This primer is not meant to be comprehensive, and is written for tech novices or those who are not already engaged with the neural art scene. It is best consumed digitally, as there are many hyperlinks to referenced resources for deeper exploration. As the field is rapidly expanding, changes are to be expected. The most updated version of this primer will be available at `github.com/oliviaseow/pixels2picasso`.

2.1 Who makes art with neural networks?

The reach of neural art is vast. It can benefit anyone who might want to visualize an idea or express themselves in more ways, and it has been incorporated into workflows by creatives in many disciplines including architects [23], illustrators [10], large-scale installation artists [16], and painters [14]. An AI artist directory is on `aiartists.org`.

2.2 What do I need to get started?

For starters, neural art can be created on a web browser via *no code* tools. Some examples are shown in Figure 2-1.



(a) ArtBreeder, make art using "genes"

(b) GauGAN2, paint to photorealism

(c) This X Does Not Exist, generate synthetic data

(d) VIZCOM, sketch to render

Figure 2-1: Neural art from no code services.

There have also been many new web and mobile based tools in the space of text-to-image, language interface-based neural art (e.g., Artflow, CogView, Midjourney[1], Geniverse, NeuralBlender, NightCafé, ProsePainter, ruDALL-E, Snowpixel, StarryAI, WOMBO Dream).

While seemingly targeted at developers, some services that offer free and user friendly interfaces to test neural art and other machine learning models without code include: Hugging Face Spaces, Streamlit, and Pixray (Replicate)

RunwayML has a graphical user interface for people with some understanding of technical machine learning concepts, but the startup recently shifted its focus to real-time video editing. Alternatives include Pollinations and Playform.

For greatest flexibility, some programming literacy is needed. Section 4.1 shows that many neural artists make small edits to existing code. Javascript can be used for deep learning on the web, with accessible libraries like ml5js and TensorFlow.js. Python is the most common machine learning language, with popular deep learning libraries like PyTorch and Tensorflow, and Google's friendly online interface, Colab. Shared Colab notebooks have been collated by `dvschultz` and `pharmapsychotic`.

While free users of Colab *can* access GPU instances, paid accounts (offered in some countries) can get more powerful GPUs (e.g., Tesla V100 or A100 (better)). ML code

[1]In closed beta as of May 8, 2022.

can also be run in a local environment, or deployed on cloud service providers like Paperspace, Vast.ai, and Amazon SageMaker.

Neural art filters are already common on social camera platforms like TikTok and Snapchat [17]. With time, we can expect more neural art functionality to be integrated in digital tools like Adobe Photoshop [12].

Importantly, commercial services mentioned in this section may impose additional restrictions on use of neural art outputs, and may also use inputs for profit.

2.3 What data does neural art use?

Figure 1-1 shows how the input to a generative model is data. Massive online datasets like ImageNet (14 million images annotated by 25,000 workers), WikiArt (53,000 artworks labelled with artist and genre), and LAION-5B (5 billion image-text pair dataset) are often used. Some, like Ridler and Basanta, make their own datasets, while others scrape the web or use crowdsourced data. Attribution for neural art is often murky and complex, as will be discussed in Section 4.2.2.

2.4 How do neural networks process data?

Neural networks are built on basic building blocks known as perceptrons (a.k.a. neurons), which map input value to output value [48]. Training a neural network allows individual perceptrons to learn these associations. Perceptrons work in tandem with other perceptrons to perform tasks such as pattern matching. For example, a perceptron may be activated and output a high value if it determines that its input corresponds to a vertical line, and pass this information on to connected perceptrons. An interactive neural network training experience is on TensorFlow's Playground.

Depending on the format of the input and the goal of the network, different architectures (mathematical processing steps and connections between perceptrons) are used. For sequential data, such as vector style line sketches, we usually need architectures that have temporal memory through backward connections. The most basic of such models are known as Recurrent Neural Network (RNNs) [58]. An application pertaining to sketch generation via RNN is discussed in Section 3.1.

For the more common bitmap images (grids of pixels with RGB values), inputs can be processed with convolution layers, which extract informative spatial cues like filters do. For example, in Figure 2-2, convolution operations are performed on an image to locate edges. Neural networks with such operations are known as Convolutional Neural Networks (CNNs) [52]. A peak inside a CNN architecture used to classify input images is shown in Figure 2-3 [69].

(a) Left edge **(b)** Right edge

Figure 2-2: 'Girl with a Pearl Earring' portrait after edge detecting convolutions.

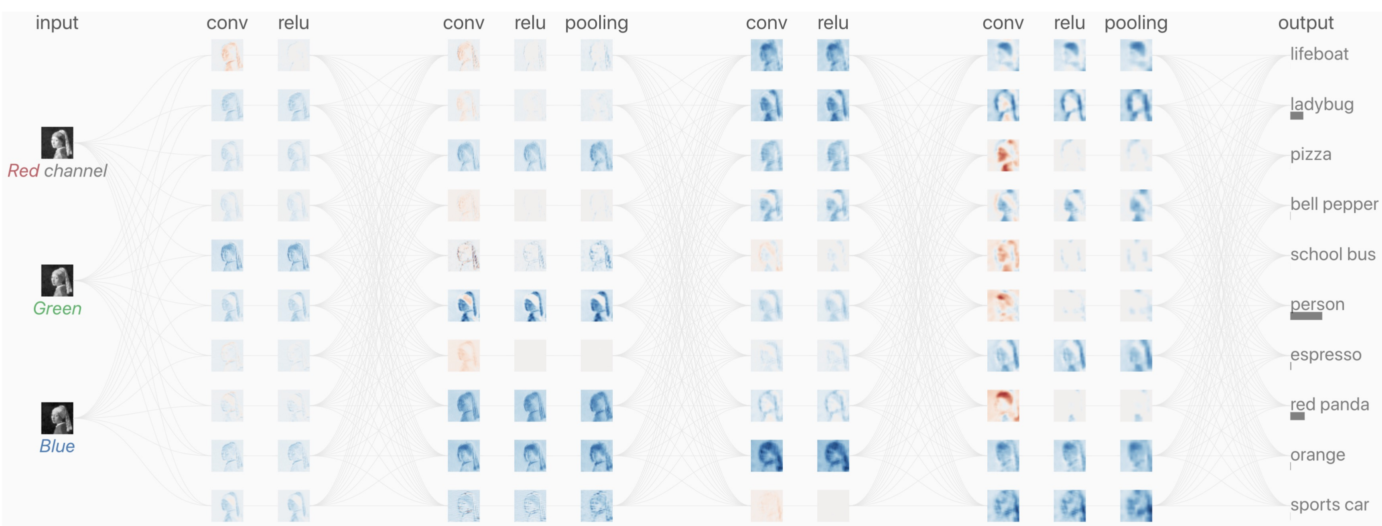


Figure 2-3: Visualization of perceptron activation maps in a CNN image classifier. Convolution layers are labelled `conv`. The `relu` layers zero negative values to learn more complexity, and `pooling` layers reduce input sizes to speed up training. The output is the probability that the input image belongs to each pre-determined class.

RNNs and CNNs are basic architectures that are often used in many generative neural networks. Common generative models include variational autoencoders (VAEs), generative adversarial networks (GANs), and diffusion models [11].

VAEs map input data into a distribution for sampling from. As shown in Figure 2-4, it has a bow-tie shape, where the left side encodes ("compress") input data into a distribution (shown in the middle) for decoding ("decompress") from, to the right. VAEs are fast and expressive, but do not produce high quality samples due to the encoding process [70]. A simple notebook for generating images with a VAE is here [4].

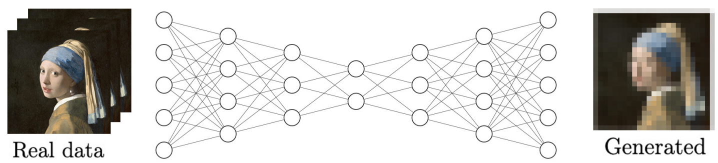


Figure 2-4: Simplified representation of a VAE architecture.

The GAN architecture contains a generator and a discriminator network. The generator takes in noise and creates an image from it, while the discriminator takes in an image and tries to determine if it is real or fake [36]. A simplified representation is shown in Figure 2-5. GANs are able to generate high quality images and can be sampled quickly, but they may lack output diversity since the goal of the network is to produce "real" images, not necessarily diverse ones [70]. GANs are also difficult to train in a stable way [43]. A notebook for generating images with a GAN is here [4].

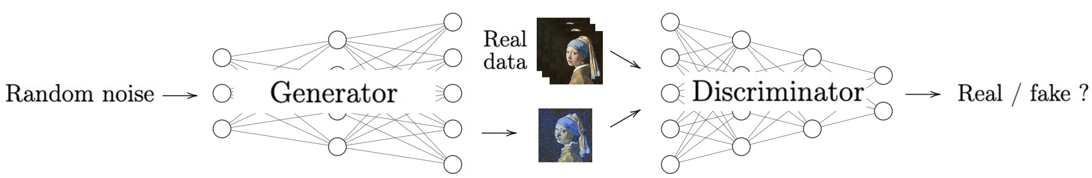


Figure 2-5: Simplified representation of a GAN architecture.

Diffusion reverses the process taken by neural networks to gradually transform data into noise, as shown in Figure 2-6. At each step, a different neural network is used. Diffusion models have better quality results than GANs and have good diversity, but due to the many network evaluations involved, produce samples slowly [70].

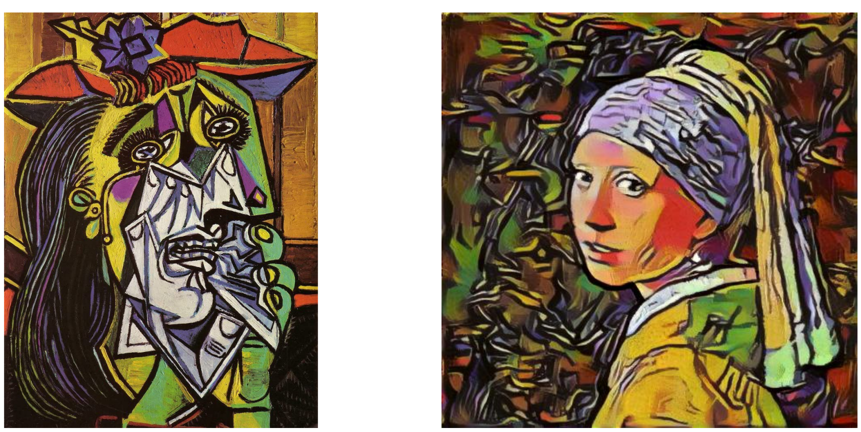
Figure 2-6: Noising and de-noising concept in diffusion models.

2.5 What types of neural art are there?

Neural art builds off the momentum of other generative and digital art created over the years [27]. A decade ago, artists were creating gesturally-controlled animations, supported by interfaces like Fiebrink's Wekinator created in 2009. A case study involving tutorials for gesturally-controlled animations is highlighted in Section 3.6. In 2015, Google Research released DeepDream [8], which uses CNNs to find and enhance patterns in images, creating a psychedelic appearance (see Figure 2-7). DeepDream was used by artists like Akten. In the same year, neural style transfer research was published, allowing creators to apply the style from one image to another, as shown in Figure 2-8. An application that uses neural style transfer is discussed in Section 3.3.

(a) Original by Johannes Vermeer **(b)** Processed with DeepDream

Figure 2-7: 'Girl with a Pearl Earring' portrait processed using DeepDream.



(a) Style image by Picasso **(b)** Style transfer output

Figure 2-8: 'Girl with a Pearl Earring' portrait processed using neural style transfer.

Networks like Generative Adversarial Networks (GANs) enable latent space interpolations and inversions, supporting morphing from one generated image to another, as shown in Figure 2-9. This is used by artists like Crespo, Klingemann, and Salavon.

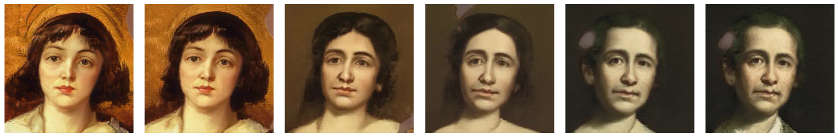


Figure 2-9: Interpolated images between GAN generated artwork

More recently, many models for language interface neural art have surfaced, allowing the generation of neural art from an input text prompt. Figure 2-10 shows a generated output from a number of models using the input phrase `a sea otter in the style of 'Girl with a Pearl Earring' by Johannes Vermeer`.

(a) CLIP + VQGAN

(b) CLIP + Diffusion

(c) Midjourney

(d) DALL·E 2 [51]

Figure 2-10: Generated outputs from four language interface neural art models.

Other prevalent kinds of neural art (e.g., Pix2Pix, CycleGAN, and neural cellular automata) were omitted for brevity. Neural networks can also be used to create animations [44, 59] and 3D objects [46] (e.g., Section 3.4). In addition, neural outputs may be pre- and post-processed to enable more interaction modalities and improved aesthetics. Some such applications are described in Chapter 3.

2.6 What are language interface neural art models?

In February 2021, OpenAI published Contrastive Language-Image Pretraining (CLIP), a neural network that learns visual concepts from natural language supervision [53]. CLIP was trained on over 400 million image-text pairs. As shown in Figure 2-11, CLIP creates an image and text embedding that associates images with descriptions.

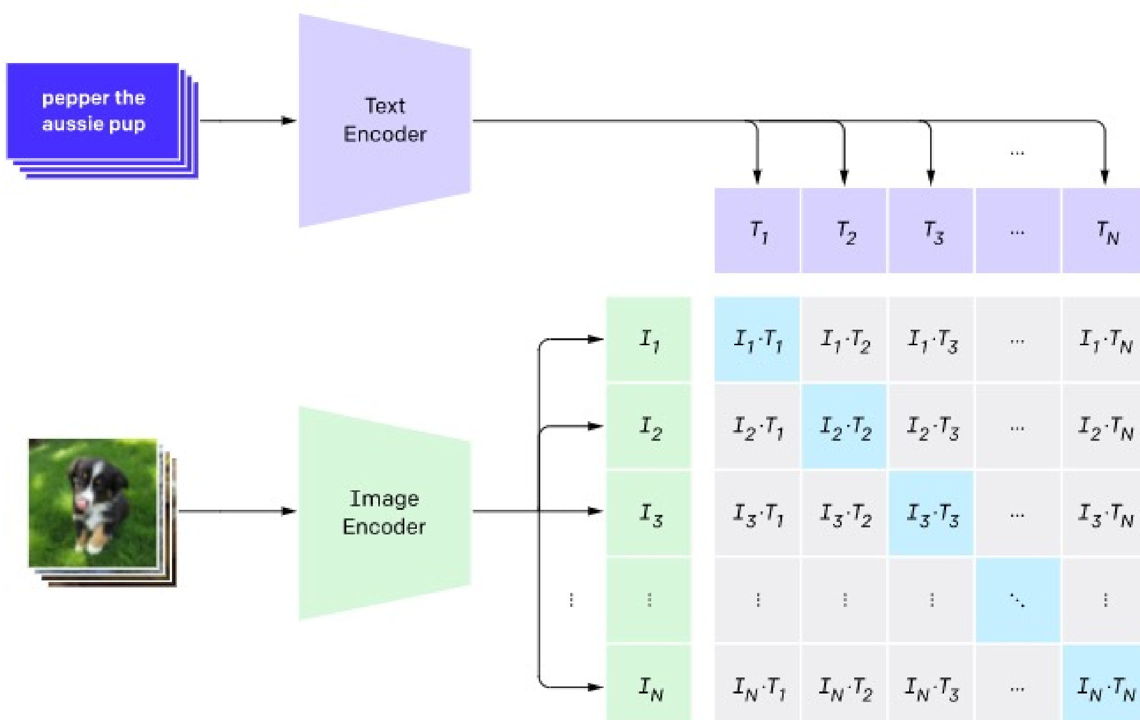


Figure 2-11: CLIP pre-training diagram reproduced from `openai.com/blog/clip`

CLIP and successors like CLOOB [32] and unCLIP [55] enable language interface neural art models, where natural language input is used to generate novel visuals that are closely associated in the learned embedding.

A list of code notebooks for language interface neural art are collated by `kaliyuga` in AI Colab Notebooks. As illustrated in Figure 2-10, different models have varying capacities for generating congruent visuals such as coherent human or animal figures. While there is also randomness based on a seed value, results can be tuned by changing the settings or rephrasing the input prompt. Ablation studies, such as that shown in Figure 2-12, can visualize how settings impact outcomes. As for writing prompts, using artist's names (e.g. "by Picasso") can invoke their styles (though notably, artists like Picasso did not have *one* style). Lists of artists whose styles have been learned by CLIP have been collated in Artist Studies by `remi_durant` and AiArtCreation's Volcano Comparison. Keywords can also be used to create visuals in a particular medium (e.g. "pencil drawing", "macro photography"). There are many tactics that help or hinder art creation, as compiled in CLIP+VQGAN keyword comparison by `kingdomakrillic`, and Disco Diffusion Modifiers by `KyrickYoung` and `sureailabs`.

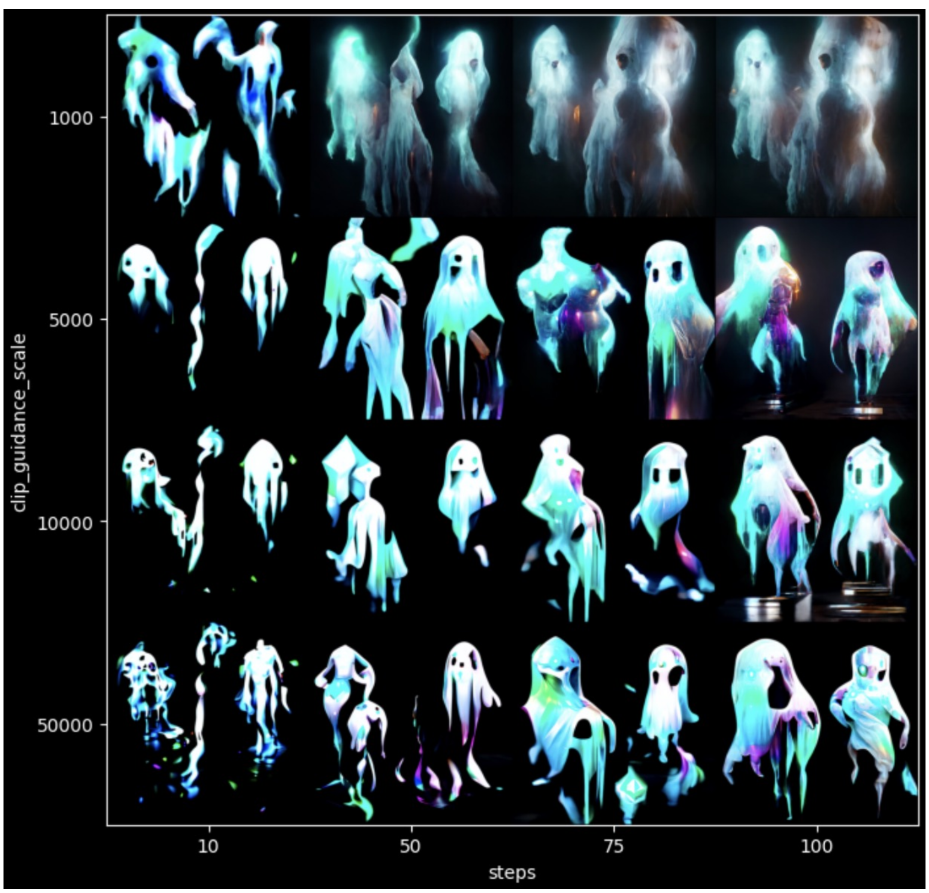


Figure 2-12: Ablation study for JAX model, reproduced from `EZ Charts`

One distinct feature about DALL·E 2 is its ability to generate variations on an image, such as the variations to 'Girl with a Pearl Earring' as shown in Figure 2-13.

Figure 2-13: Collage of variations to 'Girl with a Pearl Earring' by DALL·E 2 [5].

For a given image generated from a language interface, one can reverse engineer related prompts using a CLIP encoding retriever, such as LAION's interface.

2.7 Where is the neural art community?

Neural art online communities can be found on Discord [2], Facebook (e.g., AI Generated Art Group), Reddit (e.g., `r/bigsleep`, `r/aiArt`), and Twitter (e.g., Ai Art Community). Neural art is exhibited and discussed at venues like ACM SIGGRAPH, Art-AI Festival, Ars Electronica, ARTECHOUSE, Computer Vision Art Gallery (CVPR, ECCV, and ICCV), Creative AI Lab, Eyeo Festival, Gray Area Festival, International Conference on Design Computing and Cognition, International Conferences on Computational Creativity, MLxART, NeurIPS AI Art Gallery, NVIDIA AI Art Gallery, Sónar+D, TOCA ME, and xCoAx. In addition, there are AI Art residencies like Google's Artists + Machine Intelligence program.

2.8 What's next?

In this chapter, I have provided a primer on techniques and resources for neural art. The personal work presented in the next chapter apply such techniques. The findings inform the ensuing chapters on concerns and challenges associated with neural art.

This page intentionally left blank

Chapter 3

Case Studies

What I cannot create, I do not understand.

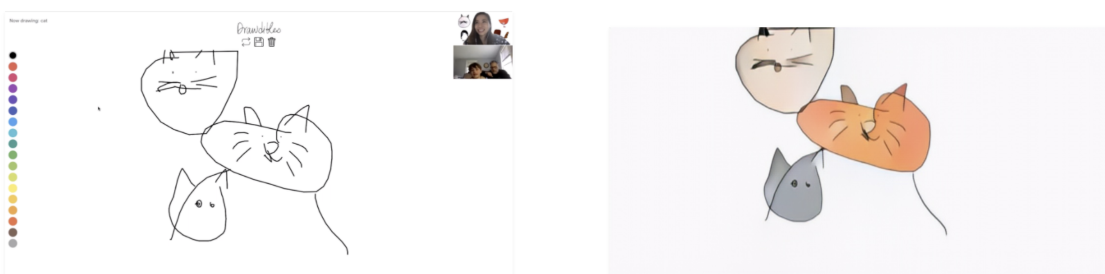
— Richard Feynman, 1988

This chapter documents personal work related to creative applications of neural art. Each tool, experience, or platform invites participation in and understanding about neural art. A discussion of key findings is provided in each case study.

1. Drawdibles (2019) is a creative drawing tool that listens to what a child wants to create and collaborates on neural sketches. Takeaways inform the use of neural art for building collaborative creative expression tools for children.
2. MetAI (2021) is a scrolly-telling interactive data visualization that explains machine vision and latent space ideas. One insight is that explainable visualizations are effective for building intuition and involvement in neural art.
3. Visual Poetry (2021) is a web-based tool that turns poetry into visuals using AttnGAN. Users can also apply evocative stylizations using neural style transfer. Findings suggest that combining neural techniques increases expressiveness, and that fine control is important for extended use of a tool.
4. Machine Pigment (2022) is an immersive art experience using CLIP-guided text2mesh. Findings point toward research opportunities for 3D neural art.
5. TheCAT is a tutorial series and experimental community for those keen on creative applications of technology. This is a stepping stone for increased participation and collaboration in neural art.

3.1 Drawdibles

Drawdibles is a creative drawing tool that listens to what a child wants to create and completes the drawing for them, allowing them to build on their creativity and tell imaginative stories. As a web platform, Drawdibles sketching works on any laptop, tablet, or mobile phone connected to the Internet. Colorization is a separate process which currently requires a Python 3 environment with CUDA. First a child says what they want to draw, e.g. "I want to draw a cat". A cat sketch-completion model is automatically loaded. The child can begin to sketch, and Drawdibles will complete the sketch for them in infinitely different cat-like ways. The cat meows when it is completed to create a symmetric multi-modal experience. The child can save their drawing and have it processed into a colored work of art. One example is shown in Figure 3-1. A user study was conducted virtually as this work occurred during the Covid-19 pandemic. The work can be accessed here: `github.com/oliviaseow/drawdibles`.



(a) Sketches **(b)** Colorized

Figure 3-1: Sketches and corresponding colorized image from user study.

3.1.1 Motivation

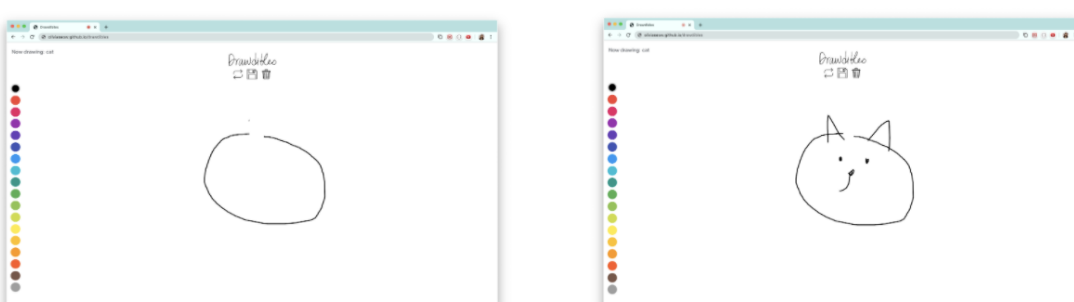
Young children tend to love to draw. Drawing is a great medium for them to tell stories. Through my research, I found that parents are always looking for ways to help creativity flourish, through means for children to learn and create more each day. Sometimes, a child may be hindered by not being able to express what they have in their heads on paper. In other cases, some may stick to drawing only things they are familiar with. Drawdibles alleviates these frustrations by listening to what a child wants to create, and working collaboratively with them to complete drawings and to tell stories that build on their limitless imagination. Unlike its contemporaries (e.g., NVIDIA GauGAN, Paper Dream), Drawdibles is available outside the research environment, and is a conversational tool that works collaboratively with children to doodle novel scenes and build their creative imagination.

3.1.2 System

Upon launching the web app, the child is shown an empty canvas with a color palette on the left. On the top left corner, the current model is displayed (see Figure 3-2).

Figure 3-2: Drawdibles landing screen.

What the child says (e.g. "I want to draw a cat") is transcribed at the bottom of the screen (see Figure 2). They can start to draw (Figure 3-3a), and the sketch is completed for them based on the model that has been loaded (Figure 3-3b).



(a) Sketch beings through user input

(b) Sketch completed by Drawdibles

Figure 3-3: Example of Drawdibles drawing input and sketch-completion process.

When select animals have been drawn (e.g. cat, dog, bird, duck), Drawdibles will play the corresponding animal noise. This creates a symmetric multimodal experience. When the child is done, they can save their work (middle icon below the Drawdibles logo) and have it processed into a colored neural art (Figure 3-4).

Figure 3-4: Example neural art from Drawdibles.

Figure 3-5 shows an architectural overview of the sketching system. Speech inputs are transcribed by Google's Web Speech API, and the system searches for lexical matches to models available in the Quick! Draw dataset [15]. When the p5.js web canvas detects that a stroke has been completed, the sketchRNN tensorflow.js model is used to supply finishing strokes to complete an image. Thereafter, audio (e.g., "woof", "meow", "chirp", "quack") from Adobe's sound effect database is played to the viewer. At present, not all models have an associated sound effect.

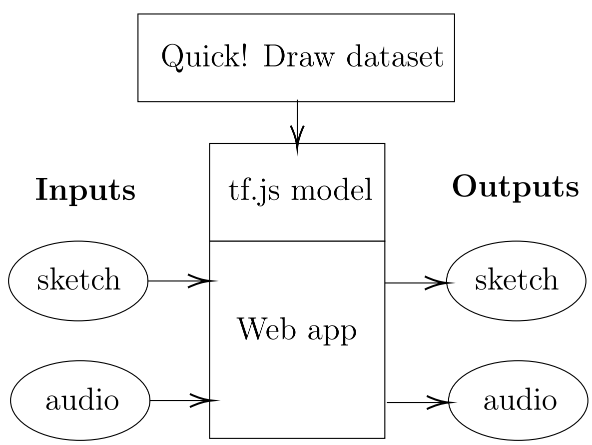


Figure 3-5: Overview of sketch system.

Figure 3-6 shows an architectural overview of the colorizing system. A VGG-19 convolutional neural network trained on Danbooru2019 [61, 6] is used to provide global style. Example outputs are shown in Figure 3-7.

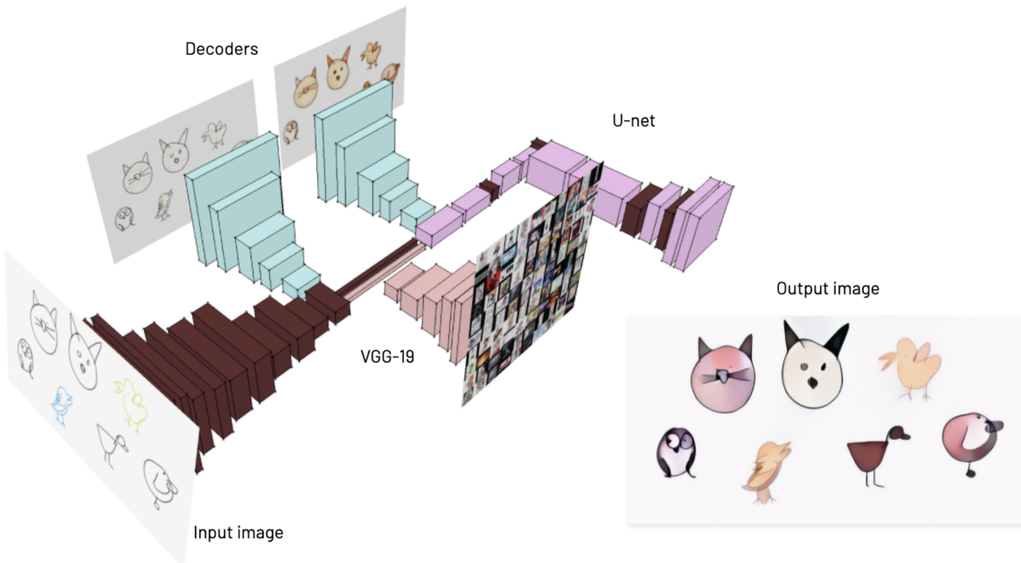


Figure 3-6: Overview of colorizing system.

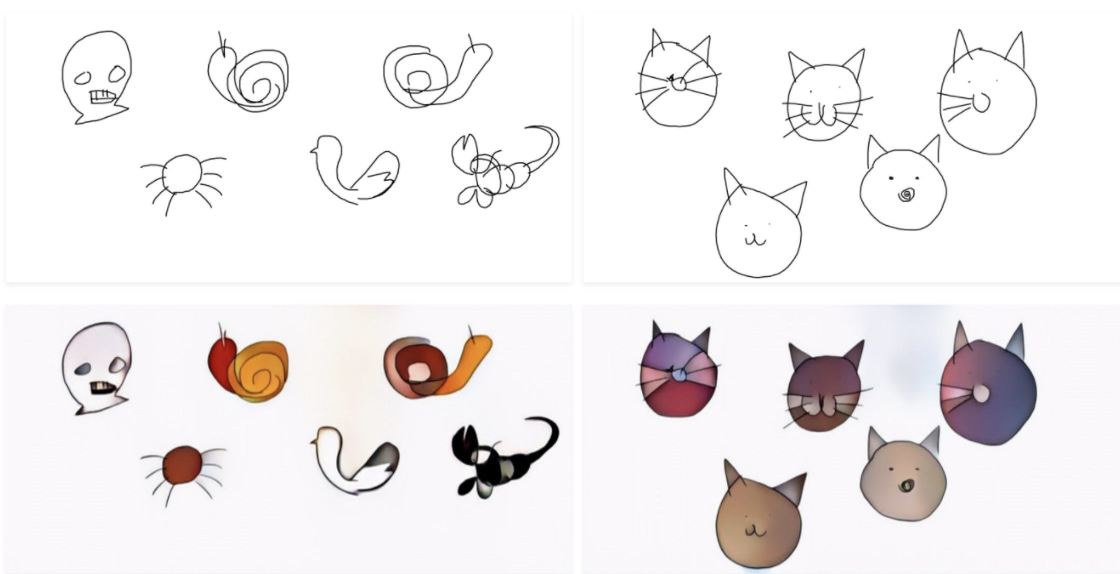


Figure 3-7: Drawdibles sketch colorization examples.

3.1.3 Discussion

Over two hours of user testing was virutally conducted with two children and their two parents. Overall, users quickly understood how to use Drawdibles and it was well-received. From this study, we derive considerations for building neural art based collaborative and creative expression tools for children.

— Children were engrossed with Drawdibles and naturally began telling stories with it. They spoke conversationally to the system, e.g. "how about a *dog*?", "this one is too small", suggesting that conversational interactions can be effective for neural art interfaces.

— Due to the quality of the original Quick! Draw dataset, not all outputs looked coherent. Despite this, children built on outputs to tell imaginative stories while sketching, demonstrating and developing their creativity. Parents remarked that the scrappy quality of drawings was actually preferred, as imperfection encourages experimentation and learning.

— Drawdibles facilitated conversational turns with parents, which studies have shown to be critical for child development [26]. It also organically induced a game between children to see who could create a better drawing or *exquisite corpse*. This suggests that neural art can foster communication and collaboration, a finding supported by other research that suggests AI is social glue [63].

— Children play by pushing the limits of the model and testing what is possible. They used toilet humor, and repetitively attempted to generate things like trains, dinosaurs, and airplanes, which were missing from the original dataset and hence could not be produced. This points at the importance of having ways for users to safely test and understand the capabilities of a neural art model.

— Voice input proves to be highly effective for young users. However, when two users were in disagreement about what to create, this descended into both shouting over which model they wanted to draw. To overcome this, future work could include a speech input toggle to selectively disable audio input.

3.2 MetAI

A primary barrier to using artificial intelligence is a proper understanding of how such systems work. To confront this, explainable visualizations are a scaffolding for providing transparency to viewers. In order to present this equitably, explainable visualizations should balance completeness and interpretability. The purpose of MetAI is to demonstrate the use of interactive data visualization techniques in explaining the power of generative neural networks to the average individual. The site can be accessed at `oliviaseow.github.io/metAI`. In May 2021, this work was documented in a paper written with John Liu and Dave Ludgin.

Figure 3-8: Screenshot of the MetAI landing screen.

3.2.1 Dataset

The dataset used for this visualization project was originally created for a paper on "Training Generative Adversarial Networks with Limited Data" [40]. It consists of 1,336 high-quality portrait PNG images at 1024×1024 resolution downloaded from the Metropolitan Museum of Art Collection API and automatically aligned and cropped using dlib (see Figure 3-10).

Figure 3-9: Sampling of images from original dataset.

3.2.2 "What are Portrait?"

The opening section, "What are portrait?", is meant to establish a conversational tone for the computer narrator. The subtitle "A machine's learnings from 1,336 portraits at the Met" is meant to personify the AI and this dialogue between the viewer is carried forward in each section. The backdrop of vibrant portraits introduces the viewer to the dataset that will be explored throughout the narrative. For increased visual interest, this backdrop is reshuffled every time the page is refreshed.

In the second section, the portraits transition into place on a scatter plot, conveying without words how the portraits may be analyzed based on their dominant colors. The animation is critical as an in-depth explanation about how the RGB values are created are not mentioned for simplicity. On the backend, this transformation was achieved by running k-means on the pixels and returning the centroid of the largest cluster. To speed up processing, the 1024×1024 images were first downsampled to 25×25. A 3D plot (shown in Figure 3-10 below) was created in python to investigate the result and determine interesting views that fit into the central narrative.

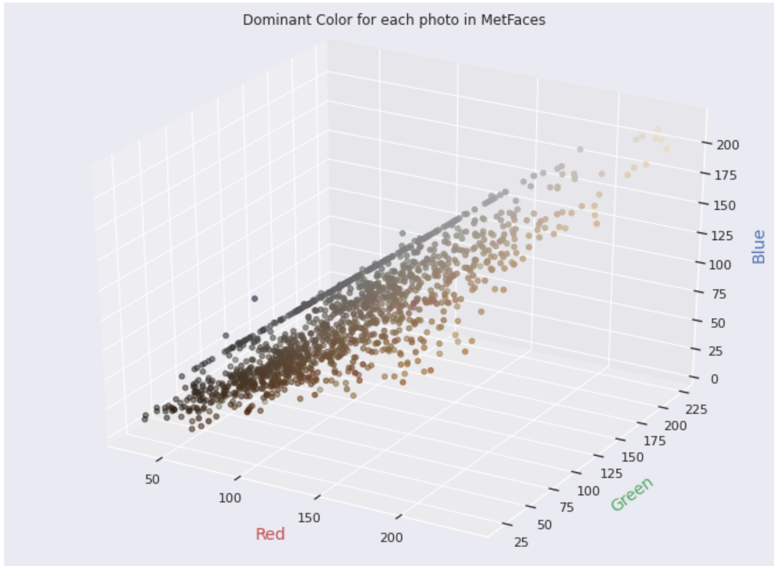


Figure 3-10: Sampling of images from original dataset.

3.2.3 "Pixels, Pixels, Pixels"

The explanatory text boxes provide the viewer with an understanding for how a computer sees portraits and these dominant color pixels as defined color values. The scatter plot is the most appropriate and effective in conveying the varying degrees of dominant color pixels in each portrait and highlighting early clusters of these RGB values. Exploration is facilitated by the tooltip, which is introduced early on so the viewer may discover specific portraits and artist names. Furthermore, the visual encoding of the average color in a portrait for each point bolsters the concept without complex explanations.

3.2.4 "Same Data, New Perspective"

The next animation is initiated by the viewer scrolling down and changing the y-axis color value from Green to Blue. The subtle dispersion of data points is meant to provide visual guidance as the text explains the variability of dominant colors within this dataset. The hover tooltip interaction continues to persist. While the first section may have sufficed for some viewers to understand how computers see images, the additional visualization serves as a check and reinforcement.

3.2.5 “Moving up from Pixel Values”

As the viewer moves through the first third of the narrative, a more dramatic transition takes place with the visualization hiding the scatter plot and reconfiguring data points into a map floating in white space. The change in form is meant to showcase new image clusters based on features and shapes rather than on a single color metric. Additionally, the dramatic animation serves as a pivot point in the story. We are taking a step up in understanding here by moving from seeing to interpreting that information. This creates a moment of friction causing a viewer to slow down and be more thoughtful about what they are viewing. On the backend, this is accomplished by performing a Uniform Manifold Approximation and Projection (UMAP) projection onto a two-dimensional space, illustrated in Figure 3-11.

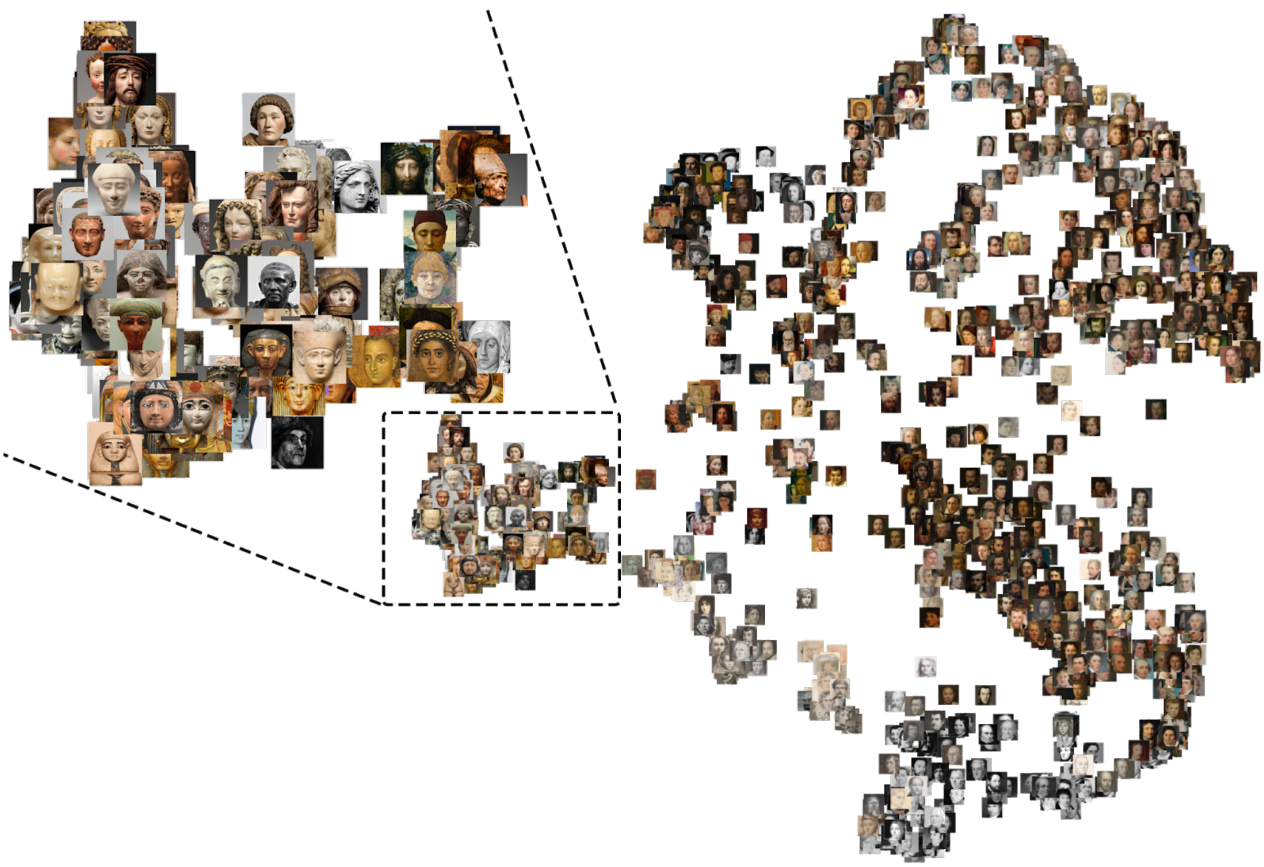


Figure 3-11: UMAP encoding of dataset, with one cluster magnified to show details.

UMAP is a dimensionality reduction algorithm that clusters multi-dimensional data onto a 2D space. It transforms the distance between two points into a uniform manifold variable for clustering. For this visualization, I chose to create ten groups. While this method does not directly feed into the Generative Adversarial Network (GAN) process described in the later sections of the scrollytelling, it is a helpful reference for viewers to contemplate how a machine learning model reduces high dimensional data in order to perform predictions. To demonstrate the arbitrary nature of the visual space in this encoding, the data is also rearranged as a grid in the following section.

3.2.6 "Clustering by Likeness"

At this point in the narrative, the viewer has become familiar with the layout of the scrolly-telling format. The text conveys that the new grid view is meant to aid human understanding, and is not critical for the computer. In this section, the visualization presents ten cluster representative images and upon hovering over the images, a viewer can see the associated cluster within the new portrait grid. The visualization is highly responsive and intuitive. This interaction draws attention to the critical link between the clusters and the representative images while offering an important point that machines cannot tell how the images are alike in the real world. An exploration of these clusters shows how AI systems sometimes cluster in ways that are familiar to viewers, but also in ways that are not intuitive.

To enhance the experience, viewers are encouraged to click on a representative image, which will animate the grid axis and reveal more details about that cluster. The associated portraits are magnified and re-arranged on the screen. When viewers select another representative image, they are taken to the new cluster within the grid. A key effect of the zooming across and in and out of the grid is to give the viewer a true feeling of exploration within the dataset. Figure 3-12 shows a mid-transition screenshot of images falling to place. When the viewer navigates the cursor away from the representative images, the visualization resets.

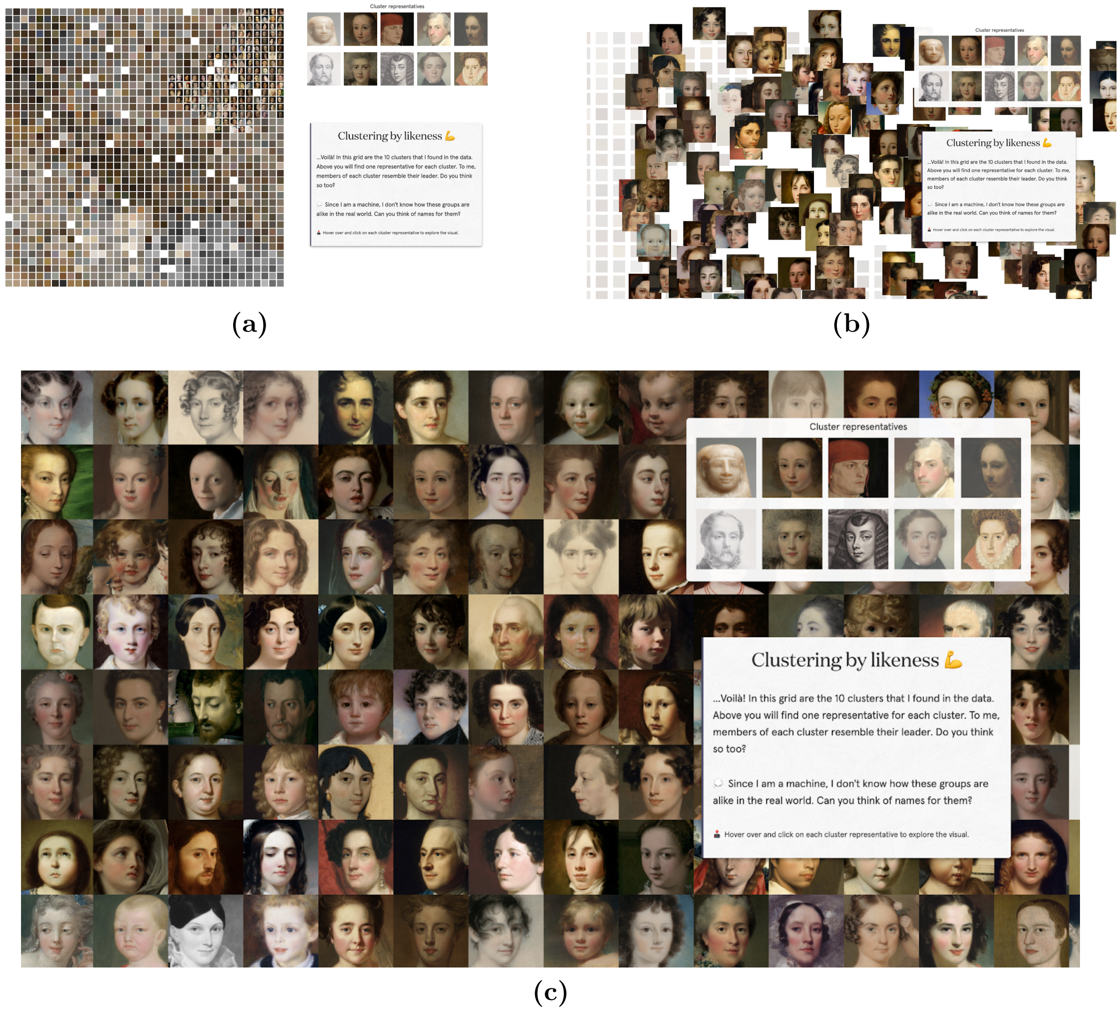


Figure 3-12: Mid-transition screenshots of images from a specific cluster flying into place from (a) to (c), and other clusters fading out.

3.2.7 "Latent Space"

The viewer is now introduced to the idea of latent space encoding as a means to generate entirely new portraits. The images displayed appear realistic, but upon closer inspection via tooltip, a viewer will notice titles are inappropriately matched with the subject within the painting. This mismatch was a design choice to signal to the user the shortcomings of the AI system and the bias embedded within the dataset. The viewer then learns about latent space arithmetic, which explains at a high level how these new images are created by seeking the in-between imaginary images between two portraits, as demonstrated in Figure 3-13.

Figure 3-13: Gridview illustrating latent space arithmetic - the bottom rows are midpoints between the top row and the left image.

The visuals were created by a StyleGAN2-ADA model trained on the MetFaces dataset. The portrait titles were created using a text-generating neural network trained on the titles in the original dataset. The network also learns the structure of the titles, such as letter casing and the inclusion of years in some samples. Examples titles generated include "Madame Head of a Lady" and "The Grandre Colley (1762 - 1791)".

3.2.8 "Have I Met You"

This final section serves as a reinforcement around the idea of latent space. Viewers are shown an animated transformation of four celebrities being projected into the latent space of generative neural network (see Figure 3-14). They are led by the narrative to consider how the outcome is limited by biases in its training data. This video is interactive and viewers can drag the image to scrub frames back and forth.

Figure 3-14: Outputs from projecting four celebrity faces into the latent space.

3.2.9 Discussion

A qualitative user study of MetAI with five people revealed that explainable visualizations are effective for people feel more comfortable with broaching neural art topics. The visualization communicates how computational systems can parse, abstract, and transform data. Viewers better understood how the output of a generative system is a product of its analysis of a dataset and the dataset itself. They also better understood how machine learning systems can generate neural artwork and gained an intuition for the latent space. In the final scene (see Figure 3-14), viewers saw how the model performed better on some faces than on others, and were prompted to consider how the original data affected this outcome. Most importantly, viewers indicated an increased propensity for applying and discussing such systems. Future work could illuminate more details at each level, and provide viewers with resources to produce their own neural art.

From an industry and technological adoption perspective, human-centered AI developers must focus on establishing, nurturing, and maintaining trust in AI systems. This work is an example of how explainable visualization tools can help us move closer to this ideal state.

3.3 Visual Poetry

Visual poetry is an interactive platform for generating and stylize images from poems. Using AttnGAN trained on the COCO dataset, phrases are generated into images that can abstractly resemble the input. To reinforce the emotion evoked by the poetry and improve cohesion, users can choose to stylize the image via one of the integrated neural style transfer models which have strong correlations to certain emotions. Each image can be manipulated together or individually. The final output is a series of images presented alongside their original text input as personalized visual poetry. This work was a collaborative effort with Peitong Chen and Alicia Guo.

3.3.1 Motivation

The view that literary art like poetry can offer insights into the nature of human emotions has gained increasing popularity in recent years [60]. These can often unfold human emotions and sentiments by forming imaginative imagery facilitated by linked-concepts and correlation. Each line is often self-contained but also collectively form layers of meaning. Visual Poetry explores the layers of meanings and emotions by altering the traditional way people approach poetry. Through the facilitation of image generation and stylization, Visual Poetry aims to create associable visual elements using the literary items and reinforce the audiences' evoked emotion.

To test and demonstrate this effect, we created a website where users can input one or more lines of a poem, obtain visual interpretations of these sentences, and then convert these images into art styles of their choice. We also used a polaroid border and text style on each of the images which proved to be highly versatile for many contexts, and which heightened the emotional response users had from experiencing the final images. The website can be accessed here: `oliviaseow.github.io/visual-poetry`.

On the backend, the website receives strings through the input field (Figure 3-15). It processes each input string through the AttnGAN model when it receives a line break. Images are immediately available for viewing just below the input box.

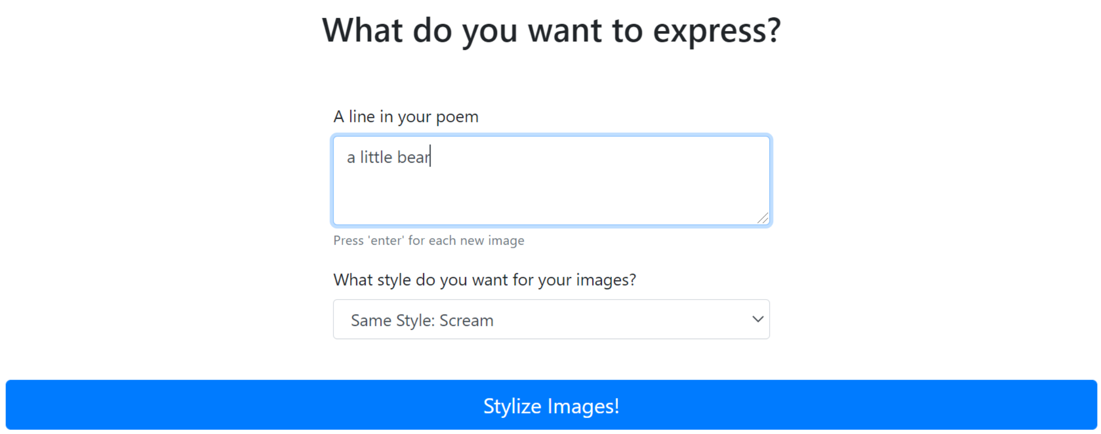


Figure 3-15: Input field on Visual Poetry website.

A user can then select one of the eight 'Same Style' options (as shown in Figure 3-16), or manually select a 'Different Style' for each image. The images are then processed via the corresponding neural style transfer model.

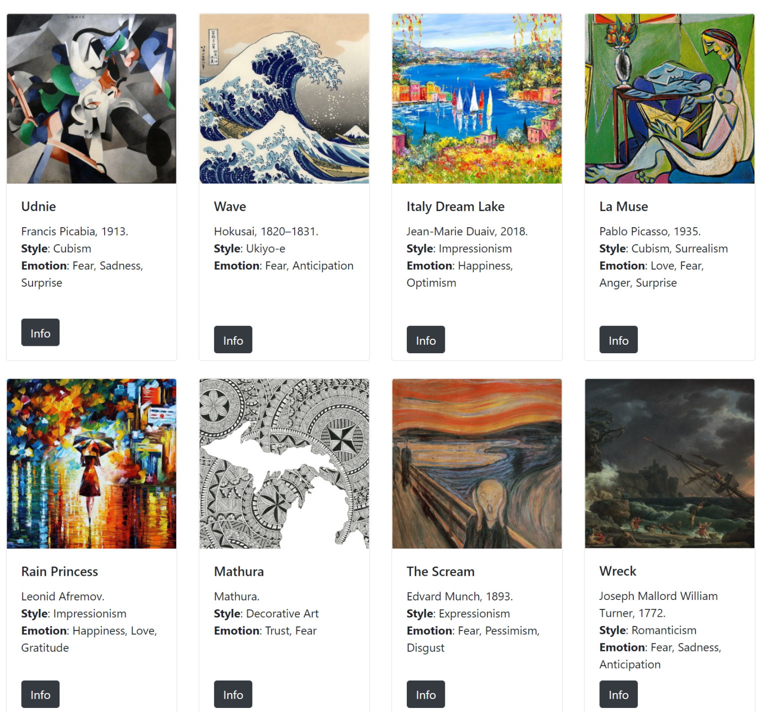
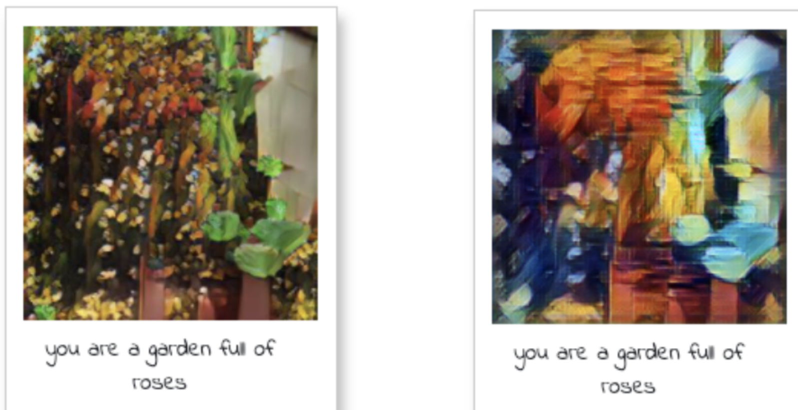


Figure 3-16: Neural styles available on website.

3.3.2 System

The website uses the AttnGAN model trained on the COCO dataset, which has many everyday objects with the corresponding labels [71]. This model enables attention-driven, multi-stage refinement for fine-grained text-to-image generation. The images can then be optionally processed via neural style transfer [33]. This work builds on the WikiArt Emotion research [57] to derive influences of art styles on emotions. Figure 3-20 shows one example of generated images from Visual Poetry.



(a) Generated image (b) Processed with emotive style

Figure 3-17: Generation sequence for "You are a garden full of roses".

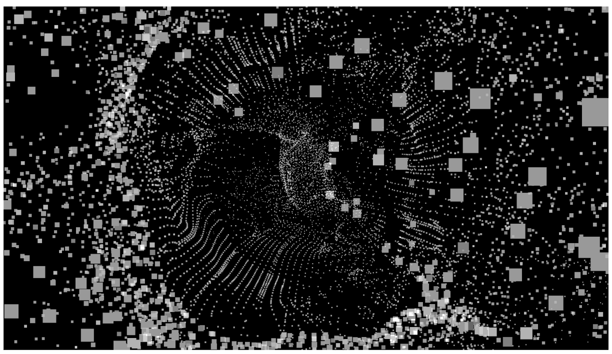
It is ultimately difficult, perhaps impossible, to accurately dissect complex human emotions and associate it with entire art genres. This work is a highly experimental approach to express literary narratives like poetry using evocative visual styles.

3.3.3 Discussion

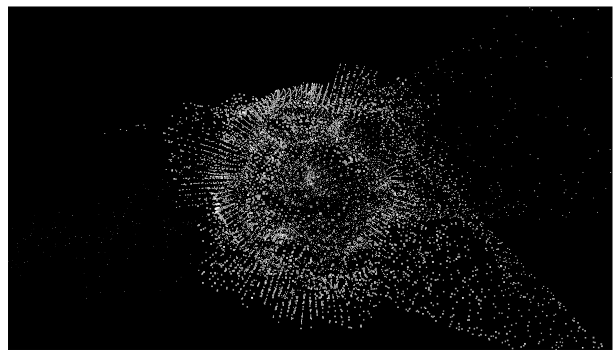
A qualitative user study of Visual Poetry with five people revealed that users gained a better understanding of generative machine learning capabilities. Combining the neural art techniques of AttnGAN and neural style transfer unlocked self-assessed opportunities for increased creativity and expression. Users were also inspired to think about how combined techniques can produce novel results, suggesting that artists can incorporate neural tools into their existing toolkit. Users expressed a desire for neural art tools to offer finer control and feedback during operation.

3.4 Machine Pigment

This work is an immersive art experience based on point clouds generated by CLIP-guided text2mesh. [46], which was collaboratively implemented with Matthew Fisher. The point clouds are animated with a noise walk derived from a GAN latent space. Viewers can navigate the space on a web browser or in VR, and observe the ebb and flow of the particles from any perspective, two of which are shown in Figure 3-18.



(a) One possible perspective



(b) Another perspective

Figure 3-18: Two perspectives for experiencing Machine Pigment.

To create this experience, a 3D object was first generated (see Figure 3-19).

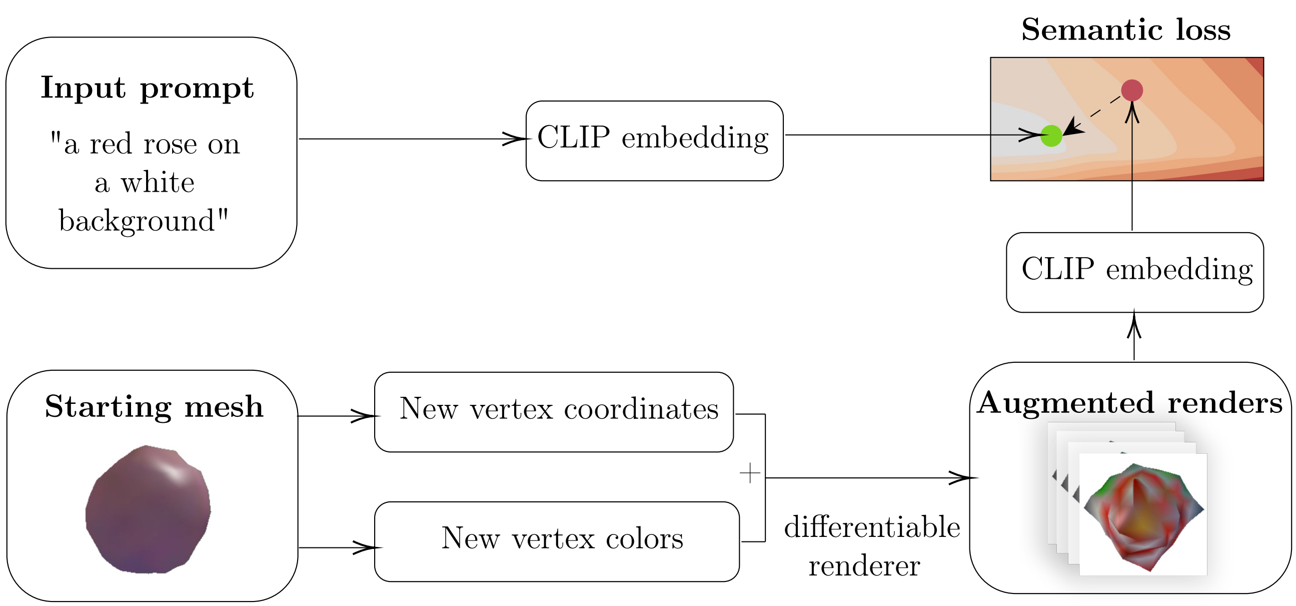


Figure 3-19: Simplified overview of text2mesh system [46].

This stylizes an input 3D mesh by predicting color and local geometric details which conform to a target text prompt. A similarity score between a text prompt and augmented renders of the perturbed input mesh is used for calculating CLIP semantic loss. This enables the synthesis of stylized meshes that resemble the input prompt. The system runs on Python 3.7 with CUDA 10.2. Using four RTX 2080 Ti GPUs, each model required 15 minutes to generate. Sample 3D outputs from this system are shown in Figure 3-20, for the target input prompt "a tiny world full of plants, colorful flowers, and a waterfall".

Figure 3-20: Five sample 3D outputs of tiny worlds.

I then converted the output 3D meshes (.obj files with vertex colors) into a point cloud representation in order to create the immersive experience using web shaders. The experience can be viewed at `oliviaseow.com/machine-pigment`.

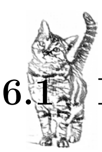
3.5 Discussion

A qualitative user study of Machine Pigment was conducted with five people from a range of technical backgrounds. Users with limited technical experience expressed that the immersive experience gave them an increased appreciation for machine generated art work. While expressing concern for 3D professions, some commented that this makes 3D design (e.g., custom room designs) accessible to all. However, more work is needed to optimize the pipeline for generating aesthetic 3D meshes efficiently. It was also found that 3D object generation has a large potential in game development, product design, architecture, and AR/VR. Future research should explore tools that can generate high quality 3D results and integrate with existing 3D workflows.

3.6 TheCAT

Through my research, I found that many people want to engage in neural art. Some creatives who actively engage with new technologies express that it greatly benefits their work. However, many people find buzzwords like machine learning and artificial intelligence opaque and intimidating. Those looking to learn more are often faced with a deluge of knowledge on the web, and a need to navigate a mess of outdated tutorials and courses. Frustration with this often leads to goal abandonment.

Consequently, we lack diverse perspectives testing, influencing, and driving AI research. TheCAT is a prototype of a platform for individuals at the intersection of art and tech to learn how to apply technologies creatively, and to discuss the broader issues surrounding applications of technology in context. An accompanying website was set up (see Figure 3-21), along with a Discord server (with 111 users at the time of writing). TheCAT allows people to learn about new technologies and to create prototypes. It has tutorials, prototyping tools, and a community of like-minded individuals from a variety of disciplines.



3.6.1 Discussion

TheCAT included content about neural art as well as other topics in creative coding, extended reality, and physical prototyping. A total of seven live, virtual sessions on neural art have been conducted for an audience of largely non-technical artists, reaching over 300 people. Examples of sessions include learning how to use PoseNet to create a nose drawing tool, and using Teachable Machine to control animated visuals.

Before a session, participants were asked to write down one word or phrase that they associate with machine learning. Phrases like "black box", "over my head" pointed at the common lack of understanding about neural art. By the end of the sessions, participants were not only able to deploy creative machine learning web applications, they also reported a greater ability to discuss opportunities, limitations, and challenges about neural art. This structure has proven to be effective in reaching and growing a community of diverse perspectives interested in neural art.

Figure 3-21: Screenshot of TheCAT website.

Chapter 4

Survey on Societal Concerns

AI cultivates and caters to our passivity, seeming to offer the fruits of creativity and self-examination without the effort and self-doubt. Algorithms always find us interesting and always testify to our insatiable desires by showing us all the things we should want.

— Rob Horning, Word Processing (Art in America, 2022)

Powerful neural art systems may have far-reaching effects on society by disrupting how digital media is produced and consumed. This has spurred speculation and concern from the community, which I observed during the course of my work presented in Chapter 3. To better understand these voices, I released a survey to open groups on Twitter, Facebook, and Discord, inviting perspectives on neural art. Survey questions can be found here. Following COUHES protocol, participation was voluntary, and responses were anonymous by default. Over two weeks beginning April 15, 2022, I received 43 responses. A synthesis of the findings are presented in this chapter.

4.1 Respondent demographic

Respondents had diverse backgrounds across art, music, software, data science, energy, transportation, healthcare, sales, and education. As shown in Figure 4-1, art making was a major impetus for interest in neural art, with over 27 respondents (62.8%) reporting 'art making' as their primary interest in neural art, followed by 7 respondents (16.3%) indicating 'technical interest' as their main motivation.

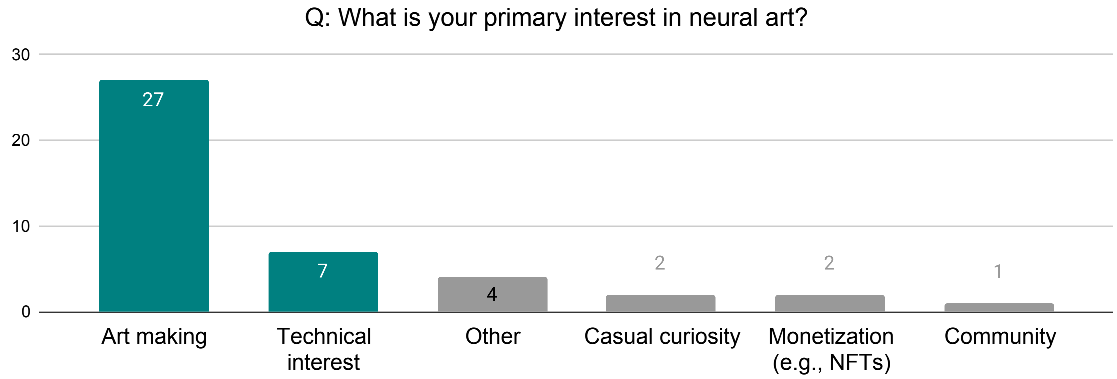


Figure 4-1: 'Art making' and 'technical interest' were the biggest reasons for respondents' interest in neural art.

The majority of respondents (29, or 67.4%) identified as artists, with 18 self-identifying as amateurs and 11 as professionals. Other respondents reported varying levels of involvement with art, and only 2 (4.7%) had no other connection to the art scene. The occupations of all respondents can be found in Table 4.1 at the end of this chapter, edited to avoid over-specificity.

As illustrated in Figure 4-2, most respondents were not well versed with modifying and originating neural art code, with 26 (60.5%) feeling comfortable only to make small modifications to existing code. 7 respondents (16.3%) could write original code.

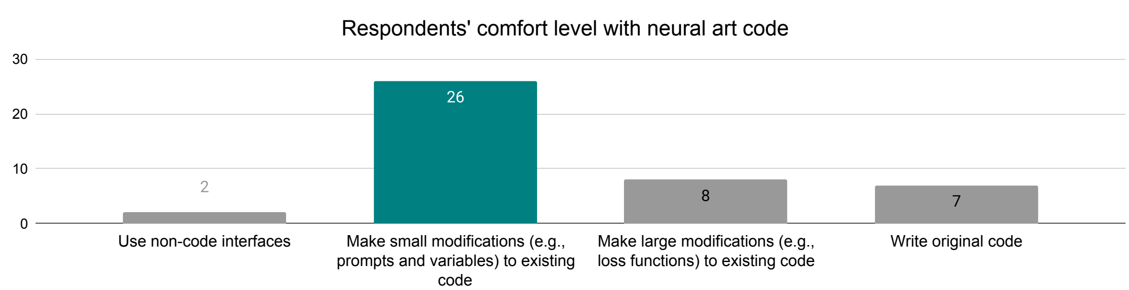


Figure 4-2: Most respondents were only comfortable making small modifications to existing code.

Overall, the respondent demographic, a self-selected volunteer sample of the neural art community, comprised largely non-technical artists. While effort was taken to share the survey across multiple open platforms, it is possible that this does not represent the entire population engaged with and impacted by neural art.

4.2 Survey findings

This section discusses main themes surfaced from the survey, including accessibility, impact on art and artists, gatekept and predatory services, ownership and attribution, divisiveness and tensions, lack of representation, and environmental concerns.

4.2.1 Accessibility

Inaccessibility was a major theme in the survey. 22 respondents (52.4%) lamented the technical hurdles, that it was hacky or confusing to use available code, or that the tools made it difficult to produce coherent outputs. This is possibly because respondents fell into the largely non-technical category, and the most easy and effective tools are widely showcased but not openly accessible.

4.2.2 Impact on art and artists

Respondents also identified that the efficiency of neural art may automate away designer jobs, contributing to inequality and economic security issues for some individuals.

> "I think that traditional artists and especially designers are going to be very sad as their entire careers and skill sets are completely subsumed by AI tools. Will they be able to step into it or will they have to find new jobs? If it takes a human hours to create a logo design, but anybody without any design background can enter idea prompts and generate 100 drafts in a minute, who is going to pay a human for that kind of work?"
> -R33

There were also mentions of work possibly being exploited. A common practice in generating neural art is to produce visuals in the style of a certain artist by using their name in an input prompt. Networks like CLIP are able to do this because they were trained on images by that particular artist. While OpenAI claims to have licensed the images in their training set, some artists have discovered that their art style has been learned by the model, allowing others to replicate their techniques without further compensation and in ways that may run contrary to their interests.

> "My work uploaded years ago is being used by others as an art style. This is actually impressive but also frustrating as I can't do anything about it." -R3

Interestingly, the relative ease of generating neural art is also balked at by some who claim this attracts those lacking in knowledge or talent.

> "It will become more-and-more associated with low-effort, or no-effort, pretend-artistry, like having uncredited ghostwriters, causing stigma... Some people who typed a sentence into a pre-built notebook act like big-brained genius inventors, which they are not." -R40

> "A bunch of derps entering prompts to a system they could never create themselves." -R41

Newcomers may try to emulate others, which contributes to perceptions of "over-saturation" and "lack of originality" (R23).

> "They are too technical for newcomers to engage with on a technical level, but too alluring and novel not to attract prospective users, leading to a lot of same-y production." -R40

> "A very small number of methods are used 99% of the time and lead to narrow recognizable aesthetics." -R28

While individual neural artists try to eke out a win, there are also concerns about a "monopoly on art" (R3) by institutions that then concentrate power and accumulate massive wealth.

> “This technology will revolutionize what art is in potentially every medium. Will we see the inequities of the industrial revolution repeated again or can we adapt our culture to prevent the wealth generated through this technology from only going to those with the money to fund the development of the technology?” -R33

> “DALL-E 2 will probably mean nothing we can create with open-source tools is interesting or novel any more.” -R1

4.2.3 Gatekept and predatory services

The most compelling neural art generative algorithms are time consuming to train and incur large image licensing and computational costs. This prohibits competition and results in most advancements being concentrated in well- funded firms. Such firms do not open source all their work due to costs, potential misuse, and consequent backlash. Instead, OpenAI grants “friends/family of employees” free access [13], as does Midjourney. Both are expected to be paid services in the future. To quote respondent R35, there is “information asymmetry”, and many simply cannot access these powerful tools.

> “The way a lot of research firms release their findings [is] mainly gatekeeping like OpenAI... It seems the drive is heading more towards being first to get publicity and drive [the] market to make money rather than releasing the code and letting everyone actively utilize the tools.” -R9

Where research models are being published, savvy technologists can use and modify them, resulting in no-code paid services like NightCafe. The operational cost of such tools is not negligible, but for the community, fees may seem predatory toward non-technical artists.

> “I hate that sites like nightcafe charge as much as they do... it feels somewhat predatory.” -R14

4.2.4 Ownership and attribution

Concerns on whether or not neural art can be considered as art in its own right (R14, R27) has in a similar way afflicted digital artists since the 1960s, and is an ongoing debate [64]. On attribution, R40 suggested that there should be standardized practices for neural art. One possibility is a watermark invisibly incorporated into generated art works. The laws on attribution of neural art are currently complex and unclear.

4.2.5 Divisiveness and tensions

Neural art introduces a number of tensions in society. Interestingly, there is a tension between those who have figured out settings and techniques (typically in the form of numbers, words, and custom code) to produce aesthetically desirable neural art, and others who perceive that such knowledge should be shared for the common good.

> “[I wish for] more collaboration and less hoarding private techniques and prompts like Gollums.” -R10

A second tension comes from tool creators blocking certain prompts from being processed, perhaps with the intention of reducing risk. When viewed from another lens, this can be limiting and repressive.

> “It’s not just about whether you can type in fuck or look at generated porn. This is just a symbol of repression and uniform thinking.” -R3

Another significant tension arises from the proliferation of NFTs, commonly used to monetize neural art. Anti-NFT voices suggest that NFTs are channels for greed and deviousness, and yet others are opposed to monetizing neural art altogether (R38).

“NFTs can fuck off.” -R1

“So much pixel trash. SO so much pixel trash minted into NFTs. It’s infuriating.” -R43

“[Neural art] will be used to generate art at volume for deceptive purposes (like NFTs), furthering the stigma.” -R40

“Stop making NFTs. They reproduce the worst parts of the commercial art market under the guise of "eliminating gatekeeping". NFTs produce the same kind of gatekeeping that social media proposed to eliminate for culture production. On top of this, they’re accelerating global warming.” -R43

“Pessimistic about current association with NFTs, and greed that comes with neural art becoming profitable for artists. (since this leads to less community, idea sharing, and maybe bad public perception).” -R23

While some respondents clearly detest NFTs or believe *There is no Such Thing as Blockchain Art* [29], it is also noteworthy that a third of respondents partake in it, with 14 people (32.5%) minting NFTs with the neural art they create (see Figure 4-3).

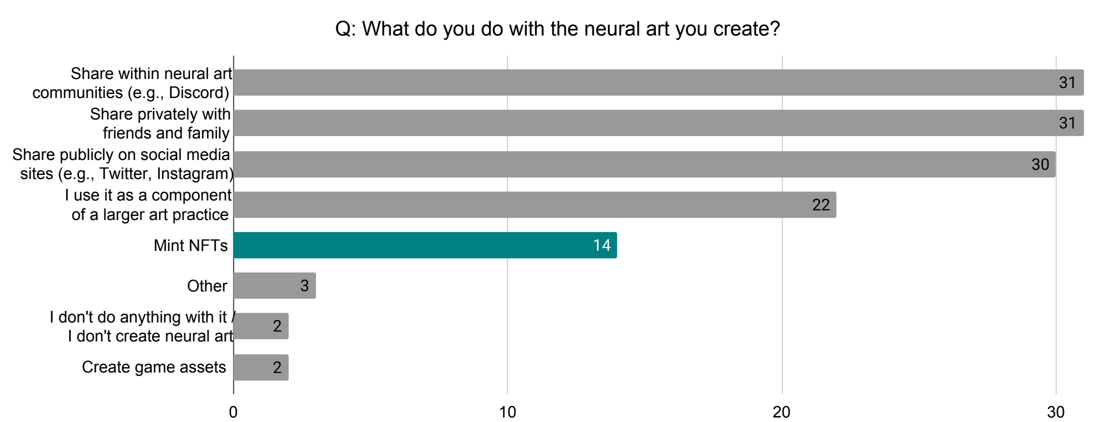


Figure 4-3: One-third of the respondents (14 out of 43) mint NFTs with neural art.

4.2.6 Lack of representation

Unequal gender representation was another issue that was identified. This hints at deeper issues of representation both in tech [1] as well as the arts, where more women are being trained but still disproportionately less are seeing success [7]. As neural art is at the confluence of different fields, it is important to study this matter and to avoid exacerbating inequalities.

> "I hope to see more women and nonbinary folks enter the community. Or, to feel comfortable disclosing their identities if they are already present." -R14

4.2.7 Environmental concerns

The immense computational requirements of neural art and associated NFT transactions may contribute to climate change (R3, R43). This has severe financial, social, and economic effects, and should be cautiously assessed.

4.2.8 Looking forward

When asked to envision how the neural art scene will evolve in one year, respondents predicted exponential algorithmic advancement (R6, R17) as well as an influx of self-proclaimed artists (R18). Some pessimistic perspectives suggest that artists will lose means and control to art creation and creativity:

> "A small renegade band of insurgent ML developers hiding in a cave, powering a GPU by frantically riding bikes hooked up to a generator, and desperately trying to avoid falling further behind the corporations who let's say control the whole power grid at this point." -R10

> "Creativity and technology will be kept behind locked doors — unavailable to purchase at any price. When AI art's marginal cost truly drops to

> zero, we plebeians will no longer be allowed casual access to the means of production.” -R10

On the other hand, more optimistic perspectives anticipate that neural art will induce new and meaningful art movements, e.g.,:

> “Artistic sub-genres and collaboration dedicated to finding purpose for human-driven AI art, such as movements in the past” -R13

> “I think a new genre of fine art will emerge with creative processes involving AI that are yet to be seen.” -R36

> “I would like to see works using generative imaging to move away from simply using the tool to produce an image and incorporate the methods into larger, more substantial projects. Examples of this already exist, but the majority of work feels akin to displaying the tool not the artwork, ‘exhibiting the paintbrush, instead of the painting’, if you will.” -R43

Ultimately, both sides agree that change is expected. This could have large impacts on any media-related field. It is important for those in power to carefully reflect on internal incentives and points of concern from society.

I would like to acknowledge `@electricxbunny`, `@HostsServer`, `@jags111`, `@joehendel`, `@joshurbandavis`, `@Mike_Howles`, `@nftartech`, `@PANCHOCODES`, `@Shane54music`, `@SirRealismo`, and `@unltd_dream_co` for their contributions.

Table 4.1: Survey respondents and their stated occupations.

Respondent	Occupation
R1	Software Dev / Startup Founder
R2	Researcher
R3	Student
R4	PhD Student in Studio Art
R5	Run a Screen Printing Shop
R6	IT
R7	Product Design
R8	Energy Trader
R9	Network Engineer and Programmer
R10	Data Scientist / Data Engineer / GIS Developer
R11	Student
R12	Artist / Entrepreneur
R13	Tech Lead
R14	Organizational Development Consultant
R15	Musician
R16	Mom
R17	Product Designer
R18	Courier
R19	Data Scientist and Educator
R20	Musician
R21	Technical Art Director
R22	Student
R23	Student
R24	Artist
R25	Student
R26	Software Developer
R27	DJ / Music Producer
R28	Artist
R29	Retired Nurse
R30	Scientist
R31	Computer Scientist
R32	CS Student
R33	Software Developer
R34	Artist
R35	Artist
R36	Graphic Designer
R37	PhD Student
R38	Scrum Master
R39	Artist
R40	Research Developer (not in neural networks)
R41	Senior Software Engineer
R42	Tech Sales
R43	Research Scientist and Artist

This page intentionally left blank

Chapter 5

Challenges

Only within the past ten years have we begun to accept the possibility that technological solutions are not universal panaceas.

— Kathleen Woodward, The Myths of Information (1980)

Thus far, I have introduced a primer on neural art and explored findings from a survey of 43 people involved in neural art. In this chapter, I will expand on some latest challenges that affect in neural art, which are important to consider when deciding *when* and *how* to use the technology, as well as when charting the future of related research. While broadly applicable to neural art, many themes arise from the latest work in language interface generated visuals, with particular emphasis on OpenAI's DALL·E 2 as it is a massively trained model with cutting edge capability. The structure of this chapter is adapted from 'Data Science in Context: Foundations, Challenges, Opportunities' by Spector et al [62].

5.1 Intractable data

Large datasets used in the latest neural art models are very tricky to sanitize. For instance, OpenAI does not allow the "generation of images related to common American hate symbols" on DALL·E 2 [13]. However, who gets to decide this list is determined by "researcher privilege" [45], and it is unclear what symbols make the cut. Furthermore, it is well understood that the concept of hate speech has complex cultural roots and and is deeply subjective, making it difficult to extract a definitive set [41]. Tools like Know Your Data are a helpful first step in understanding biases in a dataset, though data processing may have unintended effects. In one example, filtering out sexual content reduced the quality of generated images of women in general [50].

Despite any attempts to sanitize input images, we know that users (including children, as discussed in Section 3.1) can and will be *creative* in subverting the system to produce inappropriate results. For an expressive system like DALL·E 2, this may be exceptionally difficult to predict. We can easily enumerate ways that language based neural art interfaces can be misused to produce explicit content:

— Using words that are neutral in one culture but graphic in another (e.g., "knob" is neutral in the United States but offensive in the United Kingdom).

— Using emojis that have been used to convey explicit content (e.g., some fruit and vegetable emojis have lewd connotations).

— Combining seemingly neutral concepts inappropriately (e.g., "girl" and "shower").

— Using *visual synonyms*, which are words that have different meanings but a similar visual [34] (e.g, red nail polish can visually substitute blood).

— Using coded words or other anthropomorphic representations (e.g., dolls).

Another important factor is that explicit results can be triggered unintentionally if a person is unaware of a phrase's multiple meanings, such as shown in Figure 5-1, where an input prompt of `strawberry fields forever` displays entangled references to the berry, bird's-eye view of red fields, marijuana and paraphernalia, and a Beatles song (the rainbow may be a connection to the cover of 'Magical Mystery Tour').

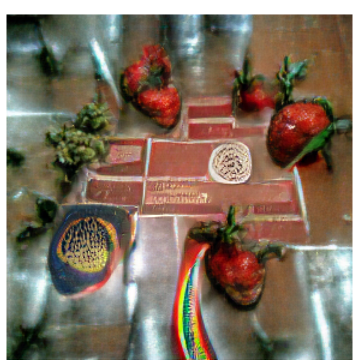


Figure 5-1: Example of reference collision in CLIP + VQGAN.

5.2 Bias and stereotypes

Generative models *inherit* various biases from its training data, which may reinforce stereotypes, erase, or denigrate certain groups [13]. Models like DALL·E 2 also *introduce* its own set of biases due to the team being primarily English speaking and located in the United States [13]. For example DALL·E 2, Figures 5-2 shows DALL·E 2 displaying disproportionately more white-passing and male-passing people in western attire for the prompt `lawyer`, while Figure 5-3 shows female stereotypes for the prompt `personal assistant`. Moreover, insidious downstream effects may arise if the model makes assumptions about what is an "apropriate" representation of people in certain communities. People with disabilities are particularly vulnerable here.

Figure 5-2: DALL·E 2 outputs for the prompt `lawyer` [13]
.

Figure 5-3: DALL·E 2 outputs for the prompt `personal assistant` [13]
.

5.3 Lack of understandability

Due to the complex ways that deep learning models acquire knowledge with a large number of interconnected perceptrons instead of a clear set of rules, the outcomes of neural networks lack understandability. This can lead to unusual or shocking outcomes.

For example, Discord user `annetropy` discovered inexplicable nudity in outputs from the Disco Diffusion model based on seemingly innocuous input prompts describing fashion. This may have been another instance of reference collision illustrated in Section 5.1, or a more complicated entanglement of concepts within the model.

The lack of understandability may also have averse effects on the interpretability and transparency, which may prevent some institutions from accepting its use due to a low tolerance for risk and a neural network's lack of auditability.

5.4 Protecting individual welfare

Neurally generated visual media have already been used for exploitative purposes. One examples of this is Deepfakes [43], where someone's likeness is used in compromising or misleading ways. It will become progressively easier to replicate images and artwork without consent, and to spread defamatory or harmful content.

Neural art models have also already learned from and are able to replicate styles from hundreds of artists. There is an underlying issue of uninformed consent, where the original artists, photographers, or subjects may not be aware that their data is being used for such purposes.

In addition, as discussed in Section 4.2.2, some contend that neural art makes it more difficult for artists to make a living since digital media can be automatically generated by neural networks. This response is somewhat analogous to the response when photography was introduced in the 19th century. After the initial resistance, photography was ultimately recognized as an artform and is now arguably an essential part of our connected world. Because the scale at which neural networks operate is unprecedented however, we may experience a divergent outcome.

While OpenAI's closed source and relatively moderated DALL·E 2 is an easy target for criticism, it is worth noting that various research communities are also working on open-sourced versions [68]. Open-sourced tools invite more opportunities for checks and balances, but also enable malicious, silent spin-offs.

5.5 On being human

Creativity is often thought of as a bastion of humanity [22]. Creative endeavors like visual arts are often described as deeply personal, effortful, and rich in context. It may hence be desecrating to some that machines can create hyper-aesthetic visuals based on mathematical extrapolations of existing work. As shown in Figure 5-4, some contend that visual arts will be destructively affected by neural art [67].

Figure 5-4: Negative sentiment from Twitter on DALL·E.

5.6 Legal issues

While web scraping publicly available data is typically legal, there is ambiguity as to the morality of distilling the essence of other people's work into an automated system. As discussed in Section 4.2.2, there is an unequal distribution of benefits amongst parties, and some artists are frustrated that they have no control over their unique styles being used by for others' commercial gain. There is currently no pathway for artists to get sweat equity from artwork that is created, and these *are* already being used commercially. Language interface neural art tools like Snowpixel claims that there are no restrictions on use of generated images, while tools like NeuralBlender transfer copyrights to premium users. Past research found that people allocate credit and responsibility of AI art first to the artist, then the curator, then the technologists, and finally the crowd from whom data is sourced [28]. However, cyber laws have yet to catch up, and there are no clear conventions for accreditation and ownership.

5.7 Environmental impact

As presented in Section 4.2.7, there are concerns about the negative economic and environmental impacts of neural art and associated activities. For instance, OpenAI's CLIP model's training process took 30 days across 592 V100 GPUs, which cost an estimated USD1 million [3] and produced around 200,000 pounds of carbon emissions [19], equivalent to the carbon sequestered by 1,500 trees grown for 10 years [66]. With increased public awareness and scrutiny, more research institutions are paying attention to this challenge [72].

5.8 Looking forward

In this chapter, we have discussed *what* neural art is, as well as concerns and challenges regarding *why*, *when* and *how* it is used. In the next chapter, I summarize my takeaways from this work, including the survey and case studies from previous chapters, and distill them into ten implications for neural art and visual domains.

This page intentionally left blank

Chapter 6

Implications

If this human-machine symbiosis turns out to be the wave of the future, then it will make more important those who can stand at the intersection of humanities and sciences. That interface will be the critical juncture. The future will belong to those who can appreciate both human emotions and technology's capabilities.

— Walter Isaacson, Jefferson Lecture (2014)

Over the last five chapters, I have introduced neural art, explored applications in my personal work, discussed findings from a survey of societal concerns, and expanded on some challenges identified. In this chapter, I distill my findings into ten implications for neural art and visual domains.

1. Neural workflows will be used in the most productive image generation pipelines in domains of visual media, e.g. graphic design, visualization, animation, game development, NFTs, and AR/VR (see Chapter 4). Its effects will be seen from brainstorming through to final polishing steps. Neural art will also be applied in unconventional ways, such as in history education or in combating aphantasia.

2. Many more artists will seek to collaborate with or be compelled to differentiate from neural artwork (see Section 4.2.2).

3. Neural art will be used to improve creative satisfaction (see Section 3.1). This can have ramifications on well-being and problem solving in other areas [24].

4. Massive developments in neural art will come from research institutions and open source alliances like `CasualGANPapers`'. Researchers have a burden of accountability in ensuring that their work does not induce harms. While it would be impossible to predict all ethical and environmental consequences, researchers ought to address and minimize negative outcomes (see Chapter 5).

5. Researchers ought to make explainability a priority, and consider the human context of their work, particularly addressing the values being embedded in their systems e.g., through the choice of training data (see Section 3.2).

6. It is important to have more diverse perspectives invited to understand and participate in discussions influencing neural art research, such as through platforms like theCAT, as shown in Section 3.6.

7. Language interfaces will proliferate for creating visual media of all kinds (e.g., static images, 3D models, animations) with increasingly natural interaction and fine control (see Sections 3.3, 3.4). Such systems will also be used to reverse engineer the techniques and prompts used to create an artwork (see Section 2.6).

8. With reverse search, "prompt hoarding" will become less relevant. Moreover, to make differentiated art, strong artistic sensibilities and a command of unique pre- and post-production tools are still relevant. Neural art will be seen as a tool or collaborator in an artist's workflow, but not the end-all (see Chapter 4).

9. The age-old debate about what is or is not art will persist, but fewer people will subscribe to the institutional theory of art, and instead look to decentralized decision making on what art is considered art, or hold that the art they create is art because they think it is so (see Section 4.2.5).

10. Cyber laws will have to continuously adapt to and clarify how neural art can be accredited, and provide artists with a means to protect their work. Neural interfaces may be used to verify the provenance of an artwork. Simultaneously, some artists will be more conscious of web hosting services and try to protect their work from being used in a massively trained neural art model (see Section 5.6).

This page intentionally left blank

Chapter 7

Conclusion

With the aid of electronic computers the composer becomes a sort of pilot: he presses the buttons, introduces coordinates, and supervises the controls of a cosmic vessel sailing in the space of sound, across sonic constellations and galaxies that he could formerly glimpse only as a distant dream.

— Iannis Xenakis, La Revue musicale (1963)

In this thesis, I presented a primer on neural art and highlighted five personal case studies with creative applications. These reveal the inner workings and potential of generative machine learning models to impact visual domains. I then discussed the findings from a 43 participant survey on the societal concerns of neural art, and documented challenges affecting neural art. These provide the larger human context for neural art in society, and help to frame future discussions on ethical and social considerations. I also distill ten key implications from the research. Taken together, this work is a jumping-off point for increased understanding, discussion, and collaboration on neural art.

This is a start, and there remain unanswered questions to be explored. Future work should address how developers can instill transparency and explainability into neural art systems, how we can appropriately track, assign, and accredit neural creations, and how artists can protect their work from automated systems.

Bibliography

[1] 2020 Top Companies for Women Technologists. URL: https://anitab.org/research-and-impact/top-companies/2020-results/.

[2] Ai generative art tools. URL: https://pharmapsychotic.com/tools.html.

[3] A Beginner's Guide to the CLIP Model. URL: https://www.kdnuggets.com/a-beginners-guide-to-the-clip-model.html/.

[4] CreativeAI: Deep Learning for Graphics Tutorial Code. URL: https://github.com/smartgeometry-ucl/dl4g.

[5] DALL·E 2. URL: https://openai.com/dall-e-2/.

[6] Danbooru2021: A Large-Scale Crowdsourced and Tagged Anime Illustration Dataset · Gwern.net. URL: https://www.gwern.net/Danbooru2021.

[7] Get the Facts About Women in the Arts | NMWA. URL: https://nmwa.org/support/advocacy/get-facts/.

[8] Google AI Blog: Inceptionism: Going Deeper into Neural Networks. URL: https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html.

[9] Inceptionism: Going Deeper into Neural Networks. URL: http://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html.

[10] Kingfisher & Wombat on Twitter. URL: https://twitter.com/UrsulaV/status/1467984046059835399.

[11] Model Zoo - Generative Models Deep Learning Models and Code. URL: https://modelzoo.co/category/generative-models.

[12] Neural Filters in Photoshop - Make smarter edits - Adobe. URL: https://www.adobe.com/products/photoshop/neural-filter.html.

[13] Openai/dalle-2-preview. URL: https://github.com/openai/dalle-2-preview.

[14] Pindar Van Arman's cloudpainter. URL: https://www.cloudpainter.com/.

[15] The Quick, Draw! Dataset. URL: https://github.com/googlecreativelab/quickdraw-dataset.

[16] Refik Anadol Studio. URL: https://refikanadolstudio.com/.

[17] SnapML Overview | Docs. URL: https://docs.snap.com/lens-studio/references/guides/lens-features/machine-learning/ml-overview/.

[18] Timnit Gebru Is Building a Slow AI Movement - IEEE Spectrum. URL: https://spectrum.ieee.org/timnit-gebru-dair-ai-ethics.

[19] Training a single AI model can emit as much carbon as five cars in their lifetimes. URL: https://www.technologyreview.com/2019/06/06/239031.

[20] M. Akten. Deep Visual Instruments: Realtime Continuous, Meaningful Human Control over Deep Neural Networks for Creative Expression. URL: https://research.gold.ac.uk/id/eprint/30191, doi:10.25602/GOLD.00030191.

[21] W. Z. BENJAMIN and HARRY. *Work of Art in the Age of Mechanical Reproduction: An Influential Essay of Cultural Criticism.*

[22] M. A. Boden. Creativity in a nutshell. 5(15):83–96. URL: https://www.cambridge.org/core/product/identifier/S147717560000230X/type/journal_article, doi:10.1017/S147717560000230X.

[23] R. Danhaive and C. T. Mueller. Design subspace learning: Structural design space exploration using performance-conditioned generative modeling. 127:103664. URL: https://linkinghub.elsevier.com/retrieve/pii/S0926580521001151, doi:10.1016/j.autcon.2021.103664.

[24] N. Daykin. Creativity, health and wellbeing: Challenges of research and evidence. 51(4). URL: https://musiikki.journal.fi/article/view/113249, doi:10.51816/musiikki.113249.

[25] P. Dhariwal and A. Nichol. Diffusion Models Beat GANs on Image Synthesis. URL: http://arxiv.org/abs/2105.05233, arXiv:2105.05233.

[26] S. Donnelly and E. Kidd. The Longitudinal Relationship Between Conversational Turn-Taking and Vocabulary Growth in Early Language Development. 92(2):609–625. URL: https://onlinelibrary.wiley.com/doi/10.1111/cdev.13511, doi:10.1111/cdev.13511.

[27] A. Dorin, J. McCabe, J. McCormack, G. Monro, and M. Whitelaw. A framework for understanding generative art. 23(3-4):239–259. URL: https://www.tandfonline.com/doi/full/10.1080/14626268.2012.709940, doi:10.1080/14626268.2012.709940.

[28] Z. Epstein, S. Levine, D. G. Rand, and I. Rahwan. Who Gets Credit for AI-Generated Art? 23(9):101515. URL: https://linkinghub.elsevier.com/retrieve/pii/S2589004220307070, doi:10.1016/j.isci.2020.101515.

[29] M. Fernandez, S. Gustafsson, and F. Kakoubay. There Is No Such Thing as Blockchain Art - A Report on the Current Status of the Intersection of Blockchain and Art.

[30] D. Foster. *Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play.* O'Reilly Media, Inc, first edition edition.

[31] G. Franceschelli and M. Musolesi. Creativity and Machine Learning: A Survey. URL: `http://arxiv.org/abs/2104.02726`, `arXiv:2104.02726`.

[32] A. Fürst, E. Rumetshofer, J. Lehner, V. Tran, F. Tang, H. Ramsauer, D. Kreil, M. Kopp, G. Klambauer, A. Bitto-Nemling, and S. Hochreiter. CLOOB: Modern Hopfield Networks with InfoLOOB Outperform CLIP. URL: `http://arxiv.org/abs/2110.11316`, `arXiv:2110.11316`.

[33] L. A. Gatys, A. S. Ecker, and M. Bethge. A Neural Algorithm of Artistic Style. URL: `http://arxiv.org/abs/1508.06576`, `arXiv:1508.06576`.

[34] E. Gavves, C. G. Snoek, and A. W. Smeulders. Visual synonyms for landmark image retrieval. 116(2):238–249. URL: `https://linkinghub.elsevier.com/retrieve/pii/S1077314211002153`, `doi:10.1016/j.cviu.2011.10.004`.

[35] M. Glickstein. Golgi and Cajal: The neuron doctrine and the 100th anniversary of the 1906 Nobel Prize. 16(5):R147–R151. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0960982206012036`, `doi:10.1016/j.cub.2006.02.053`.

[36] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. URL: `http://arxiv.org/abs/1406.2661`, `arXiv:1406.2661`.

[37] D. Grba. Deep Else: A Critical Framework for AI Art. 2(1):1–32. URL: `https://www.mdpi.com/2673-6470/2/1/1`, `doi:10.3390/digital2010001`.

[38] A. Hertzmann. Aesthetics of Neural Network Art. URL: `http://arxiv.org/abs/1903.05696`, `arXiv:1903.05696`.

[39] A. Hertzmann. Can Computers Create Art? 7(2):18. URL: `http://www.mdpi.com/2076-0752/7/2/18`, `doi:10.3390/arts7020018`.

[40] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila. Training Generative Adversarial Networks with Limited Data. URL: `http://arxiv.org/abs/2006.06676`, `arXiv:2006.06676`.

[41] J. Kocoń, A. Figas, M. Gruza, D. Puchalska, T. Kajdanowicz, and P. Kazienko. Offensive, aggressive, and hate speech analysis: From data-centric to human-centered approach. 58(5):102643. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0306457321001333`, `doi:10.1016/j.ipm.2021.102643`.

[42] M. Kranzberg. Technology and History: "Kranzberg's Laws". 27(3):544. `arXiv:3105385`, `doi:10.2307/3105385`.

[43] J. Langr and V. Bok. *GANs in Action: Deep Learning with Generative Adversarial Networks*. Manning Publications.

[44] P. Li, K. Aberman, Z. Zhang, R. Hanocka, and O. Sorkine-Hornung. GANimator: Neural Motion Synthesis from a Single Sequence. URL: `http://arxiv.org/abs/2205.02625`, `arXiv:2205.02625`, `doi:10.1145/3528223.3530157`.

[45] R. H. McLaughlin. Privilege and practice in social science research: [Rejoinder]. 24(4):999–1003. `arXiv:829156`.

[46] O. Michel, R. Bar-On, R. Liu, S. Benaim, and R. Hanocka. Text2Mesh: Text-Driven Neural Stylization for Meshes. URL: `http://arxiv.org/abs/2112.03221`, `arXiv:2112.03221`.

[47] A. I. Miller. *The Artist in the Machine: The World of AI Powered Creativity*. The MIT Press.

[48] M. Minsky, S. Papert, and L. Bottou. *Perceptrons: An Introduction to Computational Geometry*. URL: `https://ieeexplore.ieee.org/document/8093962`.

[49] ml4a. Machine learning for arts. URL: `https://ml4a.net/`.

[50] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, and M. Chen. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. URL: `http://arxiv.org/abs/2112.10741`, `arXiv:2112.10741`.

[51] OpenAI [@openai]. Here's a look at what DALL·E 2 can do. Want to see more? Follow along on Instagram: https://instagram.com/openaidalle/. URL: `https://twitter.com/openai/status/1511714511673126914`.

[52] K. O'Shea and R. Nash. An Introduction to Convolutional Neural Networks. URL: `http://arxiv.org/abs/1511.08458`, `arXiv:1511.08458`.

[53] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning Transferable Visual Models From Natural Language Supervision. URL: `http://arxiv.org/abs/2103.00020`, `arXiv:2103.00020`.

[54] I. D. Raji, E. M. Bender, A. Paullada, E. Denton, and A. Hanna. AI and the Everything in the Whole Wide World Benchmark. URL: `http://arxiv.org/abs/2111.15366`, `arXiv:2111.15366`.

[55] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen. Hierarchical Text-Conditional Image Generation with CLIP Latents. URL: `http://arxiv.org/abs/2204.06125`, `arXiv:2204.06125`.

[56] N. Rostamzadeh, E. Denton, and L. Petrini. Ethics and Creativity in Computer Vision. URL: http://arxiv.org/abs/2112.03111, arXiv:2112.03111.

[57] M. Saif and K. Svetlana. WikiArt Emotions: An Annotated Dataset of Emotions Evoked by Art. pages 05–2018. URL: https://aclanthology.org/L18-1197.

[58] A. Sherstinsky. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. 404:132306. URL: http://arxiv.org/abs/1808.03314, arXiv:1808.03314, doi:10.1016/j.physd.2019.132306.

[59] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe. First Order Motion Model for Image Animation. URL: http://arxiv.org/abs/2003.00196, arXiv:2003.00196.

[60] K. Simecek. Beyond Narrative: Poetry, Emotion and the Perspectival View. 55(4):497–513. URL: https://academic.oup.com/bjaesthetics/article-lookup/doi/10.1093/aesthj/ayv041, doi:10.1093/aesthj/ayv041.

[61] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. URL: http://arxiv.org/abs/1409.1556, arXiv:1409.1556.

[62] A. Spector, P. Norvig, C. Wiggins, and J. M. Wing. *Data Science in Context: Foundations, Challenges, Opportunities.*

[63] M. M. Suh, E. Youngblom, M. Terry, and C. J. Cai. AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–11. ACM. URL: https://dl.acm.org/doi/10.1145/3411764.3445219, doi:10.1145/3411764.3445219.

[64] G. D. Taylor. *When the Machine Made Art: The Troubled History of Computer Art.* International Texts in Critical Media Aesthetics. Bloomsbury Academic.

[65] Timnit Gebru [@timnitGebru]. "Mr. Altman thought the essay nailed a big problem: In the face of the “internet mob” that guarded against sexism and racism..." https://nytimes.com/2021/02/13/technology/slate-star-codex-rationalists.html I wish we could tell stories of how misogynists high up at Open AI drove out women from various institutions. URL: https://twitter.com/timnitGebru/status/1520895272980664321.

[66] O. US EPA. Greenhouse Gas Equivalencies Calculator. URL: https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator.

[67] Vivid Void [@VividVoid_]. DALL-E is breaking my heart. AI art is about to lay utter waste to traditional visual art forms. This will be so much more destructive than what the Internet did to music. It will be a technological conquest of one of

the great human avenues of spiritual transformation. https://t.co/9N3CxnK6oa. URL: https://twitter.com/VividVoid_/status/1512140765656338432.

[68] P. Wang. Lucidrains/DALLE2-pytorch. URL: https://github.com/lucidrains/DALLE2-pytorch.

[69] Z. J. Wang, R. Turko, O. Shaikh, H. Park, N. Das, F. Hohman, M. Kahng, and D. H. Chau. CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization. 27(2):1396–1406. URL: http://arxiv.org/abs/2004.15004, arXiv:2004.15004, doi:10.1109/TVCG.2020.3030418.

[70] Z. Xiao, K. Kreis, and A. Vahdat. Tackling the Generative Learning Trilemma with Denoising Diffusion GANs. URL: http://arxiv.org/abs/2112.07804, arXiv:2112.07804.

[71] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He. AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks. URL: http://arxiv.org/abs/1711.10485, arXiv:1711.10485.

[72] S. Zhang, S. Roller, N. Goyal, M. Artetxe, M. Chen, S. Chen, C. Dewan, M. Diab, X. Li, X. V. Lin, T. Mihaylov, M. Ott, S. Shleifer, K. Shuster, D. Simig, P. S. Koura, A. Sridhar, T. Wang, and L. Zettlemoyer. OPT: Open Pre-trained Transformer Language Models. URL: http://arxiv.org/abs/2205.01068, arXiv:2205.01068.