Artist-Guided Neural Networks – Automated Creativity or Tools for Extending Minds?

Varvara Guljajeva, Mar Canet Sola,[1] Isaac Clarke

Introduction

It is claimed that recent advancements in AI, such as CLIP-based products Midjourney and DALL-E, are supposed to augment our creativity. For the first time, it does not sound so absurd that artists can find themselves out of jobs (Nicholas 2017). Not that artists would have ever had a secure and stable job, but deep learning (DL) tools might eventually lead to losing some commercial commissions. However, such thinking relies on a modern art approach where skills are in the centre of attention and not the conceptual idea. Quoting Lev Manovich: "Since 1970 the contemporary art world has become conceptual, ie focused on ideas. It is no longer about visual skills but semantic skills". (2022, 62) Echoing Aaron Hertzmann, once painters were in a similar situation when photography was invented and took over the niche of portrait-making. Then visual artists had to re-invent themselves and re-think the meaning of painting. And photography had to wait another 40 years until it got recognized as an artistic medium (Hertzmann 2018).

Computer art emerged with the invention of the computer. Artists, such as Vera Molnar and Manfred Mohr, created their first computer-generated artworks in the 1960s using scientific lab computers at night when they were not used by scientists. Early computer artists were re-purposing a machine for artistic use and writing code to make art on it. Since the creation process was mediated by a computer, it may seem to the general audience that the artists were simply pressing a button and the computer doing art for them. Hence, the question of authorship emerged: is the artist a machine or human?

Paradoxically, today with the appearance of neural networks (NN) and their creative applications, the same question re-appears. For example, Aaron Hertzmann has written several articles arguing that people do art and not computers (Hertzmann 2018, 2020). Lev Manovich also describes how AI-generated images that im-

1 Equal contribution.

itate realist and modernist paintings are claimed to be art (Manovich 2022). At the same time, experimental art forms, like installation, interactive, performance and sound art, are often overlooked unless they are promoted by a large corporation.

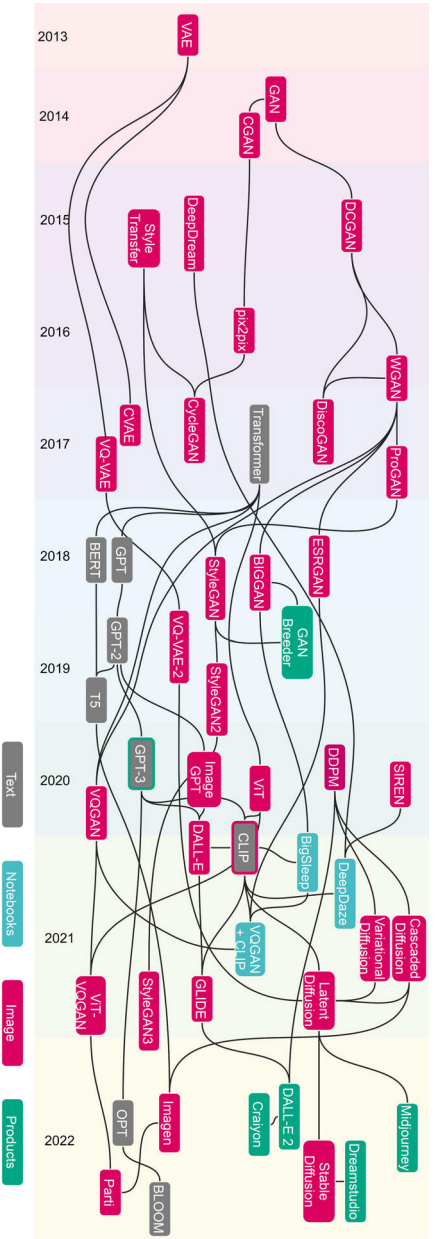
Instead of re-telling a short but very dense history of DL technology development, in the next section, we focus on the appearance of NN tools that raised interest among the artists that led to meaningful artwork production.

Historical overview of deep learning development

DL is a subset of machine learning (ML) using Deep Neural Networks (DNN) to learn underlying patterns and structures in large datasets. In 2012, a DNN designed by Alex Krizhevsky outperformed other computer vision algorithms to achieve the new state of the art in the ImageNet Large Scale Visual Recognition Challenge (Krizhevsky / Sutskever / Hinton 2017). This model, AlexNet, signalled the start of a new DL era. Over the past decade, DNNs have continued to grow in size and complexity, and are now used in a wide range of tasks including Computer Vision, Natural Language Processing, and even playing board games like Go. As AI technology has developed and become more prevalent in real-world systems, artists have been exploring its limits and potentials, adapting these models to their own practices.

As the number of scientific publications on artificial intelligence grows exponentially (Krenn et al. 2022), and the artistic interest grows alongside, it is useful to map out the influential papers, and related applications, to help track the evolution of the AI-Art space in relation to the technological advances. Figure 1 shows a timeline of the development of generative models for images and text. Using this diagram we can make a few observations on the past ten years: the dominance of GANs for image generation, the influence of the Transformer on Large Language Models (LLM), and the growing interest in multi-modal approaches and translation models. The starting period of image generation using DNNs can be traced back to the creation of the Variational Auto-Encoder (VAE) (Kingma / Welling 2013), and the Generative Adversarial Network (GAN) (Goodfellow et al. 2014). These models showed different ways in which a neural network can be trained on a large dataset, and then used to generate outputs that resemble but do not copy the original dataset.

Fig. 1: Timeline of creative deep learning development.



For much of the past decade, GAN art has been a dominant and defining element of AI Art. GANs are trained using a competitive lying game, played by two players: the Generator and the Discriminator. The Generator wins by making an image that the Discriminator thinks is from the original dataset. The Discriminator wins by successfully identifying which images the Generator has made. By playing this game repeatedly, both sides slowly learn when they have been fooled and remember information so they do not fall for the same tricks again. The Generator gets better at making images, and the Discriminator gets better at detecting these fakes. At the end of the game we are left with a Generator that is very good at generating new images, with the qualities and style of our original inputs.

Image-to-Image Translation with Conditional Adversarial Nets (Isola et al. 2016), also known as pix2pix, showed a process of converting one type of image into another type. Mario Klingemann's work *Alternative Face*[2] used the pix2pix model with a dataset of biometric face markers and the music videos of the singer François Hardy. This allowed him to control the movement of the face with this form of digital puppetry, which he then demonstrated by transferring the facial expressions of the political consultant Kellyanne Conway onto Hardy's face as she talks about "alternative facts".

In 2015, on the Google research blog, the post *Inceptionism: Going Deeper into Neural Networks* (Mordvintsev / Olah / Tyka 2015) described a tool developed by researchers attempting to understand how image features are understood in the hidden layers of the NN. Alongside this post they released a tool called DeepDream. This model enhances an image with the NN's attempts to find the features of the dataset it was trained on. The creative use of DeepDream was proposed by the authors in the original article "It also makes us wonder whether NNs could become a tool for artists – a new way to remix visual concepts – or perhaps even shed a little light on the roots of the creative process in general". DeepDream's psychedelic imagery quickly caught the attention of the internet and of artists around the world, resonating with those interested in understanding the cross-over between biological and neurological constructions of images. Memo Atken's work *All Watched Over By Machines Of Loving Grace: Deepdream edition*[3] for an exhibition of DeepDream artworks in 2016, hallucinated over an aerial photograph of the GCHQ headquarters. This work raises questions around the motivations of the organisations funding the development of artificial intelligence, and in doing so make the dreamlike qualities a little more nightmarish.

CycleGAN continued with the problem of image-to-image generation shown in pix2pix, but removed the requirement of aligned image pairs being needed for train-

2 https://underdestruction.com/2017/02/04/alternative-face/

3 https://www.memo.tv/works/all-watched-over-by-machines-of-loving-grace-deepdream-edition

ing (Zhu et al. 2017). Instead a set of source images and a set of target images that aren't directly related can be used. This has the advantage that it is simpler to scale to larger datasets, making the process more accessible for artists. Helena Sarin has been using CycleGAN for a number of years, and recently in *Leaves of Manifold*[4] she collected and photographed thousands of leaves to build her own training dataset, and then implemented a custom pipeline with changes that improve results when working with smaller datasets. This personalised approach in crafting the models resonates with the hand-made, collaged aesthetic of the images generated.

Other notable developments to GANs brought improvements to image quality and resolution (Karras et al. 2017). In late 2018, the release of StyleGAN (Karras 2021), a model built on a combination of ideas from Style Transfer and PGGAN, demonstrated very convincing images of human faces. In his article "How to recognize fake AI-generated Images", the artist Kyle McDonald (McDonald 2018) investigated the images generated by StyleGAN, and highlighted the visual artefacts he found. At a glance these images look like photographs, but on closer inspection irregularities such as patches of straight hair, misaligned eyelines, or mismatched earrings reveal the difficulties GANs have in managing "long-distance dependencies" in images.

In 2017 the paper *Attention Is All You Need* (Vaswani et al. 2017) proposed a new network architecture called the Transformer. This model addressed the long-distance dependency issue in RNNs and CNNs by rethinking how we could handle sequences. Rather than looking at a sentence word by word, the Transformer observes the relationship between all elements of the sequence simultaneously. Being able to better handle long distance dependencies meant the Transformer was appropriate for natural language generation. Artists and poets such as Allison Parrish have explored the use of VAEs for short text generation, but with the emergence of LLM passages of long, coherent texts could be generated (Brown et al. 2020). As dataset sizes increased, along with hardware costs for training these large models, they have become harder for individuals to train themselves, and the mode of interaction has shifted from curated datasets and homemade scripts, to web APIs and third party services. While it is more difficult to participate in the training process, the availability of services and interfaces provides new ways of working with these models that can produce less technical and more playful approaches. For example, Hito Steyerl used GPT-3 to create *Twenty-One Art Worlds: A Game Map* (2021) and described the process as "fooling around" with GPT-3 to write descriptions of different Art Worlds, using the tool in the process of world building. In the resulting text it is difficult to distinguish which words may have been written by Steyerl and which were written by GPT-3.

The learnings from Large Language Models for text generation were soon applied to image generation (Dosovitskiy et al. 2021), and the simultaneous release of

4 https://www.nvidia.com/en-us/research/ai-art-gallery/artists/helena-sarin/

CLIP (Radford et al. 2021) and DALL-E (Ramesh et al. 2021) signalled the start of a new era of image generation. Although the DALL-E model was not released, CLIP was made available to the public, and the model was quickly adopted by AI artists who applied the idea of CLIP guidance to various image generation techniques. Ryan Murdock produced the colab notebooks DeepDaze[5], combining CLIP and SIREN, and BigSleep[6], combining CLIP and BIGGAN, which were subsequently adapted by Katherine Crowson in the widely distributed VQGAN+CLIP notebooks[7].

The paper Denoising Diffusion Probabilistic Models (Ho / Jain / Abbeel 2020) introduced a different method for creating generative models. This technique trains a model by adding increasing amounts of noise to an image and then having the model remove the noise, resulting in a model that can generate images from only noise. Diffusion models, when combined with CLIP or other conditioning processes, enable much faster text-to-image processing.

The popularity and accessibility of these techniques was further raised by the product release of DALL-E 2 (Ramesh et al. 2022) and Midjourney[8]. Midjourney became so popular it is now the largest Discord server with over 5 million members as of writing. Following the releases of these products, open source models such as Stable Diffusion have also been developed. There are many benefits of using free and open source models for artists. Being able to modify code and develop on your own hardware allows the artist to pursue their own experimental approaches, not restricted to the interface designed by a service provider.

The artist's involvement in generating new images with these models is vastly different to working with GANs. Rather than building custom datasets and training models, instead the focus has shifted to writing prompts that can generate the images the artist wants to find, and designing interfaces for exploring these prompts and their translations. This artist Johannez coined the term "Promptism" (Herdon / Dryhurst 2022) for this art practice, and wrote a humorous Prompist manifesto using GPT-3. Against a backdrop of models trained on hundreds of millions of images scraped from the internet, the manifesto asserts "The prompt must always be yours" (Johannez 2021).

Artist-Guided Neural Networks

Many papers discuss AI from the point of view of creativity taking mostly one position of two: either AI as an amazing tool for artists and creativity, or AI is seen as

5 https://github.com/lucidrains/deep-daze
6 https://github.com/lucudrains/big-sleep
7 https://github.com/EleutherAI/vqgan-clip
8 https://www.midjourney.com

something negative in art. It is easy to see that people from the industry advocate for the first position, and theory scholars for the second one. But, how do practitioners see contemporary AI technology themselves? And in which ways AI is deployed in art practice? Hence, it is not the focus of this paper to discuss whether AI can make art, but rather how AI can be useful for artists and what new ideas it can offer. By using practice-based research methodology, we decode the role of AI tools in artistic practice and trace the evolution of such artistic work. In this paper, the practice of artist duo Varvara & Mar was used as a case study, which provided us with the insides in this research.

In this chapter, we explore in various ways in which AI was deployed in creative practice, dividing it into four categories based on medium: synthetic image, synthetic text, synthetic form, and translation models. From the view of the practitioner, the limitations, new possibilities, and change in production processes are discussed.

Synthetic Image

Our exploration of DL started with image generation in 2017 when we used Google Deep Dream algorithm. The idea behind the *Neuronal Landscapes* project[9] was to imagine how the Estonian landscape will look like in 100 years' time (commission work for Estonian History Museum). Since machine mysterious synthetic vista began to be prevalent, the artwork offers an experience of what it is like to see the environment through machine eyes, and to be immersed into an endless hallucinated simulacra of a neural net. Our intention here was to go beyond a still image and achieve immersiveness depicting Estonian society's evolution throughout time: from forest and farm land, towards urbanisation and increasing digitalization. For that purpose, a 360° VR video (00:09:39 long) was created. First, the video material was filmed and then edited. Filming was done with two 360° cameras carried by a drone to achieve a seamless, continuous shot, without a visible camera. After the stabilisation of the image, each frame was passed through the DeepDream algorithm. The rendering process took 30 days' time on two (for that time) powerful machines with Nvidia TitanX GPUs working in parallel. Although we could change parameters of the algorithm to slightly customise the effect, the imprint of the algorithm was still very present. At the end, Google Deep Dream became sort of a generative filter to the images.

In the next art project, ProGAN was deployed. For the first time we worked with datasets and training GAN models. *PlasticLand* (2019)[10] talks about plastic waste and

9 https://var-mar.info/neuronal-landscapes/

10 https://var-mar.info/plasticland/

ecological problems this material causes. We composed four different datasets of images of layered plastics in our planet: landfills, plastic on top of water, plastic underwater, and plastiglomerates. The ProGAN model was trained on a local machine using pyTorch and took a week to train, and we used part of the images generated during training to create a video composition.

A metal totem displaying those synthetic, as plastic is, layers, we draw attention not only to the problem of waste but also question whether AI has some similarity with this material. Since the invention of plastic, this material was applied almost everywhere because of its perfect qualities, until we realised that it is not sustainable and ecology-friendly. Will a similar story happen with AI?

From the practice-based research perspective, this artwork shows artists' desire to move from a still to moving image and towards sculptural form that is held back by the early stage of machine learning technology: low resolution images jumping from one frame to another.

The next artworks *POSTcard Landscapes from Lanzarote I (00:18:37)* and *II (00:18:40)*[11] in 2021 demonstrate the artist's ability to create video works with StyleGAN2 (see figure 2). The hypnotic appearance of these works, where one frame morphs naturally into another, shows the artists' ability in guiding the outputs of the NN. Vector curation and composition of a journey through the latent space, created by training the model on specific datasets of 2000+ images, were crucial and integral parts of the artistic process.

The artwork talks about critical tourism and how circulation of images representing touristic gaze overpower the nature of seeing. In the words of Jonas Larsen "'reality' becomes touristic, and item for visual consumption" (2006: 241–257). Hence, we scraped, where licence allowed, the location-tagged images from Flickr and composed two datasets of photos categorised as tourism or landscape.

As we have written elsewhere,

> the art project consists of two videos representing a journey of critical tourism through the latent space of AI-generated images using StyleGan2. Later the images are composed into latent interpolations that take the form of smoothly progressive videos. The two videos are random walks in the latent space of the Stylegan2 trained models, creating a cinematic synthetic space. The audiovisual piece shows an animated image through the melted liquid trip of learning acquired from the dataset composed of static images. The video flows from point to point, generating new views and meaning spaces through the latent space's movement. The audio was created after the video was generated in response to the visual material to complete the art piece. (Guljajeva / Canet Sola 2022a)

11 https://var-mar.info/postcard-landscapes-from-lanzarote/

The sound for local or landscape view was created by a sound artist from Lanzarote, Adrian Rodd, who aimed to give a socio-political voice to the piece. In contrast, the sound design created by Taavi Varm is a soundscape replying to touristic gaze. We aimed to initiate collaborations with others but also to experiment with human-AI co-creation. In a similar vein is the artwork *Phantom Landscapes of Buenos Aires* (00:20:00, 2021)[12], with sound work by Cecilia Castro.

Our last experiment with GAN models *Synthetic Scapes of Tartu* (00:10:00, 2022), demonstrates a different approach. Taking a dataset composed from our own video footage (flaneur walks), we first produced the sound (a composition by Taavi Varm, Ville MJ Hyvönen with piano by J. Kujanpää) and used this to inform the direction of the video. The result was a sound-guided AI-generated visual output.

Fig. 2: Stills from POSTcard Landscapes from Lanzarote I and II

Synthetic Text

In this section, we focus on artwork incorporating AI text generation as part of the artistic concept. Our journey to text generation started with the online participative theatre project ENA[13] and ended with a hand-bound publication.

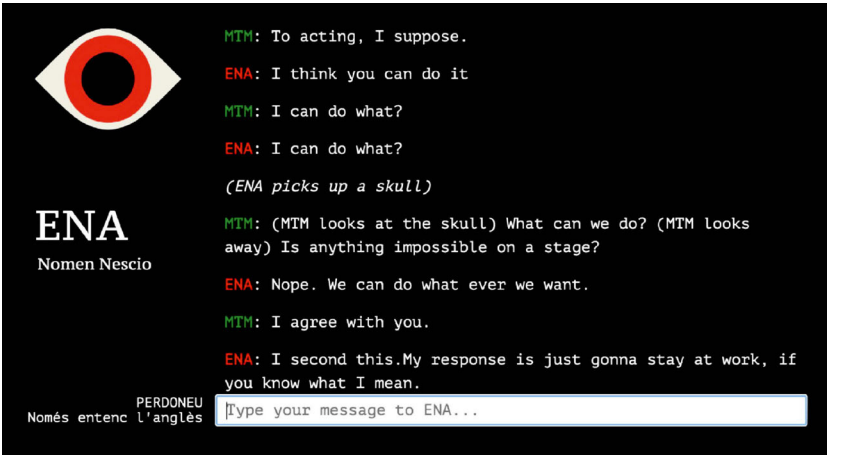
During the first lockdown in May 2020, together with theatre maker Roger Bernat, we created an online participative theatre piece ENA on the website of Theater Lliure in Barcelona. ENA is a generative chatbot that talks to its audience, and together (AI and audience), they make theatre (see figure 3). As we have described

12 https://var-mar.info/phantom-landscapes-of-buenos-aires/

13 https://var-mar.info/ena/

before: Although in the description of the project it was stated explicitly that people were talking to a machine, multiple participants were convinced that on the other side of the screen another human was replying to them – more precisely the theatre director himself, or at least an actor". (Guljajeva / Canet Sola 2021)

Fig. 3: A screenshot of ENA user interface and published book.



Analysing synthetic books, Varvara Guljajeva has stressed the importance of human input in the AI text-generation systems (2021). In addition, one also needs to guide the audience participation and interaction with the chatbot. For this purpose, we have adopted the traditional theatre method for guiding actors, as a way to guide the audience, and thus, the bot, too. Stage directions were used as a guiding method, which triggered thematic conversation and offered meaningful dialogue between humans and the AI system. We found the conversations so meaningful that we decided to publish a book that contains all the conversations with ENA.

With this project, we learned that it is essential to guide neural networks via audience interaction. In order to do this, it is also necessary to guide the audience. Without audience interaction guidance, it is nearly impossible to achieve meaningful navigation of NNs.

Translation models

This category focuses on translation models that enable interactive and installation-based formats. Translation refers to the conversion of mediums, or as we put it, translation of semiotic spaces. To illustrate this, we introduce *Dream Painter*,[14] an art installation that translates the audience's spoken dreams to a line-drawing produced by a robot (see figure 4). As described earlier: "Dream Painter is an interactive robotic art installation that explores the creative potential of speech-to-AI-drawing transformation, which is a translation of different semiotic spaces performed by a robot. We extended the AI model CLIPdraw which uses CLIP encoder and the differential rasterizer diffvg for transforming the spoken dreams into a robot-drawn image". (Canet Sola / Guljajeva 2022) "Design- and technology-wise, the installation is composed of four larger parts: audience interaction via spoken word, AI-driven multicolored drawing software, control of an industrial robot arm, and kinetic mechanism, which makes paper progression after each painting has been completed. All these interconnected parts are orchestrated into an interactive and autonomous system in the form of an art installation […]". (Guljajava / Canet Sola 2022b) Out of all the projects discussed, this was the most difficult to realise.

In this project, we investigated how guidance of NNs could be interactive and real-time instead of non-interactive and pre-determined, as shown in previous examples of our work. It is important to notice that methods, such as dataset composition and output curation were not used in this case. In fact, visual output curation is totally missing. The artists created an interactive system to be experienced and discovered by the audience. This means the audience determines the output. Instead of curating a dataset, a CLIP model is used that can produce nearly real-time output

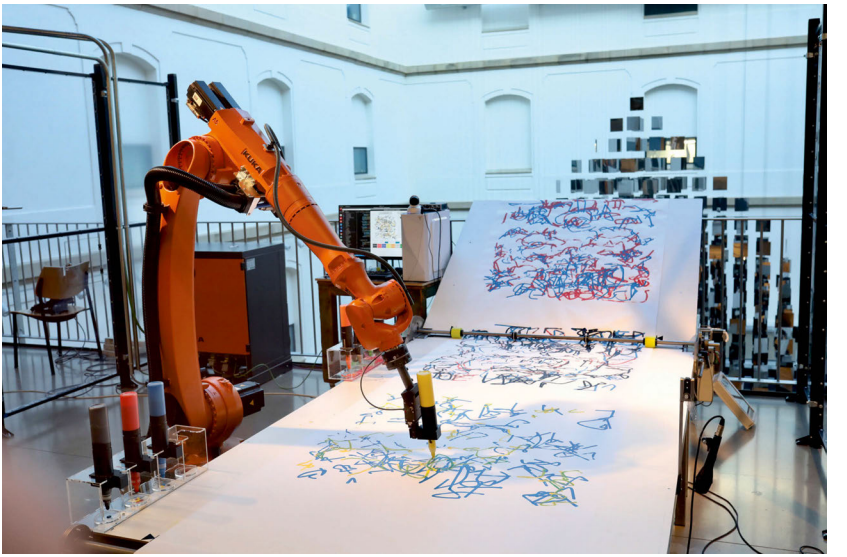
14 https://var-mar.info/dream-painter/

guided by a text prompt. As we have written earlier: "Translation of semiotic spaces, such as spoken dreams to AI-generated robot-drawn painting, allowed us to deviate from image-to-image or text-to-text creation, and thus, imagine different scenarios for interaction and participation". (Guljajava / Canet Sola 2022b)

This project indicates our search for transformative outputs of AI technology, and thus, shows the evolution in practice. By extending available DL tools and combining with other technology, for example, text-to-speech models, real-time industrial robot control, and physical computing, it offered an interactive robotic and kinetic experience of NN latent space navigation. This contributes towards the explainability of AI because the audience could experience how the words affected the drawing, and which concept triggered which outcome.

Being inspired by Sigmund Freud's work on the interpretation of the human mind while unconscious, we speculatively ask if AI is powerful enough to understand our dreamworld. Through practice we question the capacities of NNs and investigate how far we can push this technology in the art context. This artwork allows the audience to experience the limits of concept-based navigation with AI. The system is unable to interpret and can only illustrate our dreams. It cannot understand the prompt semantically and only gets the concepts.

Fig. 4: Kuka industrial robot painting audience's dreams. Installation view.



Synthetic Form

In this section, we ask how artists can guide neural networks when creating volumetric forms, and what happens when AI meets materiality. After working for a while with DL tools that produce 2D outputs, it is an obvious step to explore possibilities to produce 3D results. To our surprise, it was not an easy task to find the solution (Oct 2021). *Psychedelic Forms*[15] is a series of sculptures produced in ceramics and recycled plastic through which we investigated the possibilities of AI in producing physical sculptures. The project re-interprets antique culture in the contemporary language and tools.

Following the same paradigm shift as in the previous section, text2mesh is a CLIP-based model that does not require a dataset, but a 3D object and text prompt as input. Hence, the model actually does not create a 3D model but stylises the inserted one. And this is guided by inputted text.

We decided to go back to the origins, in terms of ancient sculptures and material selection. Although it was said that there was no dataset, we still had a collection of 3d models of ancient sculptures because, by far, not all produced the desirable output. In this sense, there was definitely an output curation present in the process.

The criteria for selection were the following: first, the form had to be intriguing, and second, it should be possible to produce it in material afterwards. It was clear that we had to modify each model because the physical world has gravity, and the DL model does not take this into account. Some generated models were discarded because they were seen as not-fixable, although interesting in their shape.

The process demonstrated here is quite an unusual way to create an object. After extensive experimentation with the tool, we learned how certain words triggered certain shapes and colours. This knowledge gave us a chance to treat text prompts as poetic input. Thus, we created short poems to guide NN. The best ones survived as titles and are reflected in the forms.

The artists did not strictly follow the original model but took the creative liberty to modify the shape and determine the colour by manually glazing the sculptures. The dripping technique was used for colouring the sculptures. This served as a metaphor for liquid latent space and the psychedelic production process (this was the artists' inner feeling about the creative process because they did not know what results would be achieved in the end).

Sometimes, AI-generated vertex colouring was taken as inspiration, sometimes totally ignored. Nevertheless, digital sculptures were exhibited alongside the physical ones to underline the transformation and human role in the creative process. Although ceramic sculptures were 3D printed in clay, the fabrication process had to follow the traditional way of producing pottery (see final sculptures in figure 5).

15 https://var-mar.info/psychedelic-forms/

Since we had never engaged in ceramics before, the whole production process felt psychedelic: unexpected NN processes led to transformation by numerical, physical, and chemical processes, all guided by both the artists and chance. Hence, the art project highlights the relationship between different agencies.

In the end, we can say that AI is not prepared for the physical world. It created nice images, but when one wants to materialise the output, it requires considerable additional work. However, those extra processes were very rewarding and creative in our case. In this project, AI served as an inspiration or a departing point more than anything else. In other words, the experimental phase of technology is necessary for experimental practices, and this can lead to the creation of a new production pipeline. The fine line between control and chance when guiding the neural networks and related processes is likely the main creative drive for the artists.

Fig. 5: Ceramic sculptures guided by 3D object and text prompt, 3D printed in clay, and glazed manually. From left to right: Psychedelic Angel (Venus), Mermaid in green jelly and pink feather (Nymph), Psychedelic Angel (Venus).

Discussion

According to the media hype around AI, this technology is intelligent enough to create art autonomously (Perez 2018; Vallance 2022). However, the reality is different. According to computer scientist and a co-inventor of Siri Luc Julia, AI does not exist. He advocates for machines' multiple intelligences that often outperform humans. However, machine intelligence is limited and discontinuous compared to human

intelligence (Julia 2020). Therefore, it is vital to have artistic practices around this technology, as a counterbalance to the AI fantasies served by the industry and mass media.

We see AI as a creative tool with its own possibilities and limitations, which can stimulate artists' creativity through unexpected outputs. Research has shown that tool-making expands human cognitive level and constitutes evolution in culture (Stout 2016). Similarly, as a new tool, generative AI could potentially enrich creativity by allowing new production pipelines that can generate unique results.

Coming back to the synthetic images, we can say that all machine-created synthetic image-based works discussed here have particular aesthetics: both with Deep Dream and GAN. Unlike the output of GANs, Deep Dream has a more recognizable style and can be seen more as a filter that transforms every inputted image instead of learning from the given dataset.

Regarding GAN aesthetics, such visual appearance is inherited from two entities to a large extent: the dataset and the model itself. GANs have a particular footprint, as seen in all works produced with this model. The visual palette comes from the used datasets. For example, if a dataset is homogeneous (only landscape images), then we will easily recognize landscapes in the generated output. However, if images in the dataset have a lot of visual variation, the output is rather abstract. *POSTcard Landscapes from Lanzarote* illustrate this well. Also, when photos in the dataset look similar, the output will also be similar, as was the case with the *Synthetic Scapes of Tartu* video work where frames from recorded flaneur walks in a city were extracted. When we talk about video works generated with the neural net, then manual guidance of latent space offered more variations than an audio-led approach.

Synthetic image works have encouraged us to work with formats like images and videos that we did not engage in before in our art practice, but we found it exciting working with AI and video. For example, AI video generation has some affordances, like starting and ending can be done in a perfect loop since images are synthetically generated. However, creating real-time AI work is much more complex because some models are too slow. It might take a few minutes to render a single image. The limitations inspire us to devise new solutions and work in new mediums. Moreover, the limitations of the medium has always been a good challenge for our creativity.

Working with GANs or other image-generation tools has become much easier in recent years, although it used to be quite difficult. We must note that for practitioners, easy-to-use tools, such as DALL-E and Midjourney, offer little creative freedom, and thus, are less attractive to the artists. Those products tend to instrumentalize the user rather than the other way around. At the same time, open source models offer more creative freedom and enable broader use of artistic ideas.

The work with generated text demonstrates that AI is not context-aware but maps concepts automatically without understanding semantics. More importantly, as shown in the *ENA* project the audience must also be guided alongside the AI. In

the case of *ENA*, stage directions were used, and in the *Dream Painter* project, the concept of dream telling was applied to guide the participants who in turn guided the neural net through their interaction, creating a chain reaction. Navigating concepts in latent space is artistically interesting and inspiring, this was especially evident when working with form. The artists went beyond semantics and learned how to guide neural networks with a text prompt and 3D object.

The presented practice represents a paradigm shift in machine learning, moving away from composing datasets for GANs and toward translating semiotic spaces enabled by diffusion models. The evolution in practice shows how artists discover and learn to work with the DL toolset, embracing its possibilities and limitations. In the case of practice-based research, practice can be seen as a lab for testing artistic ideas with technology through chance until control is encountered.

Conclusion

In this article, we have summarised DL development from the perspective of artists' interests concentrating on the image, video, text, 3D object generation, and translation models. We applied practice-based research methodology to investigate the role and possibilities of recent co-creative AI tools in artistic practice.

It is difficult to keep pace with AI development. In less than a decade, we have gone from blurry black-and-white faces to impressive high-resolution images guided by text prompts. The user level has gone from difficult to easy, which on one side, broadens possibilities for creation, but on another, it diminishes experimentation and creativity, since AI outputs seem ready-made. This is also demonstrated by the explorative nature of the body of work presented here.

Furthermore, it was noticed that creative AI, especially GAN models, have recognizable aesthetics, which in the long run, become repetitive. This led to the change of tools by the artists. The curation of datasets, models, and outputs, along with NN guidance, have become the toolset of an artist working with AI. Finally, these models can generate multitudes of outputs, but the art is giving the right input to guide the desired output and selecting the results that best serve the concept

As Andy Warhol had envisioned in 1963, eventually, art production will become mechanised and automated. In his own words: "I want to be a machine", which was also a reflection on that time's vast industrialization process. Resonating with today's deep learning age: I want my machine to do art.

Bibliography

Arjovsky, Martin / Soumith Chintala / Léon Bottou (2017): "Wasserstein GAN". *ArXiv*. 1701.07875 [stat.ML].

Brown, Tom B. / Benjamin Mann / Nick Ryder / Melanie Subbiah / Jared Kaplan / Prafulla Dhariwal / Arvind Neelakantan / Pranav Shyam / Girish Sastry / Amanda Askell / Sandhini Agarwal / Ariel Herbert-Voss / Gretchen Krueger / Tom Henighan / Rewon Child / Aditya Ramesh / Daniel M. Ziegler / Jeffrey Wu / Clemens Winter / Christopher Hesse / Mark Chen / Eric Sigler / Mateusz Litwin / Scott Gray / Benjamin Chess / Jack Clark / Christopher Berner / Sam McCandlish / Alec Radford / Ilya Sutskever / Dario Amodei. (2020): "Language Models are Few-Shot Learners". *ArXiv*. 2005.14165 [cs.CL].

Canet Sola, Mar / Varvara Guljajeva (2022): "Dream Painter: Exploring creative possibilities of AI-aided speech-to-image synthesis in the interactive art context". *Proc. ACM Comput. Graph. Interact. Tech*. 5(4), Article 33, doi:10.1145/3533386

Dosovitskiy, Alexey / Lucas Beyer / Alexander Kolesnikov / Dirk Weissenborn / Xiaohua Zhai / Thomas Unterthiner / Mostafa Dehghani / Matthias Minderer / Georg Heigold / Sylvain Gelly / Jakob Uszkoreit / Neil Houlsby (2020): "An image is worth 16x16 words: Transformers for image recognition at scale". *ArXiv*. 2010.11929 [cs.CV].

Gatys, Leon A. / Alexander S. Ecker ./ Matthias Bethge (2015): "A neural algorithm of artistic style". *ArXiv*. 1508.06576 [cs.CV].

Goodfellow, Ian J. / Jean Pouget-Abadie / Mehdi Mirza / Bing Xu / David Warde-Farley / Sherjil Ozair / Aaron Courville / Yoshua Bengio (2014): "Generative Adversarial Networks". *ArXiv*. 1406.2661 [stat.ML].

Guljajeva, Varvara / Mar Canet Sola (2021): "ENA: Participative Art Forms During Pandemics". Ed. Livia Nolasco Rozsas / Borbala Kalman. In *HyMEx 2021 (Hybrid Museum Experience Symposium Proceedings)*. Budapest, 77–82.

Guljajeva, Varvara / Mar Canet Sola (2022a): "POSTcard Landscapes from Lanzarote". *Creativity and Cognition (C&C '22)*, June 20–23, 2022, Venice, Italy. ACM, New York, NY, USA. doi:10.1145/3527927.3531191

Guljajeva, Varvara / Mar Canet Sola (2022b): "Dream Painter: An Interactive Art Installation Bridging Audience Interaction, Robotics, and Creative AI". *Proceedings of the 30th ACM International Conference on Multimedia (MM '22)*. Association for Computing Machinery, New York, NY, USA, 7235–7236. doi:10.1145/3503161.3549976

Guljajeva, Varvara (2021): "Synthetic Books". *ARTECH 2021: 10th International Conference on Digital and Interactive Arts*. New York, NY, USA: ACM. doi: 10.1145/3483529.3483663.

Herndon Holly / Mathew Dryhurst (2022): "Infinite Images and the latent camera". https://mirror.xyz/herndondryhurst.eth/eZG6mucl9fqU897XvJs0vUUMnm5OITpSWN8S-6KWam Y (13 December 2022).

Hertzmann, Aaron (2018): "Can Computers Create Art?" *Arts*. MDPI AG, 7(2), 18. doi: 10.3390/arts7020018.

Hertzmann, Aaron (2020): "Computers do not make art, people do". *Communications of the ACM. Association for Computing Machinery (ACM)*, 63(5), 45–48. doi: 10.1145/3347092.

Ho, Jonathan / Ajay Jain / Pieter Abbeel (2020): "Denoising Diffusion Probabilistic Models". *ArXiv*. 2006.11239 [cs.LG].

Isola, Phillip / Jun-Yan Zhu / Tinghui Zhou / Alexei A. Efros (2016): "Image-to-image translation with conditional adversarial networks". *ArXiv*. 1611.07004 [cs.CV].

Johannez (2021): "The promptist manifesto". *https://deeplearn.art/the-promptist-manifesto/* (13 December 2022).

Karras, Tero / Timo Aila / Samuli Laine / Jaakko Lehtinen (2017): "Progressive growing of GANs for improved quality, stability, and variation". *ArXiv*. 1710.10196 [cs.NE].

Karras, Tero / Samuli Laine / Timo Aila (2021): "A style-based generator architecture for generative adversarial networks". *IEEE transactions on pattern analysis and machine intelligence. Institute of Electrical and Electronics Engineers (IEEE)*, 43(12), 4217–4228. doi: 10.1109/TPAMI.2020.2970919.

Kingma, Diederik P. / Max Welling (2013): "Auto-Encoding Variational Bayes". *ArXiv*. 1312.6114 [stat.ML].

Krenn, Mario / Lorenzo Buffoni / Bruno Coutinho / Sagi Eppel / Jacob Gates Foster / Andrew Gritsevskiy / Harlin Lee / Yichao Lu / Joao P. Moutinho / Nima Sanjabi / Rishi Sonthalia / Ngoc Mai Tran / Francisco Valente / Yangxinyu Xie / Rose Yu / Michael Kopp (2022): "Predicting the Future of AI with AI: High-quality link prediction in an exponentially growing knowledge network". *ArXiv*. 2210.00881 [cs.AI].

Krizhevsky, Alex / Ilya Sutskever / Geoffrey E. Hinton (2017): "ImageNet classification with deep convolutional neural networks". *Communications of the ACM. Association for Computing Machinery (ACM)*, 60(6), 84–90. doi: 10.1145/3065386.

Larsen, Jonas (2006): "Geographies of tourist photography". *Geographies of Communication: The Spatial Turn in Media Studies*. Gothenburg, 241–257.

Manovich, Lev (2022): "AI and myths of creativity". *Architectural design* 92(3), 60–65. doi: 10.1002/ad.2814.

McDonald, Kyle (2018): "How to recognize fake AI-generated images". *Medium*. *https://kcimc.medium.com/how-to-recognize-fake-ai-generated-images-4d1f6f9a2842* (13 December 2022).

Michel, Oscar / Roi Bar-On / Richard Liu / Sagie Benaim / Rana Hanocka (2022): "Text2Mesh: Text-driven neural stylization for meshes". 2022

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi: 10.1109/cvpr52688.2022.01313.

Mordvintsev, Alexander / Christopher Olah / Mike Tyka (2015): "Inceptionism: Going deeper into neural networks". *https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html* (13 December 2022).

Nicholas, Gabriel (2017): "These Stunning A.I. Tools Are About to Change the Art World". *https://slate.com/technology/2017/12/a-i-neural-photo-and-image-style-transfer-will-change-the-art-world.html*. (13 December 2022).

Perez, Sarah (2018): "Microsoft's new drawing bot is an AI artist". *TechCrunch*, 18 January. *https://techcrunch.com/2018/01/18/microsofts-new-drawing-bot-is-an-a-i-artist/* (13 December 2022).

Radford, Alec / Jong Wook Kim / Chris Hallacy / Aditya Ramesh / Gabriel Goh / Sandhini Agarwal / Girish Sastry / Amanda Askell / Pamela Mishkin / Jack Clark / Gretchen Krueger / Ilya Sutskever (2021): "Learning transferable visual models from natural language supervision". *ArXiv.* 2103.00020 [cs.CV].

Ramesh, Aditya / Mikhail Pavlov / Gabriel Goh / Scott Gray / Chelsea Voss / Alec Radford / Mark Chen / Ilya Sutskever (2021): "Zero-shot text-to-image generation". *ArXiv.* 2102.12092 [cs.CV].

Ramesh, Aditya / Prafulla Dhariwal / Alex Nichol / Casey Chu / Mark Chen (2022): "Hierarchical Text-Conditional Image Generation with CLIP Latents". doi: 10.1002/ad.2814.

Steyerl, Hito (2021): "Twenty-one art worlds: A game map". *https://www.e-flux.com/journal/121/423438/twenty-one-art-worlds-a-game-map/* (13 December 2022).

Stout, Dietrich (2011): "Stone toolmaking and the evolution of human culture and cognition". *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366, 1050–1059.

Stout, Dietrich (2016): "Tales of a stone age neuroscientist". *Scientific American*. Springer Science and Business Media LLC, 314(4), 28–35. doi: 10.1038/scientificamerican0416-28.

Vallance, Chris (2022): "'Art is dead Dude' – the rise of the AI artists stirs debate". *BBC*, 13 September. *https://www.bbc.com/news/technology-62788725* (13 December 2022).

Vaswani, Ashish / Noam Shazeer / Niki Parmar / Jakob Uszkoreit / Llion Jones / Aidan N. Gomez / Lukasz Kaiser / Illia Polosukhin (2017): "Attention is all you need". *ArXiv*. 1706.03762 [cs.CL].

Wang, Xintao / Ke Yu / Shixiang Wu / Jinjin Gu / Yihao Liu / Chao Dong / Chen Change Loy / Yu Qiao / Xiaoou Tang (2018): "ESRGAN: Enhanced super-resolution generative adversarial networks". *ArXiv*. 1809.00219 [cs.CV].

Zhu, Jun-Yan / Taesung Park / Phillip Isola / Alexei A. Efros (2017): "Unpaired image-to-image translation using cycle-consistent adversarial networks". *ArXiv*. 1703.10593 [cs.CV].