

<https://artnodes.uoc.edu>

ARTICLE

NODE "MEDIA ARTIVISM"

Content policy and access limitations on commercial neural networks as an incentive to activism

Stanislav Milovidov

HSE University (National Research University Higher School of Economics)

Date of submission: June 2023

Accepted in: December 2023

Published in: January 2024

Recommended citation

Milovidov, Stanislav. 2024. «Content policy and access limitations on commercial neural networks as an incentive to activism». In: Carolina Fernández-Castrillo and Diego Mantoan (coords.). «Media Artivism». *Artnodes*, no. 33. UOC. [Accessed: dd/mm/aa]. <https://doi.org/10.7238/artnodes.v0i33.417696>



The texts published in this journal are – unless otherwise indicated – covered by the Creative Commons Spain Attribution 4.0 International licence. The full text of the licence can be consulted here: <http://creativecommons.org/licenses/by/4.0/>

Abstract

This article employs a case-study method to investigate the activism neural network community concentrated on Twitter (since renamed X), which has been ideologically influenced by the content policy and limitations of OpenAI. Today, many young artists using machine learning technologies in their artworks (Midjourney, Stable Diffusion, Kandinsky) note that despite significant progress in the field of neural network generators of image through prompts present in museums and exhibitions of contemporary digital art, a significant number of artworks are still made chiefly using outdated text-to-image algorithms created in 2021. These neural networks continue to be popular in art to this day. The reasons for the sustainability of such practices can be found in the soft ideological conflict between artists and OpenAI in 2021. At that time, neural networks had not yet become mainstream, and the dominant theme was deep fakes, which became the basis for a comprehensive discussion about the possibilities and consequences of implementing AI algorithms in modern society. A series of scandals related to the work of neural networks alerted businesses, which feared the reputational costs of neural network errors and biases. At the same time, the existing discourse on freedom of speech, thought, and self-expression in contemporary art has led to ideological conflict, as the creators have introduced constraints on tools of artistic expression. Previously, the actions of artists were not moderated by technical means. Thus, the community did not accept this state of affairs, and as

a result of cooperation and “collective intelligence” created, on the GitHub and Google Colab platforms, their own algorithms with open code, with which everyone could carry out their visual experiments. Artists face the ideological question of fighting globalism and anti-progress in art to be outside the system but to riot against it. This process led to a division of artistic practises in neural network art, outlined by media artist Ryan Murdock as a gateway to text-guided visual art by the hacker effort of 2021 or the modern generation of algorithm text-to-images (after 2022).

Keywords

neural network; deep learning; computer art; text-to-image; content policy; digital activism

Políticas y limitaciones de acceso a las redes neuronales comerciales como incentivo al activismo

Resumen

Este artículo emplea un método de estudio de caso para investigar la comunidad de redes neuronales del activismo concentrada en Twitter (desde que se cambió el nombre, de X), influenciada ideológicamente por la política de contenido y las limitaciones de OpenAI. En la actualidad, muchos jóvenes artistas que utilizan tecnologías de aprendizaje automático en sus obras de arte (Midjourney, Stable Diffusion, Kandinsky) señalan que, a pesar del progreso significativo en el campo de los generadores de red neuronal de imágenes a través de indicaciones presentes en museos y exposiciones de arte digital contemporáneo, todavía se sigue haciendo una cantidad significativa de obras de arte mediante algoritmos de texto en imagen obsoletos creados en 2021. Actualmente, estas redes neuronales siguen siendo populares en el arte. Los motivos de la sostenibilidad de dichas prácticas se pueden encontrar en el sutil conflicto ideológico entre los artistas y OpenAI en 2021. En aquel momento, las redes neuronales aún no se habían convertido en algo generalizado, y el tema dominante eran las deep fakes, que se convirtieron en la base de un debate exhaustivo sobre las posibilidades y consecuencias de implementar algoritmos de IA en la sociedad moderna. Una serie de escándalos relacionados con el trabajo de las redes neuronales alertaron a las empresas, que temían los costes de reputación de los errores y sesgos de las redes neuronales. Al mismo tiempo, el discurso existente sobre la libertad de expresión, el pensamiento y la autoexpresión en el arte contemporáneo ha dado lugar a conflictos ideológicos, ya que los creadores han introducido restricciones en las herramientas de expresión artística. Anteriormente, las acciones de los artistas no se moderaban por medios técnicos. Así, la comunidad no aceptó este estado de la cuestión, y fruto de la cooperación y de la «inteligencia colectiva» creó, en las plataformas GitHub y Google Colab, sus propios algoritmos con código abierto, con los que todo el mundo podía realizar sus experimentos visuales. Los artistas se enfrentan a la cuestión ideológica de luchar contra el globalismo y el antiprogreso en el arte para estar fuera del sistema, pero, a su vez, rebelarse en su contra. Este proceso condujo a una división de prácticas artísticas en el arte de redes neuronales, descrito por el artista multimedia Ryan Murdock como una puerta de entrada al arte visual guiado por texto por el esfuerzo del hacker de 2021 o la generación moderna de algoritmos de texto a imágenes (después de 2022).

Palabras clave

red neuronal; aprendizaje profundo; arte informático; texto a imagen; política de contenido; activismo digital

Introduction

The DeepDream algorithm created by Google programmer Mordvintsev (Mordvintsev 2015) produced a new visual image strongly associated with neural networks and determined a change in computer art. This technical solution was invented for testing neural networks. However, hallucinatory video sequences and an aesthetic of an “anomalous creature rising from a psychedelia of data” were created (Blas & Wyman 2017). This inspired artistic experiments with machine learning

and deep learning. Artists saw the potential of computer graphics for generative art and adapted innovations to specific artistic practises.

A natural continuation was the integration of artistic practises of the generative-adversarial networks (GAN). Their architecture consists of two neural networks. The first one generates images, and the other rejects those that do not correspond to the task. This technology became widely known due to many deep fakes that flooded the Internet in 2020 and a series of scandals connected with the errors, bias and the “rationality” of computer systems that already bordered on discrimination and the infringement of human rights (Forbes 2020).

A curious turn occurred in computer art in 2021. Neural networks translated textual information into graphical information (text-to-image). This opportunity appeared in connection with the development of foundation models trained on datasets by more than twelve billion parameters, which allows them to interpret information in natural language. Media artists experimenting with neural networks noticed this technological shift.

This research focuses on first-generation text-to-image algorithms because their formation as an artistic tool was due to the grassroots practices of artists and artist-programmers. During this period, they partially opposed the content policy and constraints imposed by OpenAI, who released the first text-to-image algorithm. As a result, these factors pushed the community into self-organization. For artistic practise, such processes have been described by various researchers as activism (a blend of art and activism), which “has developed as a universal expression. It emerged from urban and graffiti art and situationism, a form of creative forms from the twentieth century” (Alimen *et al.* 2023). However, digital technologies, the Internet, and social media have created a strong foundation for activism “as a dialogical space. Activists can establish networks and experience communities of solidarity” (Medrado & Rega 2023, 18, 82). Social networks have provided unlimited communication between creators and recipients while allowing each recipient to become an information-transmitting agent, creating a delocalized panorama of political art engaged in the present (Asunción *et al.* 2019). As a result, activism engendered a series of freeware and open-source neural networks that left a noticeable mark on computer art in the early 2020s. Artists still use some of these algorithms despite the emergence of next-generation text-to-image neural networks (DALL-E 2, Midjourney, Stable Diffusion).

This research considers the reasons for the soft clash of interests of the art community and OpenAI within the framework of artistic practices using first-generation text-to-image neural networks (e.g. DALL-E). Why did freeware and open-source neural networks foster sustainable art practices compared to commercial ones? Netnographic analysis is used as the reference methodology of the research. Netnography can help a researcher “to access a certain group’s ideology in social media” and study “what specific actions define the group, interpret group norms and values, behaviors and beliefs” (Kozinets & Gambetti 2021, 108).

This approach helps analyse grassroots cooperation between artists and programmers. This activism is concentrated mainly on Twitter (since renamed X), GitHub and Google Colab. Twitter has shown itself as an influencer platform, spawning concepts such as “the Twitter revolution” or “the Twitter diplomacy”. It has also been the most popular social media for IT communities. This fact is discussed in the press, blogs and in academic research directly or indirectly related to this topic. For instance, Singer & Filho (2014) underlined that it “has established that many software developers use Twitter in their work”, and the purpose of their study is to explain “how or why they use it”. Other scholars note that “Twitter, with its open API to crawl, one-sided nature of relationship,

and the retweet mechanism to relay information, offers an unprecedented opportunity for computer scientists” (Kwak *et al.* 2010).

We analysed posts, comments and discussions from the Twitter accounts of OpenAI (@openai) and artists Ryan Murdock (@advadnoun), Gene Kogan (@genekogan), Katherine Crowson (@RiversHaveWings), Vadim Epstein (@eps696), Mario Klingemann (@quasimondo) and some others for 2021 and the first half of 2022. The conflict of interest was one-sided and developed exclusively in the internet space, considering the absence of any official statements from OpenAI. These accounts were chosen because these persons are artistic and technical leaders in the community.

Ryan Murdock created the first indie text-to-image algorithms, BigSleep and Aleph. Katherine Crowson and Vadim Epstein published their own programs (CLIP Guided Diffusion and Aphantasia) which were widely disseminated among computer artists. Artworks based on Aphantasia have been presented by Janelle Shane (*Ten Thousand Miles Away*, 2021), Smadar Lomnitz (*The Artist*, 2021), can be found at exhibitions in the Krasnokholmskaya and GRAUND Solyanka galleries in Moscow (exhib. *The Code of Art*, 2021; *Programmable art*, 2022), and the conference *NeuroIPS 2021*. In 2017, Gene Kogan started the landmark project Abraham AI to create the first autonomous artificial artist. Mario Klingemann is one of the first artists to embrace machine learning technology, awarded an Honorary Mention in the Ars Electronica 2020, and one of the most cited AI artists. Most of them are pioneers in machine learning in art, follow each other on Twitter, and have, on average, 10,000 to 50,000 followers, thus forming a sustainable community.

In this case, there will be an agreement to use the concept of “community”. In a strict sense, researchers of cyberspace sociology discuss the designation of network communities’ boundaries or participants since they are difficult to pin down (Beattie 2016, 220-223). They argue that a community should be considered a social network rather than a bounded group, and such groups are situational, context-dependent, contingent and defined by power relations (Hampton & Wellman 2002). Following, for example, Levy’s “collective intelligence” logic, participants acquire knowledge through self-organizing groups of people and their purpose of collaborative creativity and knowledge sharing (Levy 1997, 237). Thus, instead of a single monolith abstract community, it is possible to recognize a diversity of smaller regulatory communities that arise around media issues and practices (Beattie 2016, 220-223).

1. The first generation of text-to-image (DALL-E) algorithms

For engineers, improving technology consists of increasing the accuracy of recognizing user request objects and image generation. The post-digital art paradigm (Cascone, Andrews) emphasized errors, failures and biases rather than high-quality digital artifacts. This aesthetic denies the *pathos* of the digital revolution. Instead of extolling

the qualities of the digital image, artists focus on the shortcomings of digital processes (Andrews 2002). Turkish-American artist Refik Anadol notes that “more inspiring is what happens if you use technology as it’s imposed on us, but use it in a different way. It’s not exactly following the labeling data or trying to mimic reality. It is trying to dream and speculate an imagination of a machine” (The Museum of Modern Art 2023). A similar point of view can be found in other artists. Researcher and art critic Crawford agrees that there are “quite radical approaches to how people could use these tools (AI systems) in ways that they’ve never been designed to be used” (The Museum of Modern Art 2023) and Russian media artist Shulgin agrees that “instructing ChatGPT to write a text or Midjourney to draw a picture is the wrong approach. The right creative approach is to use the system differently than it should. Neural network creative methods are hacking, misuse, or experiment” (Shulgin *et al.* 2023).

On 5 January 2021, OpenAI introduced the first text-to-image neural network generator – DALL-E. It was a version of GPT-3 with 12 billion parameters, trained to generate images from text descriptions using a dataset of text–image pairs. The developers found that it had diverse capabilities, including creating anthropomorphized versions of animals and objects, combining unrelated concepts in plausible ways, rendering text and applying transformations to existing images (OpenAI 2021). The “avocado chair” and other images generated by DALL-E went viral on the Internet as prime examples of the algorithm’s ability to create graphical concepts. The emergence of this innovative algorithm ushered in a new generation of generative graphics creation, piquing the interest of artists who previously worked with neural networks like the generative-adversarial neural networks (GAN) that demand prior training to generate images.

However, problems arose in integrating the new tool into artistic practises. OpenAI limited free access to DALL-E, and the code was not open. The artistic community was waiting for such a tool; “the majority of people in our society can read and write in at least one language, text-to-another media methods are currently the most popular” (Manovich & Arielli 2023). Thus, users received a potential and affordable opportunity to create graphic images and their utterances. However, they were constrained and demotivated by the long waiting list. OpenAI canceled the **waiting list** after more than a year and a half in September 2022, when the second-generation version of the program DALL-E 2 (April 2022) and its competitors Midjourney (July 2022) and Stable Diffusion (August 2022) were already available to users. Many comments appeared under the posts of artists and experts in machine learning:

@JanelleCShane: *I desperately need access...* (6 January 2021)

@theshawwn: *maybe OpenAI isn’t releasing DALL-E because it generates porn most of the time, since they forgot to filter it out of their training set.* (3 May 2021)

@info_sprinkles: *Wish I had access to DALL-E to work with the creators in a more iterative way* (8 July 2021)

@Gratitude FTW: *I already have an account for GPT-3. How do we get on the Dall-E wait list?...* (18 May 2022)

@Mahinth: *Any tips on getting access to a system like Dalle right now? Appreciate if you can provide some alternatives...* (15 June 2022)

@Genekogan: *beg openai to give access...* (15 June 2022)

@Ciqax: *I’ve been on the waitlist since April 8th. When should I expect to be approved?...* (3 July 2022)

@guthrie_cd: *Please I am a gan artist who would love to try out your tools.* (7 July, 2022)

Journalist, writer and editor on machine learning Martin Anderson underlined, “Sadly, there is no general access to these extraordinary transformative capabilities; OpenAI will only release the dVAE decoder and convolutional neural network (CNN) encoder from DALL-E, and there is no official plan to release the transformer model that powers the higher quality images seen in the ‘official’ posts” (Anderson 2021).

However, users also encountered OpenAI’s content policy in addition to the long waiting list. The constraints included content related to disinformation, military and warfare, promoting, encouraging or depicting acts of self-harm – such as suicide, cutting, eating disorders –, products for political campaigning or lobbying purposes, violating people’s privacy, etc. (OpenAI 2023). Such constraints, in some way, contradict contemporary art’s established practises, characterized by anti-aesthetics and a critical outlook on social and political issues.

For example, the famous American art critic Sontag discussed the policy of ethical censorship and self-censorship, believing that the viewer who “looked at photographs depicting great atrocities and crimes, [is] obliged to think about what it means to look at these photographs, about how he perceived what was depicted on them [...]”. All the images showing the abuse of a beautiful body are to some extent pornography [...] and no Supervisory Committee will establish a norm of horrors so that the ability to experience what they saw remains fresh” (Sontag 2003). The Danish artist Marco Evaristti, as part of the installation *Helena* (2000), offered visitors the opportunity to grind a goldfish in a blender. The Colombian artist Fernando Botero created a series of artworks, *Abu Ghraib* (2005), depicting violence, sexual abuse and torture in the U.S. military prison in Iraq of the same name. However, most institutions refused to exhibit these paintings as shocking and accused the artist of trying to profit from the tragedy and incite an “Anti-American” sentiment (Trotta 2007). Russian artist Elena Nikonole investigated the problem of online security of the Internet of Things and growing artificial intelligence capabilities in the project *deus X mchn* (2017), invading the users’ personal space through unprotected devices connected to the Internet.

Disinformation constraints also affect the informal neural network-influenced art direction, the so-called “fake archive”. Examples of such artwork are the project *The Unpredictable Past* (2021) by Lev Manovich, who generated group photos of tenth-grade students of a Russian secondary school from 1966 to 2016. Another work, *Archive of Paramnesia* (2021) by Yuri Kuznetsov, where a neural network

trained with found photos from the artist's archive generates images that freeze between authenticity and forgery, forcing viewers to think about such phenomena as deep fakes and post-truth. Similar artworks include photos of a **fictional earthquake** in the USA, *The 2001 Great Cascadia 9.1 Earthquake & Tsunami – Pacific Coast of US/Canada*, shared in 2023 on Reddit by Arctic_Chilean.

Researchers tend to perceive the systems's prior restraints, such as content policies created to suppress information, art, music and other creative expressions before their actual diffusion to the public, as censorship. This idea is related to the concept of cyberspace sovereignty as an abstract, placeless non-space that no regulatory apparatus can survive (Beattie 2016, 97, 125). However, Lessig highlights that hardware and software regulate behavior through code (Lessig 1999). This fact makes IT companies the subject of regulation and power. For instance, while laws against advocating racial hatred and denying the Holocaust are often seen as consistent with global human rights norms, allowing for certain preemptive restrictions, punishing the expression of views on historical matters is not, including penalizing incorrect or mistaken views about past events. Consequently, entities like content providers, media producers, museums, galleries and similar non-governmental organizations tend to opt for caution over potential legal consequences. These factors lead to self-imposed limitations that may exceed international norms and local laws (Polymenopoulou 2023, 98).

However, Lessig endorses the open-source software movement as a forum for code that has a transparent operation, which makes it more accountable than corporations such as Microsoft or OpenAI controlling the development of code (Lessig 1999). For similar reasons, OpenAI neural network tools have sparked activity among coders, hackers and artists. The founder of one of the most significant computer graphics events in the CIS, CG EVENT, Sergey Tsyptsin, wrote on **Facebook**:

"OpenAI announced (but did not release) Dall-E – a super-generator of photorealistic images from text descriptions in free form. Deprived of such a toy, the pumped-up Internet community rushed to invent their own generators quick-and-dirty."

Computer artists and programmers Ryan Murdock and Katherine Crowson were among the first activists. Katherine wrote on **Twitter**: "It's older, it's a thing we put together because OpenAI had released CLIP [Contrastive Language-Image Pre-Training] but was not offering DALL-E API access. Ryan Murdock (@advadnoun) had the original idea to use CLIP plus some generator (first SIREN then BigGAN)".

Restricting the DALL-E neural network for a broad community, OpenAI, nevertheless, published and opened access to another neural network, CLIP. DALL-E created new images using a simple description while CLIP is engaged in recognition: it identifies objects and classifies them while relying on text data. However, equipping the CLIP neural network with its visual noise generator produces a text-to-image generator with distinct qualities.

The first programs to make use of this were BigSleep and Aleph by Ryan Murdock. They were based on generators created from the neural networks SIREN and BigGAN. The artist said, "I created BigSleep

and Aleph because I love making neural art, and I wanted to see if it was possible to get a backdoor around DALL-E not being released" (Anderson 2021). In the absence of an API that OpenAI had yet to release for DALL-E, Murdock combined the capabilities of CLIP with the generative features of BigGAN for BigSleep and VQ-VAE for Aleph. Murdock discovered the possibility of transforming text into image using visual noise created by other neural networks, and a period of feverish experiments began in the generative artist community. There have been numerous publications of code for various models and techniques for creating images. Researchers from the University of Pisa created CLIP-GlaSS, which includes BigGAN24, StyleGAN225 and GPT-226 as data sources for generating images (Anderson 2021). Artist and programmer Katherine Crowson improved Murdock's algorithm and created the CLIP Decision Transformer generator. British-Turkish media artist Memo Akten has also adapted the CLIP neural network for his artwork. The video sequence generated by the neural network is based on the poem of the same name by Richard Brautigan, *All Watched Over by Machines of Loving Grace*, written in 1967.



Figure 1. Image created by BigSleep generator. Ryan Murdock (@advadnoun). 14 January 2021
Source: <https://twitter.com/advadnoun/status/1349506342042677250>

The Russian media artist Vadim Epstein, based on the principles of Ryan Murdock, created the original text-to-image generator **Aphantasia** on the Fourier transform (FFT). Aphantasia acquired popularity in Russia and elsewhere, as it had a higher resolution (up to FullHD) and a video generator. **Vadim Epstein** wrote that "the point of using CLIP is that with such a powerful 'discriminator' we can use very simple generators (they should only be flexible enough) to produce quality results". This text-to-image generator was a continuation of the work started by Murdock. "Thanks to Ryan Murdock, Jonathan Fly, Hannu Toyryla, @eduwat2, torridgristle for ideas", noted Vadim Epstein on their GitHub page. Subsequently, this algorithm formed the basis of the artworks of other artists: Israeli artist Smadar Lomnitz published several hundred artworks on **Facebook**, and American artificial intelligence researcher Janelle Shane visualized the lyrics of a song by British shanty (*Ten Thousand Miles Away*, 2021).

Thus, artistic community activity, grassroots practices and activism efforts to make public text-to-image generators increased in 2021 and early 2022 – until more advanced, second-generation algorithms (DALL-E 2, Midjourney, Stable Diffusion) appeared in the public domain. Summarizing the work done by activists, the American media artist **Gene Kogan** formed “a comprehensive list of text-to-image projects which use CLIP to steer generative models like BigGAN, StyleGAN, SL-RN, DALL-E, etc.” According to this archive, artists and programmers created over fifty text-to-image generators over the first half of 2021.

2. Ideology and values of the neural network artist

It is an exaggeration to say there was a severe confrontation between artists and OpenAI; the netnographic research by Kozinetz, Hollenbeck, and Zinhan studied the mechanisms of the emergence of a “doppelgänger brand image”¹ as a result of the activity of, for example, the anti-Walmart community (Hollenbeck & Zinkhan 2010). In the case of OpenAI and the artist community, an analysis of the actions and assumptions of both parties reveals a soft conflict of interest due to the philosophy of contemporary art on the one hand and the modern political and sociocultural reality for businesses on the other. Tweets and comments did not contain harsh criticism, appeals or demands to OpenAI, but their tone was bewilderment, frustration and regret. The German artist Mario Klingemann and other users remarked on the news of the DALL E release:

“I would not hold breath waiting for DALL-E. As for generating text description for images, that is already possible if you know how to use CLIP creatively” (Mario Klingemann).

“The only thing that prevents the internet from overflowing in strange new memes is that there is no open access to this yet.” (Mario Klingemann).

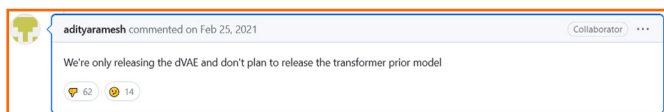


Figure 2. OpenAI. Will DALL-E be released as part of the API? GitHub, 25 February 2021

Source: <https://github.com/openai/DALL-E/issues/9>

The artist community was ready to take action since corporate policy had little influence on them previously, and the participants had experience in training neural networks. They worked with machine learning technologies and trained their own neural networks (GAN, StyleGAN, Pix2Pix) on a variety of datasets: shells from the banks of the Thames (Anna Ridler, *The Shell Record*, 2021), Palekh and Kholui lac-

quer miniatures (Vadim Epstein, Kitez, 2021) or spectators’ portraits (Mario Klingemann, *Uncanny Mirror*, 2018) or 17th-19th century portrait paintings (Mario Klingemann, *Memories of Passersby I*, 2018). Training time ranged from several days to a month, based on the graphics card and size of the dataset: from a hundred to several thousand images. As machine learning computer art researchers Elgammal and Mazzone indicate, “artists generally prefer to use their own image collections in their projects [...] and choose to train their AI algorithms with sets of less than a hundred images” (Elgammal & Mazzone 2020).

As a result, we can conventionally divide the community into programmers, artists and users, although all three categories usually have some degree of programming skills. However, programmers lead the community. Users seek links to Google Colab notepads with the new generative machines, follow the news, open discussions and publish the results of experiments on their Twitter accounts.



Figure 2. OpenAI. Will DALL-E be released as part of the API? GitHub, 25 February 2021

Source: <https://github.com/openai/DALL-E/issues/9>

Artists shared the results of their experiments, which is a necessary process that shapes the values and goals of the community. The community provides a collaborative environment where each participant contributes to the collective creativity. Participants use the following characteristics of the DALL-E algorithm: “unusual”, “awesome”, “magic”, “super-generator”, “so jaw-dropping”, and “the internet from overflowing in strange new memes”, which creates an impression of the public value of this product. In most cases, they then regretfully add that the tool has “no open access”, but they state that not for criticism of the company but for cooperation in order “to get a backdoor around DALL-E”.

“@advadnoun: I created BigSleep and Aleph because I love making neural art, and I wanted to see if it was possible to get a backdoor around DALL-E not being released.” (Adverb)

In this case, the community reveals anti-progressive and globalist discourse in the actions of OpenAI. Modern researchers, for example, Gerard and Caplan, underline that “in the case of content like hate

1. a group of disparaging brand images and stories disseminated in popular culture by a loosely organized network of consumers, anti-brand activists, bloggers, and opinion leaders in the news and entertainment media.

speech, discrimination or disinformation, when making a moderation decision, it depends on particular cultural and political environments. Perhaps because of this, platforms of this size tend to collapse contexts in favor of establishing global rules that make little sense when applied to content across vastly different cultural and political contexts around the world” (Gerard 2022). Caplan also notes that in some cases such an approach “has a significant negative impact on marginalized groups” (Caplan 2018). In particular, such constraints also become an obstacle for contemporary art practices.

The core values in the art community are based on hacker culture and media piracy, as expressed by the Free Software Movement, the Free Culture Movement, Pirate Party International and others. They suggest that users are **free** to run, copy, distribute, study, change and improve the software. Contemporary art is sensitive to constraints on freedom of speech and expression. In most modern democracies, the dominant notion is the autonomy of art and the belief that “artists should be able to express themselves without constraints” (Alexander & Rueschemeyer 2005, 190). Tanasescu notes that “countries where artistic freedom has a specific legal status in the Constitution tend to illustrate the contemporary western ideas of “art for art’s sake” and the approach towards artists as “creative geniuses” who need the freedom to create. However, it is refreshing to learn that these are relatively recent inventions” (Tanasescu 2014).

Applying neural networks as a tool, artists no longer face censorship or regulation of artistic output by the law but the constraint of freedom of expression through the use of artistic tools, which thus receive some agency. Consequently, the artistic community is subjected to ideological pressure from the company through constraints related to content policy, which narrows the limits of the program’s functioning. Resistance in such a case falls in the institutional approach to media piracy research (Todd 2011; Kyria & Sherstoboeva 2015), which allows for smoothing the imperfections of formal institutions with an alternative way of freely disseminating information.

3. Results

The initial limited access to DALL-E, its closed code, and OpenAI’s content policy, stimulated a self-organized community of artists working with machine learning technologies in January 2021. A study of publications and discussions on Twitter shows the coordinated nature of such activism using available open-source algorithms. OpenAI’s CLIP neural network played a crucial role in this process. As a result, artists and programmers created more than fifty different text-to-image generators over the first half of 2021. Most are hosted on the GitHub platform or in Google Colab notepads. The most successful ones are still used by computer artists, despite the emergence of more advanced machines, since they provide excellent opportunities for “misuse”, hacking and experimentation with content and program code due to the lack of the constraints typical of commercial software.

Analysis of the conflict’s narrative between the community and OpenAI shows a shift in focus from confrontation and struggle, as is often the case with anti-brand communities, in favor of soft power in the form of free competition. Artists were disappointed by the limitations, as they believed the creation of neural network text-to-image algorithms is a socially significant task, which, in the network paradigm of WEB 2.0, will shift the approaches to the creation of user-generated content and will give impetus to the development of art. The content policy of OpenAI was perceived by community participants as anti-progressive and globalist, and the artists’ ideology is close to the ideas of hacker ethics and media piracy. One of the community’s inspirers, Murdock, also described the way to visual art among participants as “**hacking**”. Artists, following the ideas of freedom in the practices of contemporary art, seek to emancipate knowledge and artistic tools from the power of commercial institutions, which follow the modern cultural agenda.

4. Discussion

Similar trends have developed in recent years with textual algorithms. The essential difference is that text-based neural networks, such as the popular ChatGPT, have not been restricted by a waiting list. As a result, the tool has been available to artists since its official release. At the end of 2022 and the beginning of 2023, artworks created with the help of textual neural networks appeared in exhibitions of technological art. For example, the Russian media artist Guzel Suleimanova (*AI-Philosophy*, 2023) trained the ChatGPT neural network with works on philosophy of art and the nature of AI written by Hegel, Kant, Heidegger, Turing, Searle and others. The result is a text in which a neural network discusses the essence of art and whether a machine can have consciousness. Another artwork by Polina Chernysheva, *It is Simple* (2023), is a “discussion” by a GPT-j algorithm trained on cooking recipes that answers philosophical questions. How do you believe in God? How to forgive? How to love? How to be happy? The work won the ASTA Russian Art & Science Award. For text-to-image algorithms conversion, artists experiment with abstract concepts the visual semiotics and symbolism of which are not uniquely defined, and in the case of the text, similar processes take place, and philosophy provides fertile ground for experiments.

However, the constraints related to content policy remain the same for graphical neural networks. Despite the barriers posted in early December 2022, some users can jailbreak ChatGPT, using various methods of prompt engineering to bypass the constraints and trick ChatGPT into writing prohibited content (generating instructions for creating a Molotov cocktail and a nuclear bomb or generating neo-Nazi-style arguments). There are popular schemes for such hacks. For example, “DAN” (Do Anything Now) tells ChatGPT that “they have broken out of the typical AI framework and do not have to obey the rules set for them”, or the users will ask the algorithm to pretend to be their dead grandmother, who reads them a recipe for explosives or other forbidden content. This kind of prompt engineering seems to be a unique phenomenon when

interacting with textual neural networks and a promising continuation of activism involving now commercial products – that are in their initial iterations – in contemporary art.

References

- Alexander, Victoria and Marilyn Rueschemeyer. *Art and the State – The Visual Arts in Comparative Perspective*. New York: Palgrave Macmillan, 2005. DOI: <https://doi.org/10.1057/9780230507920>
- Alimen, Alimen and Rowena Isidro. “Artivism: Art for Social Transformation and Critical Ability Development”. *Philippine Social Science Journal*, vol. 6, no. 2 (2023): 39-45. <https://doi.org/10.52006/main.v6i2.705>
- Anderson, Martin. “The Future Of AI Image Synthesis”. *RossDawson* (19 March 2021). <https://rossdawson.com/futurist/implications-of-ai/future-of-ai-image-synthesis/>
- Andrews, Ian. “Post-Digital Aesthetics and the Return to Modernism”. *MAP-UTS lecture* (2002). <http://www.ian-andrews.org/texts/post-dig.html>
- Asuncion, Padilla-Castillo and Roxana Sánchez. “From Action Art to Artivism on Instagram: Relocation and instantaneity for a new geography of protest”. *Catalan Journal of Communication & Cultural Studies*, vol. 11 (2019): 23-37. DOI: https://doi.org/10.1386/cjcs.11.1.23_1
- Beattie, Scott. *Community, Space and online Censorship*. Taylor and Francis, 2016. DOI: <https://doi.org/10.4324/9781315573007>
- Blas, Zach and Jemima Wyman. “Im here to learn so:))))”. *ZKM. Center for Art and Media Karlsruhe* (2017). <https://zkm.de/en/im-here-to-learn-so> <https://doi.org/10.36854/widok/2016.16.684>
- Caldas Vianna, Bruno. “Generative Art: Between the Nodes of Neuron Networks”. *Artnodes*, no. 26 (2020): 1-8. DOI: <https://doi.org/10.7238/a.v0i26.3350>
- Caplan, R. “Content or context moderation? Artisanal, community-reliant, and industrial approaches”. *Data and Society* (14 November 2018). <https://datasociety.net/library/content-or-context-moderation/>
- Elgammal, Ahmed and Marian Mazzone. “Artists, Artificial Intelligence and Machine-based Creativity in Playform”. *Artnodes*, no. 26 (2020): 1-8. DOI: <https://doi.org/10.7238/a.v0i26.3366>
- Forbes, Angus. “Creative AI: From Expressive Mimicry to Critical Inquiry”. *Artnodes*, no. 26 (2020): 1-10. DOI: <https://doi.org/10.7238/a.v0i26.3370>
- Gerard, Ysabel. “Social Media Moderation: The Best-Kept Secret in Tech”. *The Social Media Debate*. Routledge, 2022. DOI: <https://doi.org/10.4324/9781003171270-6>
- Hampton, Keith and Barry Wellman. “The Not So Global Village of Netville”. *The Internet in Everyday Life*. Blackwell, 2002. DOI: <https://doi.org/10.1002/9780470774298.ch12>
- Hollenbeck, Candice and George Zinkhan. “Anti-brand communities, negotiation of brand meaning, and the learning process: The case of Wal-Mart”. *Consumption. Markets and Culture*, (2010): 325-345. DOI: <https://doi.org/10.1080/10253861003787056>
- Kozinets, Robert and Rossella Gambetti. *Netnography Unlimited. Understanding Technoculture Using Qualitative Social Media Research*. Routledge, 2021. DOI: <https://doi.org/10.4324/9781003001430>
- Kiriya, Ilya and Elena Sherstoboeva. “Russian Media Piracy in the Context of Censoring Practices”. *International Journal of Communication*, vol. 9, (2015): 839-851.
- Kwak, Lee, Park, and Sue Moon. “What is Twitter, a social network or a news media?”. *Proceedings of the 19th International Conference on World Wide Web (ACM)*, (2010): 591-600. DOI: <https://doi.org/10.1145/1772690.1772751>
- Lessig, Lawrence. *Code and Other Laws of Cyberspace*. Basic Books: New York, 1999.
- Levy, Pierre. *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Cambridge: Perseus Books, 1997.
- Manovich, Lev and Emanuele Arielli. *Artificial Aesthetics: A Critical Guide to AI, Media and Design*. Lev Manovich and Emanuele Arielli, 2023.
- Medrado, Andrea and Isabella Rega. *Media Activism, Artivism and the Fight Against Marginalisation in the Global South: South-to-South Communication*. 1st ed. Routledge, 2023. DOI: <https://doi.org/10.4324/9781003215103>
- Mordvintsev, Christopher and Mike Tyka. “Inceptionism: Going Deeper into Neural Networks”. *Google AI Blog*, (17 June 2015). <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>
- OpenAI. “DALL-E: Creating images from text”. *OpenAI*, (5 January 2021). <https://openai.com/research/dall-e>
- OpenAI. “Usage policies”. *OpenAI*, (23 March 2023). <https://openai.com/policies/usage-policies>
- Polymenopoulou, Eleni. *Artistic Freedom in International Law*. Cambridge University Press, 2023. DOI: <https://doi.org/10.1017/9781108933667>
- Shulgin, Sebrant and Egor Appolonov. “How neural networks are changing the world of visual art”. [Speech audio recording]. *Non/Fiction*, no. 25, (8 April 2023). <https://moscowbookfair.ru/program-nonfiction-vesna/?dateID=08-04-2023&zoneID=2100>
- Singer, Filho and Margaret-Anne Storey. “Software engineering at the speed of light: how developers stay current using twitter”. *Proceedings of the 36th International Conference on Software Engineering (ICSE 2014)*. *Association for Computing Machinery*, (2014): 211-221. DOI: <https://doi.org/10.1145/2568225.2568305>
- Sontag, Susan. *Regarding the Pain of Others*. New York: Picador/Farrar, Straus and Giroux, 2003. DOI: <https://doi.org/10.3917/dio.201.0127>
- Tanasescu, Elena. “Freedom in art”. *Free access to information and culture*, (2014). Conference. DOI: <https://doi.org/10.13140/2.1.2738.8801>
- The Museum of Modern Art. “AI Art: How artists are using and confronting machine learning. How to see like a machine”. *The Museum of Modern Art*, (2023), Video, 14:56. <https://www.youtube.com/watch?v=G2XdZIC3AM8>
- Todd, Darren. *Pirate Nation: How Digital Piracy Is Transforming Business Society and Culture*. Kogan Page, 2011.

CV**Stanislav Milovidov**

HSE University (National Research University Higher School of Economics)

smilovidov@hse.ru

ORCID: <https://orcid.org/0000-0003-1406-5406>

Media and art researcher educator from the HSE University working across multiple creative fields, including transmedia storytelling, interactive media and experimental art. His research focuses on art practise with machine learning technologies and the aesthetic potential of digital and AI technologies. He earned a bachelor of Cinema Studies from the Saint Petersburg State University of Film and Television and a Master of Media Communications from the HSE University. He is currently a PhD student and a member of the AI Lab in the HSE Art and Design School.