

COMBATting DEEPFAKES: THE ROLE OF THE PUBLIC AND DESIGN COMMUNITY

Chemerys Hanna^{1,2} [0000-0003-3417-9910]

¹ PhD, Associate Professor, Head of Department of Design, Zaporizhzhia National University, Ukraine Anyta.Chemeris@gmail.com

² associate researcher, CEFRES (French Research Center in Humanities and Social Sciences), the Czech Republic

Abstract. The issue of identifying and differentiating neurocreativity is also addressed, and the potential use of neuroaesthetics technology to enhance the aesthetic level of generative design is explored. In the context of recent advancements in artificial intelligence, such as Midjourney, Stable Diffusion, DALL•E, and Lensa, there is a growing and active discussion on the ethical aspect of utilizing these technologies in the realm of visual art generation. The ethical dimension of using artists' works to train artificial neural networks is thoroughly examined, along with the question of copyright in the context of AI image generation. The author advocates for the need for ethical regulation in this field and proposes opportunities for the application of artificial intelligence in collaborative creative tasks with artists.

Keywords: Generative Art, Neural Networks, Artificial Intelligence, Design.

The challenge of integrating creativity into algorithms has become prominent in recent years for AI. Neural networks have demonstrated the ability to generate text, manipulate images, compose music, and even produce visual art. Some experts predict that by 2060, AI might accomplish nearly all tasks currently performed by humans. McKinsey Consulting estimates that approximately 1.1 billion jobs worldwide, with over 100 million in the USA and Europe, could be automated. Certain scientists go so far as to suggest that AI could surpass human capabilities across all domains within the next 50 years.

In a survey conducted by the Boston Consulting Group (2022), internet users expressed concerns about the impact of AI on the job market, with 61% fearing significant disruptions. Additionally, 58% believe that government intervention is necessary to safeguard employment, while 32% worry about the transparency of

decision-making and potential ethical issues related to AI. To identify roles most susceptible to automation, Danish scientist Jens Rasmussen's (1983) three-task category division is considered: skills (physical labor), rules (e.g., accountants, lawyers), and knowledge (50% analytical and creative activities). While automating «skills» may seem complex due to the substantial investment required for robot development, the «rules» category appears more straightforward. Chatbots are already analyzing contracts and challenging illegal fines, with Chat GPT being a notable example.

This analysis suggests that if 50% of current professions can be automated, there's a potential risk of workforce redundancy within a decade, as indicated by the «Will Robots Take My Job» service. This service, based on the «The Future of Employment» report by Carl Benedict Frey and Michael Osborne, predicts the likelihood of AI replacing various professions. According to their findings, both creative and technical roles may transition to computer programs. For instance, the probability of automation replacing interior designers is only 8%, whereas for accountants and financial analysts, it stands at around 50%. AI, like Stable Diffusion, employs a mathematical technique called Diffusion. This approach is utilized to store compressed versions of images, which the AI learns from to subsequently reconstruct and generate new images. In simpler terms, it operates as a contemporary collage maker.

To delve deeper into the diffusion process, let's consider the example of a spiral in the diagram below. Initially, the computer identifies the original spiral image as the foundation. It then undergoes a «diffusion» process, introducing noise until the image becomes completely distorted. Subsequently, the reverse process begins, eliminating the noise to reshape the image back into a spiral. As a result, a duplicate of the original image is produced, albeit with inaccuracies and variations, as the noise introduced during the process slightly modifies the original image (see Figure 1).

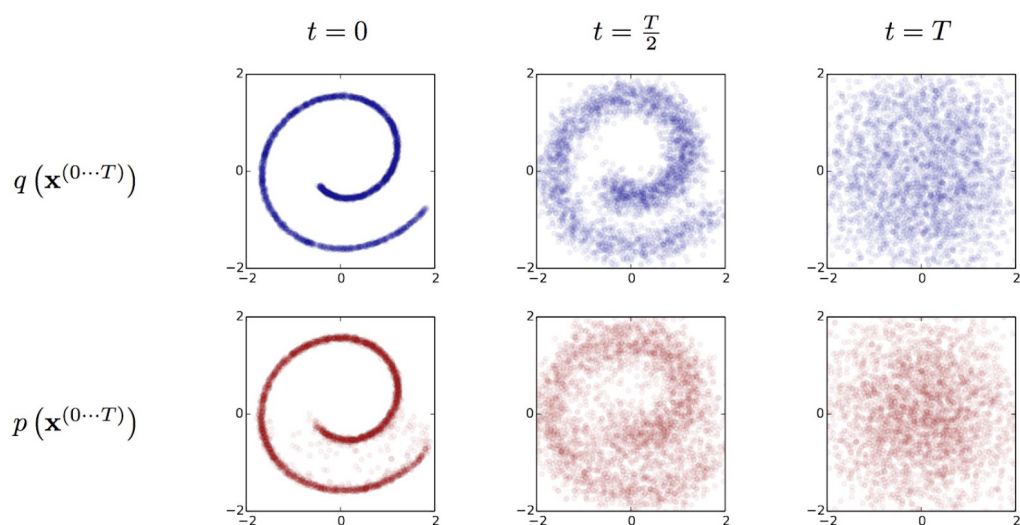






































Figure. 1. Illustration of the diffusion process on the example of a spiral. The material is taken from the artists' lawsuit against Stable Diffusion.

The present landscape of AI platforms generating visual art is both remarkable and rapidly expanding, as illustrated in Table 1. Consequently, we aim to conduct a retrospective review of the inception and evolution of the technology utilizing neural networks in the realm of visual art.

The term «artificial intelligence» was first introduced in 1956 by the American scientist John McCarthy. At that time, it emerged as a branch of informatics and a scientific field dedicated to simulating various facets of intelligence through computer systems. Concurrently, there were hypotheses suggesting the possibility of algorithmizing and mechanizing the thought process. Notably, as early as the 13th century, Raymond Lullius developed a logical system, and by the 19th century, various countries were witnessing the emergence of projects involving intelligent and computing machines.

Table 1.

The current landscape of artificial intelligence platforms that generate visual art

Type of iteration	AI Platform
Text-to-Image (T2I)	 > ROSEBUD.AI maket  pencil beautiful.ai  BOTIKA  PhotoRoom®  Facet  MODULIZE Re:cast AI  CSM   imagen Hexo AI    Stable Diffusion  Jasper  MidJourney  GauGAN2  Wonder pixray-text2image 
Text-to-Video (T2V)	 Fliki  Meta AI Google AI Phenaki   Maria  TERRA   PICTORY  Steve AI VOCHI  recut  lumen5  inworld VEED.IO  YEPIC  Basch.io  Beeptub.ai  METAPHYSIC  WOMB0
Text-to-Motion (T2M)	 MDM: Human Motion Diffusion Model
Text-to-NFT (T2I)	
Text-to-3D (T2D)	 CLIP-Mesh GET3D

Significant advancements in artificial intelligence (AI) emerged in the mid-20th century, sparking varying attitudes toward the field over the past six decades. During the Cold War, the United States invested in machine translation development. However, the complexity of the task led to its cancellation, shaking confidence in artificial intelligence and marking the onset of the so-called «winter of artificial intelligence» A resurgence occurred in the 2000s.

In the mid-2010s, the first endeavors to generate images from text materialized with the introduction of Generative Adversarial Networks (GANs) (Chamberlain, et al., 2018). GANs constitute a system comprising two artificial neural networks engaged in a competitive process. The generator network produces images based on textual descriptions, while the discriminator network assesses them. During training, the generator's objective is to deceive the discriminator by crafting a synthesized image closely resembling a real one. Simultaneously, the discriminator aims to accurately differentiate between authentic and synthesized images. An illustrative instance of such generation was crafted in 2016 (Figure 2):




Caption	Image
this vibrant red bird has a pointed black beak	
this bird is yellowish orange with black wings	
the bright blue bird has a white colored belly	

Figure 2. An example of a generation created in 2016

Subsequently, various alternative algorithms for generating images from textual queries emerged, including CNN. However, these alternatives did not exhibit a noticeable improvement in quality compared to GANs. The foundation for the Mask R-CNN algorithm, which is a type of Convolutional Neural Network (CNN), was laid (CNN, along with GUN, is responsible for object recognition). This algorithm was employed to «breathe life» into characters from children's drawings, a task that posed challenges for artificial intelligence. On one hand, these drawings did not resemble real people in photographs, yet, on the other hand, they did exhibit certain similarities such as heads, arms, and legs. CNN itself had various subspecies, such as LeNet, AlexNet, VGGNet, and ResNet, each of which underwent improvement through participation in the ImageNet competition. Significant breakthroughs in the realm of text-to-image generation occurred with the introduction of DALL•E and CLIP by OpenAI. These two solutions marked a revolutionary shift in image generation. The DALL•E neural network, based on GPT-3, a third-generation natural language processing algorithm from OpenAI, employs a transformer architecture. This architecture extends the text sequence with specialized image tokens, which are

subsequently transformed into images by another model (decoder). DALL•E demonstrated a substantial leap in the quality of synthesized images compared to previous solutions. Notably, the neural network's remarkable generalization capability allows it to generate examples not encountered during training, effectively expanding its capacity to produce novel and diverse images. An illustrative example of this is OpenAI's distinctive creation of chairs shaped like avocados, which has become a hallmark of their work (Figure 3).



Figure 3. Avocado-shaped chairs generated by OpenAI

Upon its release, CLIP garnered less attention compared to DALL•E, but it made a more substantial impact on the advancement of text-to-image generation. CLIP seamlessly connects images with text and comprises two encoders—one for text and one for images. The true surge in the popularity of artificial intelligence took place in 2022, providing users with access to AI tools capable of generating diverse images based on textual descriptions. Unlike DALL•E, developers freely shared trained CLIP resources. Following this, numerous solutions utilizing CLIP, such as BigGAN+CLIP, VQGAN+CLIP, CLIP Guided Diffusion, and others, emerged in the text-to-image domain. Nvidia's GauGAN initially operated at a slower pace and with less precision, and the program was not available for free. Subsequent improvements involved training the neural network on 5 million photos, leading to the launch of GauGAN2. With just three or four words and a rough sketch, the program now produces photorealistic images. Text recognition plays a crucial role in aligning the generated images with the author's intention (Chamberlain, et al., 2018; Ragot, et al., 2020). Users can specify not only materials, time of year, and day but also precisely describe their requirements. However, the algorithm still excels primarily with landscapes and struggles with illogical phrases (Figure 4).

Despite the ongoing debates surrounding artificial intelligence, many AI products disappoint users who have high expectations.



Figure 4. StyleGAN was trained on real photos from Icons8 to generate photorealistic portraits of non-existent people

In 2014, The Grid introduced II Molly, positioning it as a tool for websites that could «create themselves.» The crowdfunding campaign for The Grid managed to collect \$96 per person, but it took several years for the creators to unveil a finished product that ultimately failed to meet the public's expectations. In August 2022, there was a renewed surge in discussions about images generated by the Midjourney neural network. This program creates images based on text requests, and The Economist (2022) even used it to design a cover. However, it required 250 requests to generate a suitable option. Despite this, the neural network demonstrated the ability to produce detailed images in various styles of work. While corporations typically have access to greater computing power, many generative artists prefer open-source solutions. The advantage of such solutions lies in their accessibility, allowing anyone to participate in product development, experiment with its functionality, and contribute to the rapid evolution and improvement of the technologies.

The decision between open-source products and commercial solutions is often influenced by censorship concerns. Companies, having access to AI tools, often implemented «corporate filters» restricting content such as pornography and violence (Hristov, 2016). In contrast to OpenAI DALL•E 2, Stable Diffusion (Figure 5) initially lacked such filters, allowing the generation of images with any content.

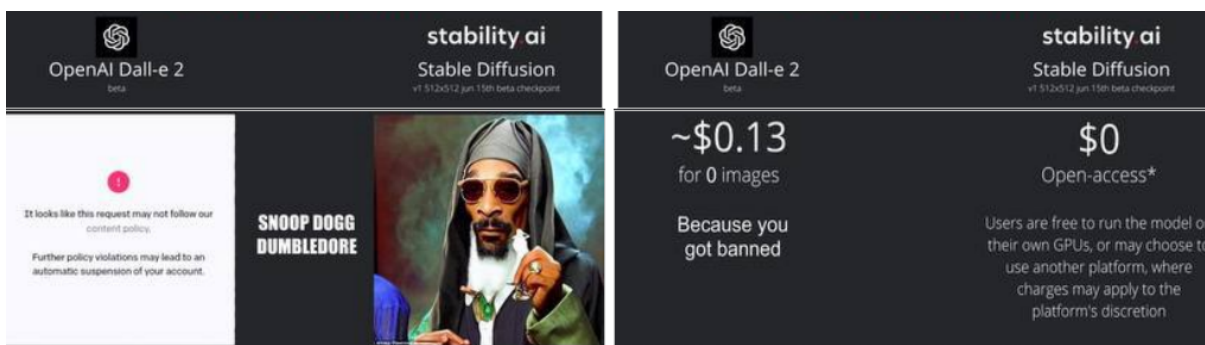


Figure 5. Censorship in DALL•E 2 and its absence in Stable Diffusion

So, for example, OpenAI prohibits the generation of images on «sensitive topics» and enforces blocks after repeated attempts to do so. Additionally, the word «Ukraine» is restricted for text input (Hristov, 2016; Gillotte, 2019).

The announcement of Jason Allen's victory in an art competition, where he created his winning piece with the assistance of the Midjourney neural network, sparked another intense discussion about the ethics of utilizing AI (Figure 6).



Figure 6. AI work by Jason Allen

AI advancements, such as Midjourney, Stable Diffusion, DALL•E, and Lensa, have sparked ethical concerns among digital artists. The main issue is the unauthorized use of artists' works to train AI models, leading to copyright violations. AI copies existing images, creating derivative works without permission, resulting in a concealed collage of various authors' works. Artists protest against unethical AI practices, copyright infringement, and cyberbullying, emphasizing their concerns about the devaluation of their work.

This is not a rejection of AI as a technology but a response to specific ethical issues. Platforms like ArtStation faced backlash for displaying AI-generated images alongside artists' originals, prompting manual blocking options. Legal action has been taken, with lawsuits against Stability AI, DeviantArt, and Midjourney for copyright infringement. Artists seek compensation and an injunction against future harm. The copyright infringement issue extends to various creative fields, though visual artists are at the forefront of grappling with AI challenges. Despite the advanced capabilities of AI, it has not replaced living artists entirely. Professionals and discerning viewers can still identify inaccuracies in the work of AI, even though it achieves a high level of image generation, can reproduce various styles, and even recreate traditional

paintings in digital art. Notably, reworkings of iconic works like «Girl with a Pearl Earring» and «Gioconda» exhibit peculiarities, such as a strange shape of the ear or unrealistic perspective (Figure 7).

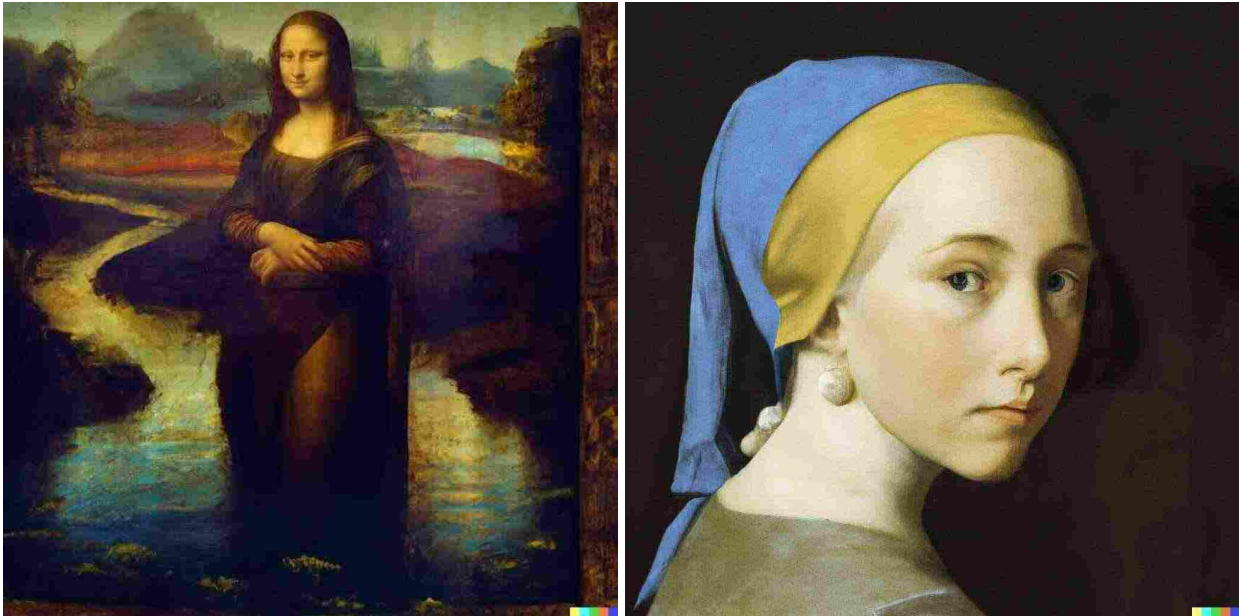


Figure 7. The result of the generation of digital images imitating oil painting technique and style

AI excels in generating design options but lacks empathy and emotional intelligence to understand customer emotions and motivations. Designers, crucial for interpreting ideas, currently face challenges with unregulated AI activities, urging the need for ethical norms in image generation to prevent copyright violations. A solution involves training AI on an artist's personal works to avoid copyright issues, allowing artists to visualize ideas uniquely. Collaborative processes, with neural networks and artists working together, facilitate rapid creation for various applications. Ethical challenges include using authors' works for neural network training and identifying neurocreativity. Training AI to recognize generative art may yield human-like images. Neuroaesthetics technology holds promise in refining AI-generated designs, enhancing aesthetic appeal through insights from neuroscience.

Funding. The work is performed within the research CEFRES, UAR 3138 CNRS–MEAE.

References

1. Chemerys, H. (2023). Truth & Trust in the Age of Deepfakes: Recognize & Overcome. *Українські студії в європейському контексті: зб. наук. пр.*, 7, 403-407. DOI: 10.31110/2710-3730/2023-7.
2. Chemerys, H., Briantseva, H. V., & Briantsev, O. A. (2021). The Urgency of the Problem Synthetically Reproduced Media Content. *International scientific conference «Interaction of culture, science and art in terms of moral development of modern European society»: conference proceedings*. Riga, Latvia: «Baltija Publishing», 85-88. DOI:

- 10.30525/978-9934-26-178-7-20.
3. Boston Consulting Group. (2022). 60% of Employees Using AI Regard It as a Coworker, Not a Job Threat (BCG). URL: <https://www.bcg.com/press/1november2022-employees-using-ai-regard-as-coworker-not-job-threat>.
 4. Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions On Systems, Man, and Cybernetics*, 3(3). URL: <https://www.iwolm.com/wp-content/downloads/SkillsRulesAndKnowledge-Rasmussen.pdf>.
 5. Chamberlain, R., Mullin, C., Scheerlinck, B., & Wagemans, J. (2018). Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychology of Aesthetics, Creativity, and the Arts*, 12(2), 177.
 6. Ragot, M., Martin, N., & Cojean, S. (2020). Ai-generated vs. human artworks: A perception bias towards artificial intelligence? *Extended abstracts of the 2020 CHI conference on human factors in computing systems*, 1-10.
 7. The economist, (2022). How a computer designed this week's cover. URL: <https://www.economist.com/news/2022/06/11/how-a-computer-designed-this-weeks-cover>
 8. Hristov, K. (2016). Artificial intelligence and the copyright dilemma. *Idea: The Intellectual Property Law Review*, 57(3), 431. URL: <https://heinonline.org/HOL/LandingPage?handle=hein.journals/idea57&div=21&id=&page>
 9. Gillotte, J. L. (2019). Copyright infringement in AI-generated artworks. *UC Davis Law Review*, 53, 2655.
 10. Andersen, S. (2022). The Alt-Right Manipulated My Comic. Then A.I. Claimed It. *The New York Times*. Retrieved from <https://www.nytimes.com/2022/12/31/opinion/sarah-andersen-how-algorithm-took-my-work.html>
 11. The Digital Millennium Copyright Act of 1998 U.S. Copyright Office Summary. (1998). Retrieved from <https://www.copyright.gov/legislation/dmca.pdf>
 12. Yanisky-Ravid, S. (2017). Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era: The Human-like Authors Are Already Here: A New Model. *Mich. St. L. Rev.*, 659.
 13. Zhang, B., & Romainoor, N. H. (2023). Research on Artificial Intelligence in New Year Prints: The Application of the Generated Pop Art Style Images on Cultural and Creative Products. *Applied Sciences*, 13(2), 1082.