

# Martina Šimkovičová

## On the Parallel Struggles of Photography and GAN-generated Imagery

---

### **Abstract**

This article analyzes certain aspects of the reception and acceptance of the emerging technology of GAN generated visuals (DALL-E, Midjourney, and Stable Diffusion) by the general public as of mid-2021 to early 2023. It will deal specifically with the negative and positive expectations of it and put them into the context of social acceptance and expectations surrounding the formation of the photography medium in the early 19th century. Both of the inventions mark a milestone in the timeline of a data-driven society. Both tools utilize a reference to reality or its visual coding. GAN image generators, as well as photography, were expected, as a product of science, to deliver almost mythically perfect, objective outputs. While photography had previously been used in pseudoscientific projects to support racist claims, the public appears surprised by the bias of the GAN datasets, and their creators are currently looking for ways to eliminate it through process moderation. The goal of using this comparison is to explore a common trait of both epochs: the

affinity for devices, specifically the idea of “a black box”. It will aid us in describing the relationship of the data society to itself based on machine-based creation of visuals with varying degrees of autonomy and generativity. Finally, we will examine the practical implications of the discussed aspects in the near future.

### **Key words**

19th Century. Accumulation of Knowledge. Black Box. Data Society. GAN Visuals. Lens-Based Imagery.

### Instead of Introduction

Let's use the term "photography" as a label for a set of chemical (photosensitive materials and means of developing and fixing the latent image) and physical (optical devices, electronic transfer of light) processes around which the social practice of photography started to emerge.

Its formation in the first half of the nineteenth century was aided by the more or less coordinated efforts of several individuals and teams. 1839 is marked as the official year of its birth in the history books, with its "father" being Louis Daguerre, although technically that was the year it was accepted as a patent by the French Academy of Sciences and then disclosed as a free patent to the general public of the world by the government (with an exception of Great Britain, where Daguerre's agent patented it).

At first, it adopted certain visual languages and conventionalized forms (genres) of other visual fields, mostly painting, to establish itself as a genius-driven autonomous art field.

A new aspect of the social practice of photography was the subject-object relation instrumentalized by the camera. The power dynamic between who photographs and who is photographed – the influences of one on the other – was no longer centered on unmediated eye-to-eye communication, but rather on and with a black box.

The black box also became a place of dissociation – of the detachment

of one's own image from their physical presence. It introduced the possibility of these multiple images existing independently in the subject's life. The previously impersonated knowledge begins to detach, and the things "previously given" (data) – factual information – begin to float. We can collect them and store them separately from the content. This resettlement influenced the emergence of data as a decision-relevant entity in twentieth-century society. The term "data" is a mid-17th century Latin philosophical term for the "facts given"<sup>1</sup> in photography – the content. Metadata, as "data on data", in photography refers to the formal, technical, structural, administrative, and legal information that gives us further context and becomes relevant after the amount of data reaches a certain extent and political meaning. The notion of interpreting words, analyzing their patterns and relations in the GAN generators appears to be in line with the goals of the metadata society, as Matteo Pasquinelli defines it. Pasquinelli<sup>2</sup> dates the birth of the metadata society to the late 1990s, when big data was being accumulated in large datacentres, both state-owned and private: *"The establishment of these large datasets as primary sources of cognitive capital and political power*

*marks the birth of the metadata society, because it is precisely the meta analysis of data—mapping and interpreting their patterns, trends, and forecasting their tendencies—that marks the birth of the metadata society"*<sup>3</sup>.

As with photography summarizing centuries of researched phenomena, GAN continues the research of neural networks, a concept that started in the 1960s.

The (hi)story of GAN (generative adversarial networks) begins in 2014, when an academic and computer scientist named Ian Goodfellow and his team invented the concept. A GAN consists of two neural networks: a generative network, which generates suggestions, and a discriminative network, which selects from them. The generative network usually maps its results from a latent space to a data distribution of focused interest, while the discriminative network compares them to the real distribution of data.

During the process, both the content and the formal aspects of the images are analyzed in the generative/discriminative process. These generations and discriminations are most common in computers and servers, which are remote from the user's screen and themselves, while they think it's a good prompt to type into the generator. There we have another black box, detached from its user, in which direct computational analysis

1 *Data*. [online]. [2023-01-30]. Available at: <<https://www.oxfordlearnersdictionaries.com/definition/english/data?q=data>>.

2 PASQUINELLI, M.: Metadata Society. In BRAIDOTTI, R., HLAVAJOVA, M. (eds.): *Posthuman Glossary*. London : Bloomsbury, 2018, pp. 253-256.

3 Ibidem.

of image referents takes place. User activity – the prompts, evaluations, and selections – becomes another material detached and harvested in this process.

The year 2021 marked the widespread usage of the Google *Collab* interface to generate images. *DALL-E*, *Stable Diffusion*, *LAION*, *Midjourney*, and others, both open source and commercial, offered this tool, and consequently, social media has become flooded with new visuals.

To deliver results that are usable in the human world, this process is “trained” using datasets: databases of visuals and their meaningful descriptions (partially done by machine vision and machine learning, partially through human input) and controlled by users’ text prompts and/or a visual input. Photographs available online through servers such as *Flickr* or *Google Search*, thumbnails of stock photography websites, professional and amateur photography, and historical footage scraped by the dataset creators to do machine learning comprise a significant portion of the images used in datasets. One of the most used datasets is *ImageNet*, which contains 14 million images with annotations combined with *Wordnet*, a lexical database of semantic relationships between words.

### 1 Scarcity in the Beginning, Overflow in the End

In the beginning, the aforementioned tools were invented

by experts, and their usage required specific knowledge. Similarly to photographic devices in the second half of the nineteenth century, these tools became available for general, non-specific use to people with no prior expert knowledge – programming, formal training in the arts, or art theory.

Although the official birth year of photography is 1839, a more or less articulated race for an image of light captured in materiality has been ongoing since the 18th century. Contributors to the final race for the invention were frequently aristocratic descendants (William Henry Fox Talbot), scientists (Sir John Herschel), and entrepreneurs with long-term but not precarious ties to the aristocracy (Nicephore Niépce). Photography hadn’t been their sole focus or source of income (Herschel was a prominent astronomer, Niépce lived off family wealth, and before experiments to capture and stabilize an image, he focused on other inventions, and Talbot was an academic scientist in several fields).

After the final invention of the daguerreotype had been examined by the French Academy of Sciences, its basic details disclosed to the public, Louis Daguerre and Nicephore Niépce’s son Isidore were granted lifetime pensions by the government.

Photography as a technological approach took over the previously formed genres of images (still life, portrait, landscape, news, and documentary). Some of them were transformed by its unique features, such as the ability to capture events

in fractions of seconds, which was critical in sports and events. Precision and the ability to capture the wavelength spectrum outside of the naked eye brought further advances in physics, astronomy, medicine, chemistry, and biology. By the end of the nineteenth century, most people in the Northern Hemisphere could afford to have their own portraits and everyday lives captured and preserved for future generations. It meant the democratization of visual memory.

It could not challenge the subjectivity frameworks of its users, such as racial studies or colonial research, because it was a tool devoid of subjectivity. After decades of living with photography, diverse practices have been legalized while others have been limited. Today’s articulated frameworks can be structured in several layers: the conditions of creating the image, the informed consent of the subject(s), and the conditions of the image’s dispersion.

The need for further monetization led photographers in the early 1920s to establish stock photography agencies. Stock photography, as vast pools of photographs with descriptions, are incorporated into machine learning training datasets 100 years later. In the early 21st century, the entities behind the GAN technology are mostly teams, be it academic research teams (for instance, the multi-levelled team behind the *ImageNet* database, which is used with the most up-to-date GAN generators as a training dataset),

self-funded teams (*Midjourney*), non-profit organizations connected with companies (*OpenAI*) or open-network based projects offering the datasets and working by donation (*LAION*). Their monetization models further rely on a combination of free tiers and user subscriptions and, eventually, AI consulting for other projects.

GAN-generated imagery, like photography, is a technological approach that employs previously formed image genres. It is built on large-scale data scraping based on the detaching and reattaching of metadata in the generation and discrimination processes. Any individual with access to a computer can use these tools without prior research. Clouds and maps of data about the world are the primary references, not the physical world itself.

Well-known artists of the past and scarcity-driven projects exist alongside anonymous *Flickr* albums for the GAN to be trained. Although there has been some consideration by the creators of the systems, such as the CEO of *Midjourney*, to incorporate some kind of consent or opt-in information from the copyright holders of the images scraped, it is practically and historically inapplicable to do it on a large scale, neither automatically

nor manually<sup>4</sup>.

Since the release of *VQGAN+CLIP*<sup>5</sup> in 2021, there has been statistical over- and underrepresentation of certain nouns in relation to certain ideas. This so-called bias, e.g., male-female representation in connection to job positions, contextual minorities, is already embedded in the datasets, based on the content and its tagging.

The teams have incorporated diverse features to de-bias the outcomes during the process of new image generation. It is based on the evaluation of the text prompts, input images, both from the dataset and the user, during the generation process and before sending the final result back to the user. During this process, for instance, when using *OpenAI*, third parties might be involved<sup>6</sup>. Users are asked to disclose the use of AI when publishing and using the visual outputs in real life, especially in the healthcare, therapy, medicine, and legal environments.

Some public institutions recognize

and regulate legally problematic situations involving the use of AI-generated material. Since January 2020, the American state of California has prohibited the use of deep fakes in pornography and election campaigns<sup>7</sup>.

## 2 The Black Box. Glancing Away From GAN Imagery

Capturing the visual, sensory world with a camera, and exacting its methods and outcomes arrives just in time for a new interpretation of it as “evidence”. Photography as a medium is formed and announced amidst the publication of Auguste Comte’s philosophy texts on positivism, which are based on rational evaluation of the outside world and logic based upon sensory experience. The camera thus challenges the idea of science, and at the same time, it appears to be a black box, where “only inputs and outputs matter”<sup>8</sup>, and the processes inside remain mysterious, hidden, as Bruno Latour put it.

What is hidden inside the box comprises not only computational activity, but also human input. To be more specific, it hides the labour of people hired through *Mechanical Turk* to tag the *ImageNet* dataset, the authors of

4 SALKOWITZ, R.: *Midjourney Founder David Holz On The Impact Of AI On Art, Imagination And The Creative Economy*. [online]. [2023-01-29]. Available at: <[www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imaginati-on-and-the-creative-economy/?sh=1808fa122d2b](http://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imaginati-on-and-the-creative-economy/?sh=1808fa122d2b)>.

5 Author’s note: The notion of embodiment and togetherness in the pandemic was the subject of the artist book *new\* not normal* (2021). The project *Affective Metadata* (2022) focused upon reduction of perceived life into data. It was created using *VQGAN+CLIP* image generation. 2022 participation on synthetic media at Uroboros Festival Praha (moderated discussion *Sensing the Synthetic*) and Humain Brno (lecture *Visitors and the Image of Soul*).

6 *Service Terms*. [online]. [2023-01-28]. Available at: <<https://openai.com/api/policies/service-terms/>>.

7 MIRELL, D. E., GELLER, J.: *AB 602 and AB 730: Curbing “Deepfakes” in Pornography and Elections*. [online]. [2023-01-30]. Available at: <<https://www.jdsupra.com/legalnews/ab-602-and-ab-730-curbing-deepfakes-in-30447/>>.

8 LATOUR, B.: *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge: Harvard University Press, 2003, para. 7.

visuals scraped from the internet. Last but not least, it processes the user activity of writing prompts, evaluating results, and evolving the outcomes. The terms of service of the aforementioned AI tools do not mention if user activity (outside of beta testing) is used for further development of the tools and for further research into human perception of word-image connections.

The fear of dissociating one's image from the physical being was observed when photography was used in cultures without rationalist traditions<sup>9</sup>. After 200 years of following a positivist paradigm, society is confronted with the impact of visual representations gaining their own trajectories.

The idea behind the in-built corrective features of GAN generators is to prevent the user from creating pornographic content or graphic depictions of violence. De-biasing GAN visuals results in directing usage frameworks in an attempt to create a balanced, friendly image of a society. However, whether human-driven machines can be free of bias is an epistemological question. Another concern is if the autonomous features of such GAN generators do bring a certain level of subjectivity – should we speak

rather of co-creation with AI than of a medium or a tool?

*AlgorithmWatch* warns that trust in the autonomous behaviour of black box(es) based on its idea of objectivity is already causing concerns in fields other than visual art – in public policies, such as digital surveillance of public spaces and algorithmic biometric data evaluation. Isn't the constant "surprise by bias" already disproving this notion?

### Conclusion

The general public is concerned about whether GAN image generation will render certain professions obsolete. David Holtz, the founder of *Midjourney*, says professional use is not the intent or plan. *Midjourney*, which runs via a *Discord* channel with around two million users, is meant more for "imagination"<sup>10</sup>.

However, its use in the moodboarding, prototyping, and drafting phases in creative industries can speed up the whole process. Looking back at photography's answers to these anxieties, we can conclude that it rendered some work positions unnecessary while creating others. The disproportion of wealth is rather a feature of the sociopolitical system

than something to be solved by a tool without subjectivity. Similarly, we can think of reconfiguring work with GAN-based visual imagery, but precarity and poverty, similar to wealth, will be taking on new forms. Authorship is needed mostly for social capital and for monetizing (as well as the concept of scarcity). The possibility of being excluded from data scraping and therefore creating a feeling of value based on scarcity is similar to creating scarcity by delivering limited editions of reproducible art. Another approach is to incorporate AI into already stable forms of image distribution, as Shutterstock, one of the largest stock photography agencies with 300 million images available (and claiming to grow by another 200,000 daily), is attempting to do. Blockchain technology, as already used in the NFT world, can mark an "original" to maintain this function if we, as a society, decide we want it. Although the significance of the specific applications for GAN image generation mentioned above might lessen in the years to come, they mark steps toward new encodings of reality and new aspects of visual literacy.

9 STROTHER, Z.: "A Photograph Steals the Soul: The History of an Idea." In PFEFFER, J., CAMERON, E. (eds.): *Portraiture and Photography in Africa*. Bloomington : Indiana University Press, 2013, pp. 177-212. [online]. [2022-10-07]. Available at: <[https://www.academia.edu/36719998/A\\_Photo-graph\\_Steals\\_the\\_Soul\\_The\\_History\\_of\\_an\\_Idea](https://www.academia.edu/36719998/A_Photo-graph_Steals_the_Soul_The_History_of_an_Idea)>.

10 SALKOWITZ, R.: *Midjourney Founder David Holz On The Impact Of AI On Art, Imagination And The Creative Economy*. [online]. [2023-01-29]. Available at: <[www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=1808fa122d2b](https://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=1808fa122d2b)>.

## Bibliography

Data. [online]. [2023-01-30]. Available at: <<https://www.oxfordlearnersdictionaries.com/definition/english/data?q=data>>.

LATOURE, B.: *Science in Action : How to Follow Scientists and Engineers Through Society*. Cambridge : Harvard University Press, 2003.

MIRELL, D. E., GELLER, J.: *AB 602 and AB 730: Curbing "Deepfakes" in Pornography and Elections*. [online]. [2023-01-30]. Available at: <<https://www.jdsupra.com/legalnews/ab-602-and-ab-730-curbing-deepfakes-in-30447/>>.

PASQUINELLI, M.: Metadata Society. In BRAIDOTTI, R., HLAVAJOVA, M. (eds.): *Posthuman Glossary*. London : Bloomsbury, 2018, pp. 253-256.

SALKOWITZ, R.: *Midjourney Founder David Holz On The Impact Of AI On Art, Imagination And The Creative Economy*. [online]. [2023-01-29]. Available at: <[www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=1808fa122d2b](http://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=1808fa122d2b)>.

Service Terms. [online]. [2023-01-28]. Available at: <<https://openai.com/api/policies/service-terms/>>.

STROTHER, Z.: "A Photograph Steals the Soul": The History of an Idea". In PFEFFER, J., CAMERON, E. (eds.): *Portraiture and Photography in Africa*. Bloomington : Indiana University Press, 2013, pp. 177-212. [online]. [2022-10-07]. Available at: <[https://www.academia.edu/36719998/A\\_Photo-graph\\_Steals\\_the\\_Soul\\_The\\_History\\_of\\_an\\_Idea\\_](https://www.academia.edu/36719998/A_Photo-graph_Steals_the_Soul_The_History_of_an_Idea_)>.

## Author

Mgr. art. et Mag. Art. Martina Šimkovičová, ArtD.  
simkovicova@fedu.uniba.sk

Faculty of Education  
Comenius University Bratislava  
Šoltésovej 4  
81108 Bratislava  
SLOVAK REPUBLIC

## Profile of the Author

Martina Šimkovičová (1988) is a graduate of photography and visual arts at the academies in Bratislava and Vienna. At the moment, she teaches photography related subjects at the Department of Fine Art Education at the Faculty of Education, Comenius University in Bratislava. She co-founded Robota - Center for Advanced Studies with Kristian Lukić.