Thomas Smits

Visual Memory of Protest in the Age of Generative Artificial Intelligence

1 Introduction

Looking confidently into the camera, a young man is taking a selfie on a wide street. Instead of the expected cars, we see a tank on a zebra crossing. The image looks photorealistic but taking a closer look we notice several visual inconsistencies. For example, why do the cars on the right side of the street seem to blend into each other, and what purpose does the excessively large streetlight serve? The image looks real, familiar, and 'off' at the same time. Around 20 September 2023, the 'tank man selfie' stirred a brief controversy on social media.[1] Following its posting by the anonymous user Ouroboros696969 on Reddit, the image, generated using the image generation AI model Midjourney, quickly went viral on the internet. For a brief period, the Google Knowledge Graph, which connects web entities, displayed the image as the top visual result for people searching for 'tank man': the iconic photograph of a man defying a column of tanks during the Tiananmen Square protests of 1989.

As a result of the rapid development of image generation AI, highly realistic artificial images are rapidly proliferating on the internet. The increasing use of these models blurs the line between what is human-made and artificial or AI-generated in the online world. Although Google promptly removed the tank man selfie from its knowledge graph, its brief appearance foreshadows a future in which differentiating between 'natural' and generated cultural content might become impractical for the corporations that control access to the online world and, as a direct result, exceedingly hard for individual users. This chapter delves into a particular facet of the rise of AI-generated images: their impact on our visual recollection of the past. More specifically, it discusses how these new image generation capabilities might influence the visual memory of protests: the role that images play in the memory *of* and memory *in* social movements.

1 https://web.archive.org/save/https://www.reddit.com/r/midjourney/comments/129pcy8/selfie_tank_man_tiananmen_square_1989/.

https://doi.org/10.1515/9783111439273-052

2 AI-generated images

In popular discourse, image generation models are often perceived as autonomously producing texts and images, separate from human involvement. Work in critical AI studies has argued that, in actuality, humans and models write texts or create images together; they both have agency in the generation process (Roberge and Castelle 2021). As a result of this distributed agency, together with Rik Smit and Samuel Merrill, I have previously argued that AI systems can only remember in collaboration with humans (Smit et al., under review). If we want to examine the role of image generation models in this human-AI remembering, we must first understand how they work.

Like any other form of machine learning, image generation models require substantial amounts of training data. Although we might assume they are trained solely on visual material, their training datasets consist of millions of image-text combinations, pairing images with textual descriptions. While they have different brand names, such as Midjourney, DALL·E, or Stable Diffusion, most image generation models are based on the phenomenon of diffusion. In the realm of data, diffusion describes the process of data transitioning from a simple to a complex distribution. Image generation models start by learning how to diffuse the image-text pairs in their training data, gradually adding noise to them until they approach a random data distribution. After this so-called forward process, they attempt to reverse the diffusion, learning the mathematical operations that return the diffused image-texts to the simple original distribution.

While image generation models are frequently described as *generating* images, it would be more apt to describe them as *denoising* images. The combination of images and text in the training data is essential for the denoising process. When we provide a prompt to an image generation model, it uses this textual description as a key to denoise an image of random visual noise. Because random visual noise is never the same, the image generation model can generate endless denoised images based on the same prompt.

The quality of training data influences how well an image generation model can denoise specific types of images based on particular prompts. For example, because models have seen thousands of different digitized versions of Van Gogh paintings, they are very capable of producing images 'in the style of Van Gogh.' More problematically, models also learn all sorts of biased data patterns. Given the prompt 'a photograph of a doctor,' models are, for example, likely to produce white male medical doctors (Nicoletti and Bass 2023). As Birhane et al. (2021) show, bias often stems from common combinations of images with texts on the internet. Words like 'latina' or 'nun' are often associated with pornographic images. During training, image generation models learn to reproduce these kinds of problematic

patterns. The companies behind image generation models use several ways to mask this kind of problematic behavior. The automatically generated prompts used to generate the figures of this chapter are a good example of this practice. OpenAI, the company behind DALL·E, automatically adds words like 'people of various genders and descents' (Figure 1) to prompts to make the images more diverse. Without these additions, the model is likely to depict 'persons' as white males.

It is important to note that image generation models learn countless patterns that are not considered to be problematic. Just as an image generation model has a visual idea of a doctor (white/male/lab coat), it also has an idea of 'image of protest.' Take, for example, the reference to 'historic protest images' in the prompt used to generate Figure 2. Based on the work of Bender et al. (2021), who studied text generation models (GPT3), we can describe image generation models as visual stochastic parrots. Unlike human image producers, such as painters or photographers, diffusion models lack an understanding of image creation and specific visual intentions; they simply utilize probabilistic patterns to denoise images based on text prompts. As a result, the 'creativity' of image generation models is thought to be limited to combining visual styles and patterns that can be found in their training datasets. To put it in other words, models are very good at transferring a specific style to another image – a manga-style version of Picasso's *Guernica* – but are unable to produce an entirely new visual style. We can, of course, wonder how limited this kind of creativity is. After all, most humans are also unable to produce images in a new visual style. We can explore this question by examining the potential of image-generation models in the visual memory of protest.

3 AI-generated visual memory of protest

In her seminal article "Remembering Hope," Rigney notes that, at first glance, memory and activism may seem poles apart: "the former [being] oriented towards the past and the latter towards the future" (2018a, 371). Upon closer examination, we quickly realize that they are "deeply entangled" (371). Rigney describes this entanglement as a 'memory-activism nexus,' which consists of three interrelated elements: memory activism, the struggle to steer future remembrance, memory in activism, which describes how the memory of earlier movements shapes the tactics and demands of movements in the present, and memory of activism, which deals with how movements are remembered.

Social movement scholars have shown that protests are profoundly visual in character (Doerr et al. 2013; Mattoni and Teune 2014). The political power of social movements is derived from their visibility, as it forces political actors, such as poli-

ticians or political parties, to acknowledge and act on their demands (McGarry et al. 2020). As a result, images play an important role as "carriers of memory" (Rigney 2015a) in the memory-activism nexus: they shape how movements are perceived and create links between different movements in time and space.

What role can AI-generated images play in the visual memory of protest? It is important to acknowledge that most research in this field, including my collaboration with Professor Rigney (Rigney and Smits 2023; Smits 2021), has primarily focused on photographs. However, it's crucial to remember that paintings, illustrations, and other non-photographic visual mediums have also played significant roles in the intersection of memory and activism. For instance, consider Eugene Delacroix's renowned *Liberty Leading the People* (1830). Originally painted to commemorate

Figure 1: AI-generated image (DALL·E). Original prompt "Liberty leading the people, but the figures are climate activists." Automatically-generated prompt: "Illustration in the style of a classic oil painting depicting a scene reminiscent of 'Liberty Leading the People.' A woman of African descent stands tall, leading a group of diverse climate activists. Her dress is intricately made of plants, flowers, and greenery, symbolizing nature. She waves a flag with the Earth on it. People of various genders and descents surround her, each passionately advocating for the environment with placards and eco-friendly symbols. The backdrop features a blend of nature and urban setting, with wind turbines and solar panels."

the July Revolution of 1830, the painting has inspired numerous artists and activists to make versions that relate to causes and movements in their own time (Smits 2020). The role of Delacroix's painting – and its many adaptions by later movements – prompts us to question the distinction between real and artificial visual carriers of memory. As the previous section argued, we should neither over- nor underestimate the agency of AI systems in the generation of images. Just as Delacroix used a paintbrush, activists might use an image generation model to place their cause in the chain that *Liberty Leading the People* started (Figure 1).

While it may be too early to predict their precise impact, it seems likely that image generation models will change the agency of humans in the visual memory of activism. Even if we are not brilliant painters, these models allow us to visually

Figure 2: AI-generated image (DALL·E 3). Original prompt: "Photo echoing the powerful dynamic of the 'Taking a Stand in Baton Rouge' photograph but set against the backdrop of the Paris Commune of 1871." Automatically-generated prompt: "Photo of a defiant woman with light skin confronting a barricade of French soldiers from the era of the Paris Commune of 1871. The soldiers are in historical military attire, holding their rifles at the ready, against a backdrop of the Parisian cityscape showing signs of upheaval and rebellion. The woman is dressed in contemporary formal attire, standing out amidst the turmoil of the time, her expression one of unyielding determination reminiscent of historic protest images."

interpret past activism in ways that adhere to how we 'see' protests in the present. The Tank Man Selfie is a good example of this. Instead of the original anonymous protester, whose identity remains disputed to this day, image generation models empower us to envision him as a self-assured individual who actively shapes his own image. The possibilities are endless. Maybe we want to imagine the iconic 'Taking a stand at Baton Rouge' photograph during the Paris Commune of 1870–1871 (Figure 2) or maybe we need a design of a mural that shows a 'famous Cuban revolutionary' – most models will refuse to generate images of famous persons – as an eco-socialist (Figure 3).

Figure 3: AI-generated image (DALL·E 3). Original prompt: "A photograph of a mural showing a famous Cuban revolutionary as an eco-socialist." Automatically-generated prompt: "Create an image of a mural on a brick wall depicting a revolutionary figure in a beret, with a stylized beard and long hair, who has a resemblance to a 1960s Latin American guerrilla leader. The figure is portrayed as an eco-socialist, surrounded by lush green foliage and holding a globe in one hand to symbolize care for the environment. The mural includes symbolic elements such as wind turbines, solar panels, and a variety of animals, representing a harmonious relationship between humans and nature."

4 Conclusion: endlessly stochastic?

While the possibilities of image generation models might be endless, they are also limited by the stochastic patterns in the data on which they were trained. Bridle describes how this leads machine learning models to play an active role in projecting the present (and recent past) onto the future: "That which is gathered as data is modeled as the way things are, and then projected forward – with the implicit assumption that things will not radically change or diverge from previous experience" (2018, 35).

Bridle's notion of the inherently conservative nature of machine learning models may temper our enthusiasm for applying them to various elements of the memory-activism nexus. Most notably, their use could potentially result in a stagnant and static visual culture of protest, which, in turn, might lead to ineffective social movements that are unable to make their demands visible. In the *Visual Memory of Protest* (2023), we argue that aesthetics play a crucial role in understanding why some protest images become iconic while others are swiftly forgotten. To gain attention, images must possess a striking quality, being surprising enough to truly capture our notice. However, this capacity is always relational. After all, being striking "entails making a difference with respect to a tradition" (Rigney and Smits 2023, 13). Consequently, powerful images of protest activate the visual memory of earlier movements but also deviate from it: "Images of earlier protests create the horizon of expectations for later ones and offer a benchmark for noting new departures" (Rigney and Smits 2023, 13).

Image generation models make it easier than ever for movements to communicate visually. However, a critical question remains: can (memory) activists use these new models to strike the aesthetic balance "between déjà vu and the strikingly new" that has produced so many powerful images of protest (Rigney and Smits 2023, 14)? At the end of this essay, it is important to note that the possible (mis)use of image generation by activists will play out in praxis rather than theory. In April 2023, Micah White released ProtestGPT: an "activist AI [. . .] that generates unique and unconventional protest ideas on any given topic" (White 2023). White suggests that image generation models could similarly be harnessed to create "political images" capable of triggering "waves of uprisings" on social media (White 2023). Based on the stochastic nature of these models and the aesthetic quality of memorable images of protest, it seems improbable that AI could indeed generate "unique and unconventional" pictures (White 2023). While humans can utilize image generation models to generate countless stochastic variations on a theme, it appears challenging for them to craft a fresh visual narrative tone that resonates both with activists and society at large.