**Daein Cheong**

# Experiencing Programmed Magic : AI Image Generation and the Photo-Roman

## Introduction

This text does not aim to define artificial intelligence theoretically, nor does it seek to offer a systematic exposition of Vilém Flusser's thought. I am not a Flusser scholar, but an architect who is exploring ways of unsettling the conventional assumptions through which vision has defined space, through perspective, photography, and various media apparatuses. My work and research begin from the premise that vision is not neutral, but one that has been constructed through specific representational systems and technological dispositive. Within this research context, this text takes the form of a working report based on the production of an eleven-minute photo-roman[1] presented in my recent solo exhibition in Seoul. It documents how what Vilém Flusser described as the "programmed magic" of technical images becomes particularly visible in AI-based image generation, where a short textual prompt can suddenly produce a visually coherent image while the complex computational processes that generate it remain inaccessible to the user.

## Project Background

The photo-roman, titled *Eyes of Epoché*[2], takes the form of a thought experiment inspired by Étienne Bonnot de Condillac's *Treatise on Sensations* (1754), which imagines perception through the figure of a statue that gradually acquires its senses one by one; Denis Diderot's *Letter on the Blind* (1749), which rethinks the relationship between perception and knowledge through the experience of blindness; and Frank Jackson's *Mary's Room* (1982), which exposes the gap between knowledge and experience

[1] For Vilém Flusser, the photo-novel represents a paradigmatic form of technical images, in which narrative and text are absorbed into the operational logic of the apparatus. As he writes, "the magic and ideology inherent within them are translated into the programmed magic of technical images (for example in the form of photo-novels)" (Flusser 2000: 19).
[2] https://www.youtube.com/watch?v=up5fvlDzkA4

through the figure of a subject who possesses complete scientific knowledge of color without ever having seen it. Briefly summarized, the film centers on a woman who, from birth, has been completely paralyzed as a result of a scientific experiment and who has perceived the world solely through vision within a fully controlled environment. After later regaining bodily movement, she finds herself unable to intuitively comprehend spatial relationships such as distance, depth, and orientation, having spent her life perceiving the world from a fixed viewpoint. An architect attempts to teach her how to see through architectural representations such as plans, sections, and elevations, but these attempts repeatedly fail. Through this failure, the photo-roman functions as a thought experiment that reveals how the modern assumption that vision naturally defines space is not a universal capacity, but a system constructed through specific conditions and forms of learning[3].

## Storyboard as an Intermediate Apparatus

Once the script of the fiction had acquired a certain structural outline, two stages of storyboards were produced in order to visualize each scene before developing it into the final photo-roman format. In the first stage, the storyboard was conceived as a collage based on collecting, cutting, and juxtaposing existing architectural and artistic images, into which rough sketches drawn by myself were also inserted (fig.1.). Rather than being organized according to a unified graphic rule, it functioned as a provisional device in which fragments of images from different sources were loosely assembled in order to explore the atmosphere and structure of the gaze that each scene should possess. Images were not so much objects newly produced as objects selected and arranged, and the author remained in the position of judging and constructing the relationships between these fragments.

The second storyboard did not fundamentally alter this working method. Rather, it was closer to an attempt to maintain the direction established in the first version while slightly expanding the range of references and unifying the overall background tone in black, thereby forming a more condensed atmosphere (fig.2.). In this sense, the difference between the two versions was less a change in the logic of production than a matter of adjusting the same visual hypothesis with more density. The visual tension of black and white, the sense of a suspended gaze, and the atmosphere of unreality

[3] A broader historical and theoretical discussion on vision and perspective, including the formation of perspectival systems from ancient Greece through the Renaissance and Baroque, Panofsky's interpretation of perspective as a symbolic form, visual experiments such as anamorphosis, and phenomenological approaches to visual experience in architecture, has been developed in greater detail in the author's previous study (Cheong 2024). That study examines how different visual regimes shaped architectural representation and spatial imagination prior to the emergence of technical images. The present paper instead focuses specifically on AI-based image generation and its relation to Flusser's notion of programmed magic.

became more pronounced in the second version, but the collage-based method of collecting and juxtaposing images itself remained essentially the same.

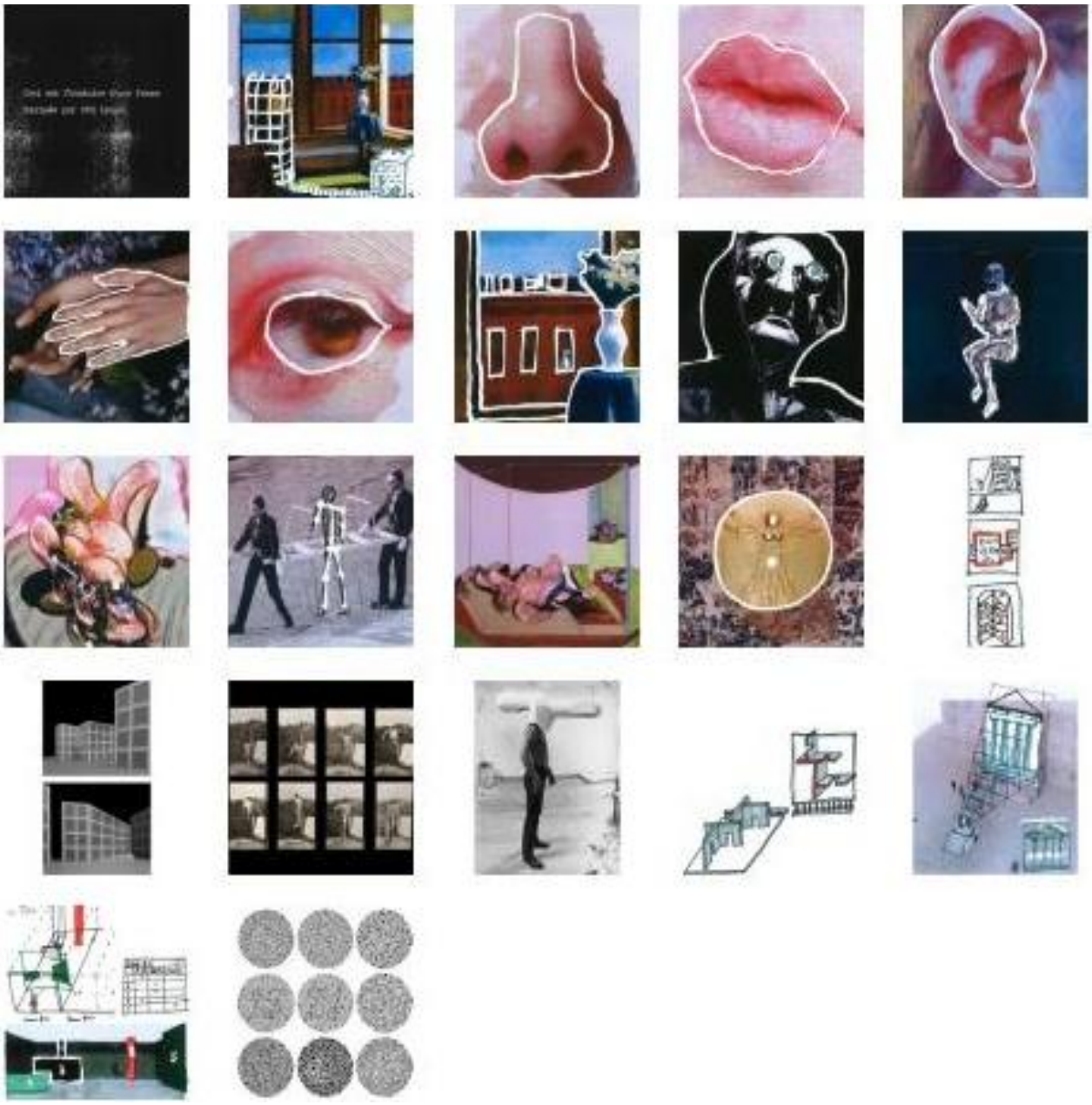

fig.1. : Storyboard, first version.

fig.2. : Storyboard, second version.

These storyboards presupposed the possibility of later developing into a photo-roman produced by staging real spaces and figures and photographing them. In other words, the collages and sketches were not the final results but could function as an intermediate stage for physically constructing the mise-en-scène, setting the camera viewpoint, and photographing each scene. Had this approach been realized, the photo-roman would have consisted not of hand-drawn lines or the juxtaposition of collected images, but of photographic images generated through the technical apparatus of the camera.

However, when I attempted to begin the actual production of the photo-roman after the two stages of storyboards had been organized, it became clear that although these images were sufficient

to suggest the atmosphere and direction of the scenes, they were not precise enough to lead directly to a concrete composition that could be photographed. I had continually kept Chris Marker's *La Jetée* (1962 ) in mind as an important visual reference, and the storyboards likewise reflected, to some extent, its black-and-white tension, suspended gaze, and atmosphere oscillating between reality and science fiction. Yet when the production process actually began, such references proved sufficient as a guide for imagining the mood of the scenes, but not as production drawings capable of organizing the actual staging of the scene and the camera composition. In particular, because I had no prior experience in filmmaking or photographic direction, when I attempted to produce the photo-roman it was difficult to grasp how to concretize the positioning of actors, the organization of space, the setting of viewpoints, and the connections between scenes.

## AI Image Generation in Practice

In this situation, image generative AI was introduced into the production process. It was not decided from the outset that AI would be used as a means of producing the final images. At first, it was used only experimentally, and I originally considered the possibility of employing it as a kind of third storyboard. However, once images were actually generated, they began to present the atmosphere of the scenes in a more concrete and spatially articulated manner than the scenes I had only vaguely imagined during the storyboard stage. Above all, the dreamlike and poetic science-fiction sensibility characteristic of La Jetée was reproduced with surprising effectiveness (fig.3.). At this point, AI began to function not merely as a tool for the auxiliary visualization of abstract visual hypotheses, but rather as a kind of productive apparatus that allowed me to grope toward scenes that had not yet been clearly defined.

Let us take the first scene as an example. In the first storyboard, this scene began with an image based on a painting by Edward Hopper, onto which I directly drew a grid-like chair structure and mechanical devices. In the second storyboard, images of Frida Kahlo's body and photographs by Herlinde Koelbl were added, each functioning as an analogy for an immobile body and a fixed condition. However, the concrete method for organizing these references into a single scene within a black-and-white tension and an unreal atmosphere similar to that of La Jetée remained unclear. It was precisely at this point that, when images were generated through AI, the fragmentary references that had remained at the level of collage and sketch began to appear as a single condensed scene more easily than expected. Above all, it became possible to avoid the uncertainties, labor, and technical inexperience that would have had to be confronted in the process of physically constructing a mise-en-scène and photographing it, and the work progressed relatively quickly.

fig.3. : AI-generated images.

However, this process was by no means entirely automatic or smooth. The collages and rough sketches of the storyboard were sometimes too ambiguous for the generative model, while in other cases they were interpreted too literally. The black-and-white tension, the suspended gaze, and the uneasy atmosphere oscillating between reality and science fiction that I considered important were not fixed through a single attempt. Even when the same prompt was repeated, the spatial relationships of the scene, the positions of the figures, the viewpoint, and the contrasts of lighting easily shifted. In some cases, the prompt had to be revised nearly a hundred times in order to obtain a desired image, and since four images were generated with each prompt, it was necessary to judge and select which result came closest to the scene I was seeking. In this sense, AI image generation was not the simple automation of transferring the responsibility of production to a machine, as is often imagined.

Rather, this process functioned as an exploration practice of adjusting what might constitute an appropriate scene, as well as an experience of active intervention. For this reason, the concept of the functionary that Flusser introduces in his analysis of the camera also appears relevant for describing

AI[4]. As in the case of the camera, in the process of AI image generation the user controls the exterior of the apparatus, namely its input and output, while at the same time being controlled by the apparatus due to the opacity of its interior.

## Photography and AI: Two Regimes of Image Production

AI clearly presented a strong incentive, in that it allowed photorealistic images to be obtained without having to undertake the physical process of photographic production. In the end, I chose the AI-generated images, originally intended to function like a third storyboard, as the final results. The introduction of AI was both an experiment with a new tool and a choice to bypass the complexity and burden of actual production. However, something was also clearly lost through working with AI.

Above all, the ambiguity, incompleteness, and the multiple layers of interpretation that had emerged from the tension between images in the earlier stages of collage and sketch were replaced by a more clearly defined sceneness. The diffuse atmosphere and possible narrative trajectories that had previously emerged from the interplay of different visual references became fixed as a more condensed and immediately legible scene. Yet this was less the faithful execution of an intention I had fully predetermined than the convergence of one possibility among many through a particular apparatus.

Of course, even if I had directly carried out the photographic shooting for the photo-roman, much of the specificity of the original storyboard would likely have been diluted. In this sense, both photography and AI-generated images share similar conditions as technical images that stabilize references into a single scene and transform loosely open relationships into a more integrated surface.

However, there is also an important difference between photography and AI image generation. This difference lies in the manner and speed through which such convergence and stabilization take place. In photography, a series of stages—shooting, editing, and selection—are temporally separated, and within this process hesitation, revision, and reconfiguration can persist for a relatively long time. One first constructs an actual space, arranges figures and objects, and records the scene through a camera, and only afterward reviews the result. Judgments therefore take place across several stages. This temporal separation does not simply indicate a difference in procedure. Problems that were not sufficiently visible at the moment of composing the scene may become apparent only after the photograph is taken, and the judgment of whether the captured image is appropriate may only be made

[4] "The functionary controls the apparatus thanks to the control of its exterior (the input and output) and is controlled by it thanks to the impenetrability of its interior. To put it another way: Functionaries control a game over which they have no competence. The world of Kafka, in fact." (Flusser 2000: 28).

during the later stages of editing and selection. In this sense, within photography a state of instability is not immediately resolved but can persist for a certain period as it passes through these separate stages.

In contrast, within an AI image generation environment the resulting image appears almost immediately, and judgment and adjustment follow rapidly afterward. Even when the generated image does not sufficiently reach the atmosphere of the scene, the spatial relationships, the viewpoint, or the arrangement of figures that I was seeking, this does not remain as a prolonged uncertainty but is quickly transformed into a condition for the next attempt. In other words, failure is treated not as a trace but as data for adjustment. In this sense, AI image generation resembles an apparatus that converts incompleteness into rapidly replaceable variables rather than preserving it. For this reason, by choosing AI image generation instead of photography, the sense of personal intervention involved in physically composing and photographing an image was diminished, which in turn weakened the sense of authorship formed through the physical process of creation.

## Conditions that Limit Play in the AI Apparatus

For Flusser, play does not mean simply following or rejecting an apparatus, but rather a practice that attempts to generate unexpected differences and combinations within its program[5]. Under the conditions in which the apparatus of technical images tends to converge the world toward ever greater efficiency and predictability, such play can be understood as a negentropic possibility that introduces exceptions and differences into that flow. In this sense, what matters for Flusser is not a fixed human type positioned opposite the functionary, but the practice itself of handling the apparatus playfully from within and testing the limits of its program.

However, in the process of producing the photo-roman discussed in this paper, I was closer not to an artist performing such play, but to a user who employed image generative AI and thus assumed the position of a functionary within the apparatus. In this process, there was no deliberate

[5] "With one exception: so-called experimental photographers — those photographers in the sense of the word intended here. They are conscious that image, apparatus, program and information are the basic problems that they have to come to terms with. They are in fact consciously attempting to create unpredictable information, i.e. to release themselves from the camera, and to place within the image something that is not in its program. They know they are playing against the camera. Yet even they are not conscious of the consequence of their practice: They are not aware that they are attempting to address the question of freedom in the context of apparatus in general." (Flusser 2000: 81)

attempt to disrupt or subvert the apparatus. This raises the following question. As experimental photographers sought to free themselves from the apparatus of the camera, is playful creative activity through AI also possible? The reasons why this is very difficult can be explained from three perspectives.

The first concerns technological acceleration and the resulting instability of the apparatus. Generative AI is developing at an extremely rapid pace, with new models, interfaces, and modes of use constantly appearing and disappearing. Rather than exploring a single apparatus over time and testing the limits of its program, users are placed in a position where they must continually adapt to a changing environment. Under such conditions, it becomes difficult for a stable relationship with the apparatus to accumulate and for practices that playfully test its internal program to emerge. The drastic reduction in glitches and malfunctions in AI-generated images over the past two years supports this observation. Furthermore, research in artificial intelligence has already expanded into discussions surrounding the possibility of AGI (Artificial General Intelligence). This uncertain technological horizon renders fluid the very judgment of where current generative models remain tools and where they might begin to shift toward a new cognitive system. In such a rapidly shifting technological environment, users are more likely to remain in a position of adapting to new systems rather than attempting to disrupt or explore an apparatus over time.

Second, the issue lies in the scale of the data handled by AI and the conditions under which that data has been collected. In the earlier experimental environments of AI, the possibility of directly constructing or training a system was relatively more visible. By contrast, the generative AI widely used today consists of commercial models that are pre-trained on large-scale datasets[6]. AI systems that interact with users by extracting patterns from massive amounts of information in advance and performing enormous quantities of repetitive computation are, in this sense, an even darker black box than the camera, making intervention far more difficult. Compared to cameras, which can be relatively easily disassembled and whose components can be understood in order to allow experimental intervention, human-scale intervention in AI tends to be limited to the adjustment of inputs and outputs, and responding to the probabilistic computational structures within the apparatus is structurally constrained. The experience of programmed magic that Flusser describes is therefore intensified through AI.

[6] A considerable body of critical scholarship has already questioned the belief that big data represents an objective and absolute totality of reality. In particular, scholars have argued that the very notion of "raw data" is itself an oxymoron, given that the processes through which data are collected, selected, classified, and processed are shaped by specific social and technical conditions (Gitelman 2013).

The third concerns the ontological uncertainty of AI's status. Today, generative artificial intelligence is still often understood as a "tool", yet it remains unclear whether it should remain at the level of a tool or be understood as a new form of intelligence that partially competes with or substitutes for human cognitive activity. The architectural historian and theorist Mario Carpo has emphasized that generative AI might function less as a rule-based problem-solving device than as a system that probabilistically recombines a vast archive of precedents and forms[7]. From this perspective, AI can be understood less as a creative subject than as a powerful tool for rapidly searching and rearranging forms that already exist. On the other hand, some recent discussions argue that "intelligence" cannot be reduced to the ability to generate many possible answers. Rather, intelligence also presupposes the capacity to recognize the limits of one's own knowledge and to reflect upon that ignorance, a condition described as learned ignorance[8]. From this perspective, what is commonly called AI may not properly qualify as intelligence at all, precisely because such systems lack any awareness of what they do not know. Precisely for this reason, it remains unclear whether AI should be treated merely as a type of apparatus or understood as a more complex cognitive counterpart, and as a result the possibility of treating AI as an object of play in Flusser's sense remains theoretically unstable.

## Conclusion

Nevertheless, these conditions do not mean that the role of the author has entirely disappeared from creative practices involving AI. At least in my own photo-roman project, the moment at which I most strongly experienced authorship lay not in modifying or selecting the outputs produced by the apparatus, but in composing the inputs that generated them. In particular, these inputs were not neutral or impersonal commands. The collages and sketches produced prior to the use of AI image generation, as well as the selection and arrangement of the reference images that made them possible, had already been carried out at a highly personal level. Likewise, the thought experiments concerning the atmosphere of the scene, spatial relationships, the structure of the gaze, and the direction of the narrative were shaped by my own interests, preferences, and imagination. In this sense, the prompt functioned less as a purely functional instruction than as a textual interface through which these accumulated

[7] "As a welcome side effect, this use of computer-generated art has revived notions and theories of imitation in the arts, and of artistic styles, which had been eluded or repudiated by modernist art theory." (Carpo 2023)

[8] The concept of learned ignorance itself originates in Nicholas of Cusa's *De Docta Ignorantia*, and it has recently been invoked again in contemporary AI debates that emphasize the connection between human knowledge and the awareness of its own limits. (Svensson 2023)

personal reflections and references were translated into the conditions of technical images. The generated images were ultimately selected outcomes, but reaching those outcomes required prior personal judgments about what to imagine and which relationships to consider significant. At this point, the sense of authorship remains present, and creative work in the environment of generative AI becomes less an act of directly producing images than a process of reflecting on what questions to pose and what conditions to establish.

In this sense, the age of AI image generation may in fact signal a moment in which textual intervention, in other words the ability to organize the structure of thought and the direction of imagination with precision, becomes extremely important. The prompt is not merely a command but an interface that mediates the relationship between the apparatus and the user, a channel through which the user's intentions and imagination are translated into technical images. While it is often predicted that generative AI will simply replace human labor more quickly and at lower cost, it is precisely for this reason that differences which are difficult to replace, such as the accumulation of personal references, the depth of thought, and the manner of posing questions, may become even more significant. Of course, such interventions do not automatically lead to the form of play described by Flusser. However, if play within the apparatus of generative AI is possible, the only level at which it may emerge is likely through the unexpected differences and combinations that arise at the level of input. In this sense, the task ahead lies not in mastering the technology that produces final images, but in exploring how textual practices that pose new questions within the program of the apparatus and open unforeseen possibilities for images might be developed.

## References

Carpo, Mario (2023). A Short but Believable History of the Digital Turn in Architecture. e-flux Architecture.

Cheong, Daein (2024). The Gaze: An Architectural Fiction Reconceiving Sight in Architecture. AMPS Proceedings Series, Issue 41.

Flusser, Vilém (2000). Towards a Philosophy of Photography. Trans. Anthony Mathews. London: Reaktion Books. First edition in German 1983.

Gitelman, Lisa (Ed.). (2013). Raw Data Is an Oxymoron. Cambridge, MA: MIT Press.

Svensson, Jacob (2023). Artificial intelligence is an oxymoron. AI & Society 38(1) (2021): 363–372.