Sophia Chefalo and Vinzenz Aubry
The Non-deterministic Apparatus
From Index to Icon

Introduction

The photographs we create have been undergoing a quiet yet radical transformation in their relation to the world. Digital cameras, particularly smartphone cameras, which we take as our guiding example for this essay, reveal how far indexicality is being displaced. With a recent Samsung Galaxy smartphone, one can no longer photograph the real moon[1]. Instead, when trying to do so, the result contains details that were never optically present. An artificial moon synthesized from a training dataset of previous moons is unknowingly superimposed onto the final image. These moons no longer point to a singular time and place in the physical world—they appear real through photographic visual language, invoking "photographicness," yet function through probabilistic generation.

What happens when the photographic image becomes statistical rather than indexical? Photography has long operated through what Roland Barthes called 'that-has-been' (ça-a-été)[2] the trace of a specific moment fixed from time. Computational and AI imaging transform this relationship from 'that-has-been' to 'this-could-be': from singular trace to probabilistic synthesis. The gesture of single-bodied image-making, which once produced the photographic trace, is evermore superseded by processes that require no bodily mediation, removing the individual's need to search, decide, or even press a shutter.

This essay traces the exponential erasure of the gesture of photography[3], the embodied human-apparatus relation on which Vilém Flusser grounded his theory of *Technical Images*. We follow this transformation through computational supplementation, the replacement of the shutter via text prompts, and machine learning models that average singular traces into statistical multitudes. This ontological shift introduces self-referential, post-*corporeal*[4], post-light icons that only intensify the question of how to read images.

[1] While this feature can be disabled, the prevalence of default settings suggests most users encounter it in its active state.
[2] Roland Barthes, *Camera Lucida: Reflections on Photography*, trans. Richard Howard (London: Vintage, 2000), 77.
[3] Vilém Flusser, *Gestures*, trans. Nancy Ann Roth (Minneapolis: University of Minnesota Press, 2014), 72.
[4] Sophia Chefalo and Vinzenz Aubry, "After Technical Images: Towards a Theory of Post-Technical Imaging," *Flusser Studies* 40 (November 2025).

The Program Transcends from Camera to Prompt

To understand how contemporary machine learning (ML) affects image generation, indexing, and judgment, we will first revisit what Flusser meant by the "apparatus" and its embedded "program," before examining its contemporary development towards dataset integration.

Flusser identified the camera as an apparatus distinct from tools and machines. For a tool — a hammer, a brush, a chisel—the human is constant, and the tool is variable. The tool extends human capability but remains subordinate to human intention. For a machine — a loom, a printing press, an assembly line — the machine is constant, and the human is variable. The machine imposes its rhythm and logic on human labor. Flusser introduced the apparatus as a necessary third category which exists in an "intricate co-relation of functions" where the apparatus seems to conform to the operator's intentions; however, the operator can only want what the apparatus can do.[5] The photographer does not use the apparatus as a neutral tool, nor do they simply operate it as a machine. Rather, they work *inside* it, playing with and against its programmatic constraints. Flusser named the concealed processes and constraints within the apparatus the "black box".

The photographer can see the input (the scene before them, visible to the human eye) and the output (the resulting photograph), yet the process between these two points is illegible and complex. The photographer is not (and never has been) in full control of the photograph: optical physics determines how light bends as it passes through lenses; mechanical constraints limit shutter speeds and aperture ranges; chemical processes govern how silver halides react to photons. This black box is amplified in digital cameras, where sensor design additionally dictates how light-sensing cells sample, quantize, and interpolate information from photons. Then, processing algorithms determine what is enhanced or discarded, and finally, compression formats intervene between the real world and the photographic image.

Flusser claims that the camera already contains all possible photographs within its technical constraints, which he called the photographic universe. The photographer does not create but rather plays within a predetermined space of possibilities, appearing as creative freedom. Flusser argued that photographers become *functionaries* of the apparatus, believing themselves to be free agents while actually serving the program embedded in the camera. The photographer explores its possibilities, but cannot transcend its boundaries: apertures, shutter speeds, film stocks, and lenses can be adjusted, but these are all moves within the game, not changes to the game itself. They are free to discover what the program already contains, but not free to produce what the program excludes. The camera mediates between the photographer and the world, actively shaping what can be seen, captured, and imagined as an image.

[5] Vilém Flusser, "Comments on 'Generative Photography: A Systematic Constructive Approach," *Leonardo* 20, no. 2 (1987): 357.

Computational Smartphone Photography

Today, the camera's program has expanded to proactively transform the information that it records. In 2018, Samsung introduced a "Scene Optimizer" feature that uses machine learning to recognize objects and adjust the camera settings accordingly. Since the Galaxy S21 series, it has been able to recognize the moon as an object. When capturing, *Scene Optimizer* uses AI processing to determine which detail enhancement engine to apply. It takes multiple pictures, synthesizes them into a single high-dynamic-range image, and reduces noise. The smartphone applies deep learning algorithms, which introduce external data and thus "enhance" photography beyond what the hardware alone can achieve.[6]

On Samsung's support page, a single sentence quietly acknowledges what has occurred: "Samsung is continuously improving Scene Optimizer to reduce any potential confusion that may occur between taking a picture of the real moon and an image of the moon."[7] This distinction is significant. The *Scene Optimizer* does not capture *the* moon; it generates a synthetic representation based on the *idea* of the moon—it is no longer an index but a semiotic icon—merely resembling it, drawing on training data that meets expectations for what a moon photograph should look like.

When *Scene Optimizer* denatures a moon photograph, it does not amplify the signal captured by the sensor. It displaces the indexical input with statistical composites, generated from an aggregate of all previous training-data moons. The photographer still witnesses indexical input, real light reflecting from the real moon enters the lens, but the output, the image, departs from the original optical presence. The output claims a causal connection to *the* moon through the use of a camera, yet the visual result is of *a* moon. The cognitive dissonance reported by users stems from the fact that Samsung's "optimizer" is less of an enhancer and more of a replacer. This reveals a broader concern in photography, which is, for the most part, about producing images that belong to the person who pressed the shutter. The smartphone user's moon belongs to no one while mimicking personal memory.

This shift is not exclusive to Samsung smartphones; Apple's iPhone XS, released the same year, is advertised to perform one trillion operations per photograph, with its dedicated Neural Engine and Image Signal Processor making computational decisions on the image output[8]. These cameras still "see" in an indexical sense, light still enters the lens, photons still strike the sensor, yet

---

6 Samsung, "How Samsung Galaxy Cameras Combine Super Resolution Technologies with AI Technology to Produce High-Quality Images of the Moon," Samsung Mobile Press, accessed February 6, 2026, https://www.samsungmobilepress.com/articles/how-samsung-galaxy-cameras-combine-super-resolution-technologies-with-ai-technology-to-produce-high-quality-images-of-the-moon

7 *Ibid*

8 Apple Inc., "Apple Special Event," YouTube, September 12, 2018, 1:08:53, accessed April 23, 2026, https://www.youtube.com/watch?v=wFTmQ27S7OQ.

the output represents something else: a new method of optical capture emerging from an accumulation of *elsewhere* and *elsewhen*.

The *elsewhere* and *elsewhen* function as extensions of the photographic program, because the algorithms can only generate images already contained within them. Changes to the physical moon would go unrealized unless that information is updated and reflected in the dataset. After averaging information during ML training, one gradually falls out of sync with reality. The trace of the night scenery is preserved to a certain degree. The surroundings are perceived as authentic, the stars are in their proper places, but the moon's trace is starting to fade. The "this-has-been" of the moon traverses an aggregate of "this-has-been-many-times," and ultimately becomes a "this-could-be."

As this example demonstrates, the selective replacement of information based on prior photographs is quietly becoming the baseline for photographic image production. Our image culture is transforming into an averaged aggregate of the past, moving from optimization towards substitution. A substitution that may take the form of replacing discrete areas within a photograph, or that may operate on the image as a whole, as in ML text-to-image generation tools.

The Prompt Replaces the Shutter

Computer-generated images represent a complete subversion of the camera apparatus. Where computational photography still begins with optical input, text-to-image generation eliminates this requirement entirely. There is no shutter to release, no moment of capture, no scene before the lens. The camera, as a device for recording radiation from the world, becomes vestigial; it gets replaced with an interface for entering text and a model that translates text into synthetic pixels.

To understand this shift, we return to Flusser's analysis of the history of media, where he traced a specific sequence: first, direct experience of the world; then, writing to abstract and encode that experience into linear text; and finally, technical images, which emerged *from* texts. Scientific treatises described optical principles; technical manuals specified chemical processes; texts made photography possible. The photograph is a realization of textual knowledge.

Once photographs existed, they began to replace the textual thinking that produced them. People would encounter a photograph and mistake it for an unmediated window onto reality, such as when experiencing traditional images, forgetting that the layers of encoding embedded within them are already an unavoidable component of the apparatus.[9] In photography, the world is pro-

[9] Vilém Flusser, *Towards a Philosophy of Photography* (London: Reaktion Books, 2000)

cessed through the apparatus that produces a photograph, which then requires text, such as a caption, to anchor its meaning. The text comes *after* the indexical capture, serving to explain, contextualize, or direct the interpretation of what has already been frozen in time.

AI imaging alters this relationship. Now, text comes *before* and determines the generated output. The prompt replaces physical positioning as the primary input. Where the camera's program consists of mechanical and optical constraints, the AI model's program consists of textual-statistical constraints: you can only generate what you can describe, and you can only describe what exists in the language-image associations of the training data.

The text no longer orients the image in the world retrospectively, as a traditional caption does; it now prospectively describes an image to be generated. The prompt is a search command into an existing visual archive, reduced to mathematical vectors. This retrieval mechanism is intentionally unstable; the same prompt produces a different image each time it runs due to an additional seed function[10]. The text does not point to a single position in vector space, but to a statistical area of possibilities. What is retrieved is not an indexical representation, but an averaged manifestation of a probabilistic distribution.

This marks the transcendence of what we call the *corporeal limit*[11] within the program. Photography is defined by embodied presence—a single body at a location, at a specific time, experiencing light through human-scale perception. This physical embodiment is explored by Flusser in *The Gesture of Photographing*[12], where photographs are "caused by phenomena"[13] and created through the search for a place, manipulation of the situation, judgment, and, finally, the release of the shutter[14]. The image-generating prompt floats free of these corporeal constraints. One could generate an image read as a photograph of the moon from a basement, in daylight, without one's body being in any particular place, even if the moon were never visible again. The linguistic act replaces the physical act. The apparatus no longer requires human presence at a scene; it only requires human language directed at a model.

Creating Multitudes Through Machine Learning

In ML image generation, the origin no longer corresponds to a single index. Each photograph in the training dataset originally carried its own indexical specificity: the moon photographed at a

[10] A random number variating (offsetting) the output to simulate organicness
[11] Chefalo and Aubry, "After Technical Images.", *op. cit.*
[12] Vilém Flusser, *Gestures*, trans. Nancy Ann Roth (Minneapolis: University of Minnesota Press, 2014), 72.
[13] *Ibid.* p.73
[14] *Ibid*, p. 73. & 77.

particular time and place, captured through a particular lens under particular atmospheric conditions. However, this specificity does not survive the training process, where images are paired with text descriptions. The system learns statistical relationships between language and visual features, not what something *is*, but what pixel patterns are probable when a word appears in a description.

The mathematical principle that allows data to be compressed while still describing reality accurately is called the manifold hypothesis. It states that high-dimensional data, such as images with millions of pixels, actually lie on a much simpler, lower-dimensional structure. Photographs do not randomly fill all possible pixel combinations; instead, they follow underlying patterns and constraints. The ML training process maps billions of images onto a lower-dimensional manifold, extracting only statistically common features while discarding variation.

Indexicality itself becomes classified as disposable noise. What makes each photograph singular — its specific time, place, and causal chain from world to sensor — is precisely what gets mathematically reduced away. The model treats spatiotemporal coordinates as high-dimensional variation to be compressed out in favor of statistical regularities. What remains is not "this moon photographed here, now" but "moon-ness", a numerical essence distilled from multitudes.

This process significantly differs from the previous photographic apparatus. Both the camera and the ML model are black boxes; however, the camera's black box processes a singular incoming event, while the ML model processes statistical distributions abstracted from prior images. Flusser writes "[The photograph] and what it represents are of the same ontological level: what is represented is the cause, the [photograph] is the effect."[15] The ML model severs this chain. The text-to-image ML model's program does not process incoming light; it only processes words and averages, which are decidedly not specific.

The Post-Indexical Image

The statistically-generated image represents an ontological shift. Photography transforms information at each step but maintains an indexical connection—causal, traceable, and singular. AI image generation asserts itself as post-indexical; it performs "photographicness", but its results do not reference any specific location, time, or light. It is self-referential, mimicking causality while operating through iconic correlation.

Dataset integration causes the index to collapse and introduces the possibility of imaging entirely without it. The apparatus has expanded from indexical capture to probabilistic generation, introducing a temporal shift in which the image cannot be traced back to a specific point in time.

[15] Vilém Flusser, "How should photographs be deciphered?," Flusser Archive, accessed January 2026

The visible is reduced to its "object-ness," an abstraction from the training data, visualizing a synthetic accumulation from outside temporal coordinates.

These images appear photographic, parasitic[16] on an authority they cannot produce. They are visual icons whose meaning derives solely from their appearance, relieved of their trace, and are a-symptomatic. The post-indexical image does not stand (vorstellen) before a world at all; it has no specific world to stand before. The apparatus no longer mediates between reality and image; it generates images from the compressed aggregates of previous mediations, activated by linguistic commands issued from anywhere, at any time.

This essay has traced a progression from indexical capture, bound to the physical world, to generation unbound from world, body, and time. The transitions are neither sudden nor complete, with each stage intensifying what came before. What emerges is *post-photographic* (it references photographic conventions while abandoning photographic conditions), *post-indexical* (it points to no specific external referent, only statistical distributions), *post-temporal* (it exists outside time coordinates, reduced to object-ness rather than this-object-at-this-moment), and *post-corporeal* (it requires no body at any scene, no presence experiencing phenomena).

We believe we are free to generate anything we can describe; however, we are limited by language and how it functions within the program. We think we are creating, but we are navigating possibility spaces already structured by billions of prior images, a lower-dimensional structure that encodes what is statistically probable and excludes what is not. The boundaries of the post-photographic universe are statistical rather than physical, linguistic rather than optical, and far more difficult to perceive or resist.

The shift from 'that-has-been' to 'this-could-be' is not only a change in temporal and spatial reference but an ontological transformation from trace to statistics, from index to icon.[17]

[16] The parasitic image will be developed as a concept in a future essay.

[17] We extend our sincere thanks to the Flusser Archive, run by Prof. Dr. Maren Hartmann and Mirjana Mitrović at UdK Berlin, and to Muyang Lyu for their generous support in granting access and guiding us through the collection. We were thrilled to work with the original materials and manuscripts.

Bibliography

Barthes, Roland. *Camera Lucida: Reflections on Photography*. Translated by Richard Howard. New York: Hill and Wang, 1980.

Bazin, André. "The Ontology of the Photographic Image." In *What Is Cinema?* Vol. 1, translated by Hugh Gray, 9–16. Berkeley: University of California Press, 1967.

Chefalo, Sophia, and Aubry, Vinzenz. "After Technical Images: Towards a Theory of Post-Technical Imaging." *Flusser Studies* 40 (November 2025). https://www.flusserstudies.net/sites/www.flusserstudies.net/files/media/attachments/after-technical-images.pdf.

Flusser, Vilém. "A New Imagination." *Artforum* 26 (April 1988): 14–15.

Flusser, Vilém. "How Should Photographs Be Deciphered?" *Philosophy of Photography* 2, no. 2 (2012): 210–14. https://doi.org/10.1386/pop.2.2.210_7.

Flusser, Vilém. "Letter to the Editor." *Leonardo* 19, no. 1 (1986).

Flusser, Vilém. "On Typography." *Printletter* (Zurich), https://www.flusserstudies.net/sites/www.flusserstudies.net/files/media/attachments/flusser-on-typography.pdf, n.d.

Flusser, Vilém. "Three Times." *Artforum* (February 1987).

Flusser, Vilém. *Towards a Philosophy of Photography*. Translated by Anthony Mathews. Göttingen: European Photography, 1984. [Originally published in German as *Für eine Philosophie der Fotografie*, 1983.]

Jäger, Gottfried, Herbert W. Franke, and Jean Stken. "Generative Photography: A Systematic, Constructive Approach." *Leonardo* 19, no. 1 (1986): 19–25.

Krauss, Rosalind. "Notes on the Index: Seventies Art in America." *October* 3 (Spring 1977): 68–81.

Krauss, Rosalind. "Notes on the Index: Seventies Art in America. Part 2." *October* 4 (Fall 1977): 58–67.

Samsung Mobile Press. "How Samsung Galaxy Cameras Combine Super Resolution Technologies With AI Technology to Produce High-Quality Images of the Moon." March 15, 2023. https://www.samsungmobilepress.com/feature-stories/how-samsung-galaxy-cameras-combine-super-resolution-technologies-with-ai-technology-to-produce-high-quality-images-of-the-moon.