NOAM M. ELCOTT

Artificial intelligence: The phrase's discursive force is equaled by its confounding imprecision. Too often, critics oppose "artificial" (*ars* + *facere*) to "human" even as we recognize that artistry or art-making (*ars* + *facere*) is a defining feature of humanity. Other critics aim to sanctify "intelligence" as uniquely human at precisely the moment scientists and philosophers recognize a heterogeneity of intelligences across a range of organisms, technologies, and hybrid systems. What is more, the definition of AI is forever in flux, with successful applications invariably dismissed as no longer "intelligent" (the so-called AI effect as evinced, for example, in optical-character recognition (OCR) technology, which ranks among the great achievements of machine learning yet has long relinquished any claim to the "AI" moniker). AI is too broad and fraught a terrain for succinct analysis. I will focus, instead, on an integral component of most advanced deep-learning neural networks, namely, latent space: the highly abstract, multi-dimensional space constructed by deep neural networks to facilitate tasks such as pattern recognition, feature abstraction, and synthetic data generation. In the paragraphs that follow, I hope to explicate the technical emergence of latent space, situate it as a grid-like cultural technique, and relate it to a prerequisite epochal media transformation: the rise of mass photography. No serious discussion of AI is possible without some grasp of latent space—its constructions and potentialities, opacities and limitations.

Foundation models such as OpenAI's GPT-4 and Google's Gemini are deep neural networks pre-trained on massive, unlabeled datasets comprising text, images, audio, video, and computer code. The "raw" digital data is "high-dimensional," like words in a text or pixels in an image, vaguely analogous to uncompressed image formats like TIFF, as distinct from compressed formats like JPEG. By and large, data cannot be processed efficiently or effectively in this high-dimensional form (just as it is difficult to work with a TIFF file in many contexts). Deep neural networks comprising thousands of layers, each composed of thousands of neurons, yielding billions or trillions of parameters, process the "raw" data in order to extract and transform features from the data. Early layers of the network tend to extract low-level features (like edges in images or phonemes in speech). As data progresses through the layers, the features become more abstract (such as complex shapes or semantic concepts); eventually the data is abstracted into forms that are unrecognizable to human senses or minds. This process also involves reducing the dimensionality of the data, distilling the most relevant information and discarding redundant or irrelevant features. By the time data reaches the deeper layers of the network, it has been transformed into a highly abstract, compact form and is embedded in a multi-dimensional, topological space that coordinates its meanings, relationships, and contexts. This is the latent space. Here, the data is represented by a set of latent variables—"latent" because they capture underlying patterns or features not immediately apparent in the high-dimensional data—in a "lower" dimensional space that nonetheless comprises hundreds of thousands of dimensions.

Latent space is an abstract, lower-dimensional representation of the original data, its relations and patterns. In this multi-dimensional manifold, proximity is generally a proxy for similarity: Similar concepts are embedded closely together and disparate concepts are located further apart. But the configuration of latent space is largely non-intuitive (dimensions often do not correspond to anthropocentric and sensorially perceptible concepts such as color or shape) and defies direct representation. In generative models, latent space can be sampled to generate new data instances, such as images or prose. In other types of networks, it might be used for classification, prediction, or other tasks.

Understood as a cultural technique, latent space is related to the grid. Building on Rosalind Krauss's foundational analysis,[1] media theorist Bernhard Siegert distinguishes between representational grids (such as Alberti's system for linear perspective), topographic grids (such as the checkerboard urban planning behind Latin American colonial settlements), speculative grids (such as the U.S. Land Ordinance of 1785, which undergirded expansion into the Northwest Territories), and three-dimensional architectural grids (such as the one promulgated by Bauhaus architect and Walter Gropius disciple Ernst Neufert in his hugely influential *Bauordnungslehre* [1943], published in English as *Architects' Data*), among others. In each instance, the grid was a crucial technology or cultural technique—an essential procedure and even agent—for the conquest of ever more spaces. Presciently, Siegert asked: "Can the expansion of Western culture from the sixteenth to the twentieth century be described in terms of a growing totalitarianism of the grid?"[2]

For Siegert, the grid has a triple function: (1) It is "an imaging technology that by means of a given algorithm enables us to project a three-dimensional world onto a two-dimensional plane"; (2) it is a diagrammatic procedure that can be implemented in the real and the symbolic; (3) it helps "constitute a world of objects imagined by a subject," serving as the equivalent to Heidegger's *Gestell* or enframing. The grid, in short, "is a medium that operationalizes deixis. It allows us to link deictic procedures with chains of symbolic operations that have effects in the real."[3] Latent space can be understood as a hypertrophic grid; but compared to earlier grids, differences of degree become differences in kind. For all their complexity, a number of generative image models—including generative adversarial networks (GANs), variational autoencoders (VAEs), and emergent multi-modal latent-diffusion models (LDMs)—are operationalized deixis: The models point to coordinates in the latent space, where hundreds or thousands of dimensions intersect, specifying the content, style, and other visual attributes indicated in the nat-

1. Rosalind Krauss, "Grids," *October* 9 (Summer 1979), pp. 50–64.

2. Bernhard Siegert, *Cultural Techniques: Grids, Filters, Doors, and Other Articulations of the Real* (New York: Fordham University Press, 2015), p. 98.

3. Ibid.

ural-language prompt.[4] Nonetheless, the gap between latent space and the representational, cartographic, topographic, and related grids analyzed by Siegert is substantive: For example, unlike a regular grid with uniformly spaced intervals, the distances in latent space are not uniform and can represent complex, non-linear relationships between different features.

Siegert's triple function can be rewritten as follows: First, in the case of images, latent space enables the reduction from hundreds of thousands or more dimensions to tens or hundreds rather than from, say, three to two dimensions, as in linear perspective. Unlike linear perspective, moreover, the lower-dimensional representation of data does not "submit the representation of objects to a theory of subjective vision";[5] quite the contrary, latent representations are generally not representable, let alone representational, in human or art-historical terms. (Try to picture specific coordinates in a space with over one thousand dimensions.) Second, like other grids, latent space is a diagrammatic procedure that can be implemented in the symbolic and in the real. For example, latent space is crucial to many real-world surveillance technologies, facial recognition in particular. What is more, in text-to-image models (especially those that rely on Contrastive Language-Image Pre-training [CLIP]), latent space facilitates the apparently effortless movement between the symbolic and the imaginary, a phenomenon made preposterously obvious in photo-realistic images of cats riding bicycles on the moon and painfully evident in abusive deepfakes.[6] Finally, latent space constitutes a world of objects imagined not by a subject but by digital multitudes: collective efforts like Wikipedia, oeuvres of individual artists, the recycled refuse of innumerable operational images, and the digital detritus of all kinds—all scraped from the Internet and used to train deep neural networks to recognize and generate even more images. It is not a pretty world.[7] We will return to this disintegrated world and absent subject below.

Every image produced by a GAN or an LDM is an image *from* latent space. But they are hardly images *of* latent space. Computer scientists have developed a number of techniques to construct visualizations of latent spaces, however partial

4. On the relation between the index (photography) and indexing (deep neural networks) as well as that between images and words, see Antonio Somaini, "Algorithmic Images: Artificial Intelligence and Visual Culture," *Grey Room* 93 (Fall 2023), pp. 77–78 and passim. For GANs to generate images based on natural-language prompts, they must be conditioned on text or paired with text-processing mechanisms.

5. Siegert, p. 98.

6. A more fastidious Lacanian analysis of deep neural networks would locate the symbolic in the structured, rule-based aspect of the network, including latent space; the imaginary in the creation and interpretation of "images" (data representations) that are reflections of the real world but are themselves constructs or simulations within the network; and the real in the chaotic, unstructured, and raw aspect of data and reality that can never be fully captured or represented by the network.

7. See esp. Kate Crawford and Trevor Paglen, "Excavating AI: The Politics of Images in Machine Learning Training Sets" (2019), https://excavating.ai; Emily Denton et al., "On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet," *Big Data & Society* 8, no. 2 (July 2021).

and distorted, including t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP). Indeed, among the virtues of latent space is its capacity to allow researchers to peek—imperfectly and incompletely—into the black box of deep neural networks.[8] But the prospects for comprehensive representations of latent space are remote.[9] And even computer scientists rely on intuition when working with neural-network architectures and their resultant latent spaces. We cannot encounter latent spaces directly.

For an aesthetic encounter with the alterity of the computer vision enabled by deep neural networks, artworks are among our greatest resources. Exemplary are Trevor Paglen's *Adversarially Evolved Hallucinations* (2017), where he forced GANs to produce images that are intelligible to the network but nearly inscrutable to humans. Works like *Angel (Corpus: Spheres of Heaven)*, *Rainbow (Corpus: Omens and Portents)*, *Porn (Corpus: The Humans)*, *A Prison Without Guards (Corpus: Eye Machine)*, and *The Great Hall (Corpus: The Interpretation of Dreams)* provide glimpses—purposely and necessarily imperfect and incomplete—into the neural networks' hidden layers, that is, into their wittingly contorted latent spaces. Paglen's invocation of Freud is apt: "'*Unheimlich*' is the name for everything that ought to have remained . . . hidden and secret and has become visible."[10] Computer vision—like artificial intelligence—is uncanny precisely in its collapse of freakishness and familiarity. Or at least it was. Many products of generative AI—whether text, image, or audio—have or will soon have traversed the uncanny valley and appear indistinguishable from human-generated texts, images, or audio. Paglen's

8. Many scholars believe that the internal operations of deep neural networks, especially in latent space, are necessarily opaque or "black boxed." Nonetheless, the black boxing of AI must be understood not only technologically but also historically and politically. As Matthew Jones has argued, "Opacity needs its history." Matthew L. Jones, "Decision Trees, Random Forests, and the Genealogy of the Black Box," in *Algorithmic Modernity: Mechanizing Thought and Action, 1500–2000*, ed. Morgan G. Ames and Massimo Mazzotti (New York: Oxford University Press, 2022), p. 192. On AI opacity understood in terms of intentional corporate or state secrecy, technical illiteracy, and technological necessity, see Jenna Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society* 3, no. 1 (2016), pp. 1–12.

9. The research arm of the AI company Anthropic recently announced a breakthrough in mapping its model's inner workings, specifically its "concept space," which can be understood as a subset of latent space where interpretable features (or concepts) are identified and analyzed. But they cannot reconstruct why or how the model constructs a certain sentence or image any more than—to use a dangerously anthropomorphic simile, which should not be taken literally—an MRI can reveal why or how a human arrived at a specific formulation. And as the authors readily concede, the prospects for comprehensive representations of the concept space are remote: "We do not believe we have found anywhere near 'all the features [i.e., concepts]' that exist in [Claude 3] Sonnet [a large-language model similar to OpenAI's GPT series], even if we restrict ourselves to the middle layer we focused on. We don't have an estimate of how many features there are or how we'd know we got all of them (if that's even the right frame!). We think it's quite likely that we're orders of magnitude short, and that if we wanted to get all the features—in all layers!—we would need to use much more compute than the total compute needed to train the underlying models." See Adly Templeton, Tom Conerly, et al., "Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet," *Anthropic* (2024), https://transformer-circuits.pub/2024/scaling-monosemanticity/index.html.

10. Schelling, quoted and developed in Sigmund Freud, "The Uncanny," in *The Uncanny* (New York: Penguin Books, 2003), p. 132.

Adversarially Evolved Hallucinations register not only striking incunabula of deep neural networks but also and more importantly our capacity for productive alienation therefrom.

Nearly a century ago, Siegfried Kracauer registered a similarly productive alienation facilitated by media technologies:

> For the first time in history, photography brings to light the entire natural cocoon; for the first time, the inert world presents itself in its independence from human beings. Photography shows cities in aerial shots, brings crockets and figures down from the Gothic cathedrals. All spatial configurations are incorporated into the central archive in unusual combinations which distance them from human proximity. . . . This is how the elements crumble, since they are not held together. The photographic archive assembles in effigy the last elements of a nature alienated from meaning.[11]

Rather than lament a disjointed nature, Kracauer recognized an opportunity: "The images of the stock of nature disintegrated into its elements are offered up to consciousness to deal with as it pleases. Their original order is lost; they no longer cling to the spatial context that linked them with an original."[12] Already in 1927, Kracauer understood that this "consciousness" was best actualized through cinema, or what his contemporary Jean Epstein called "the intelligence of a machine."[13] "If the disarray of the illustrated newspapers is simply confusion, the game that film plays with the pieces of disjointed nature is reminiscent of *dreams* in which the fragments of daily life become jumbled. This game shows that the valid organization of things [in the general inventory] remains unknown."[14] Kracauer's inert world is our artificial intelligence; his alienated and disjointed nature, our alienated and disjointed computer vision; his centrally archived, unusual, and nonhuman spatial configurations are our latent spaces; and, for better and for worse, his cinematic dreams are our deep neural-network hallucinations. What is crucial is that we remember that the valid organization of things remains unknown.

NOAM M. ELCOTT is an associate professor of art history and archaeology at Columbia University.

11. Siegfried Kracauer, "Photography" [1927], in *The Mass Ornament*, ed. and trans. Thomas Y. Levin (Cambridge, MA: Harvard University Press, 1995), p. 62. On the archives of stock photographs and neural networks, see Roland Meyer, "The New Value of the Archive: AI Image Generation and the Visual Economy of 'Style,'" *Image* 37, no. 1 (2023), pp. 100–111.

12. Kracauer, "Photography," p. 62.

13. See Jean Epstein, *The Intelligence of a Machine*, trans. Christophe Wall-Romana (Minneapolis, MN: Univocal, 2014); Noam M. Elcott, "The Master of Time: Jean Epstein's Nonhuman Time Axis Manipulation," in *Time Machine: Cinematic Temporalities*, ed. Antonio Somaini and Marie Rebecchi (Milan: Skira, 2020), pp. 163–83.

14. Kracauer, "Photography," p. 63.