(Deep) Representation: a Layered and Interconnected Mediation

Lorenzo Aimo
lorenzo.aimo@unimi.it
University of Milan and
University of Bologna, Milan, Italy

Ludovica Schaerf
ludovica.schaerf@uzh.ch
Digital Visual Studies, UZH-MPG,
Zurich, Switzerland

Sebastian Rozenberg
sebastian.rozenberg@liu.se
Linköping University, Linköping, Sweden

Ellen Charlesworth
ellen.charlesworth@uni.lu
University of Luxembourg, Luxembourg

Representation, in its most rudimentary sense of 'standing for' something else, operates as a fundamental yet profoundly complex mediation within contemporary AI models. These systems do not simply process information but re-elaborate it through a sequence of transformations, producing what we call deep representations: encodings that are simultaneously layered, interconnected, and polysemic. Drawing on Stable Diffusion as our case study, this work traces the representational journey from world to model, examining how each layer of mediation reshapes what 'standing for' means. We begin with the mediations that precede the model: data as *capta* rather than given, digital image as cultural object and computational substrate, dataset as epistemic border, and pre-processing as structural constraint. We then turn to transformations within the model itself: the training, loss functions, and material economies that shape learned representations, differing fundamentally from standardized compression. We conclude by questioning whether the inherited vocabulary of representation remains adequate to describe what these systems actually do.

1 Introduction

In 1965, conceptual artist Joseph Kosuth exhibited his artwork *One and Three Chairs at MoMA* (in Figure 1). Influenced by the new theories of language and signification, his work draws on Minimalism but engages with themes the movement had sidelined: how ideas, signification, and meaning are constructed. The central idea of the artwork is to explore the nature of representation itself: the concept of a chair is, in fact, represented through three different utterances: as an actual chair, a photographic image of the actual chair, and a dictionary entry of the word "chair". The viewer is then confronted with "three" chairs, each represented and experienced through its distinct properties.

Sixty years later, given the rapid advent and spread of generative AI, we can update Kosuth's artwork with a fourth form of representation, on the right in Figure 1, one that not only is becoming technically and culturally crucial but that combines, merges, and transforms the two representational modalities employed in Kosuth's artwork, the image and the text. Alongside the actual chair, the related dictionary entry and photographic image, inspired by Kevin Esherick's *Hyperportrait*, we insert the embedding of the chair, a deep learning representation that corresponds to the concept within, in this case, the textual branch of the model CLIP. This representation, as we explore in this article, is shaped simultaneously by the collection of numerical representations of the training data, by their digital transformations, by the models them-

Keywords: Representation, Digital Image, Data, Latent Diffusion, Mediation.

DOI 10.34626/2026_xcoax_016

selves, and by their objectives. All these layers ultimately contribute to the so-called latent vector representation.

Fig. 1. Updated version of Joseph Kostuh's *One and Three Chairs* with the visualization of the embedding of "chair" (on the right).

We propose to call this fourth chair a *deep representation*: a term that captures both its technical lineage (deep learning) and its defining characteristics. Deep representations are simultaneously layered (undergoing multiple cascading transformations), interconnected (each element defined relationally rather than intrinsically), and polysemic (carrying functional, technical, and relational meaning concurrently). Unlike Kosuth's three chairs, each occupying a distinct ontological register (material object, visual image, linguistic definition), the deep representation collapses and reconfigures these registers. It is at once image and text, material and abstract, individual instance and statistical aggregate.

Within the contemporary context of machine learning and deep neural networks, representation is a key term with a specific technical denotation, resulting from the layered history of representing information within digital computational technologies. The conception, definition, and implementation of the term derive from and belong to the specialist language of computer science, where it refers to the way information is presented and described within digital computational systems. As Paul Dourish argues, "there is no information without representation": it is the outcome of an assemblage of "hardware, software, data representations, diagrams, algebras, business needs, spatial practices, principles, computer languages, and related elements, linked by conventions of use and rules" underpinning an historically situated practice of information processing (2017, 27-28).

The assemblage the term underpins asks for a more nuanced investigation of what its meaning implies in the contemporary state of AI models. The concept of representation in deep learning models is deeply polysemic. While this is not the only interpretation of the word that is crucial to neural networks, in this work, we present a discussion of representation as re-presenting something else. We interpret representations in neural networks as re-elaborations of the input or output data into an encoding that 'stands for' the data in complex ways, as a media-

tion—hence our adoption of the term deep representations to mark this specific mode of standing-for.[1]

In what follows, we trace the representational journey from world to model output, examining how each layer of mediation transforms what 'standing for' means. To do so, we take into consideration the specific case of latent diffusion models as proposed by Rombach et al. (2021), implemented in commercially available text-to-image software, such as Stable Diffusion, released by some of the authors after joining Stability AI. Stable Diffusion was chosen as it is the most popular open-source image generation model, which helps us ground the discussion into something that is at the same time widely known and currently being used in practical applications.

Section 2 examines the layers of mediation that precede the model: from data as capta, through the digital image as computational object, to the dataset as epistemic filter, and finally, preprocessing as architectural constraint. Section 3 turns to representations within the model itself, showing how the training process and architectural choices produce a fundamentally different mode of compression than standardized formats like JPEG. We conclude by questioning whether the inherited vocabulary of representation—standing for, mediating, encoding—adequately describes what deep representations do, or whether they constitute a productive mode of presentation that has moved beyond referential logic.

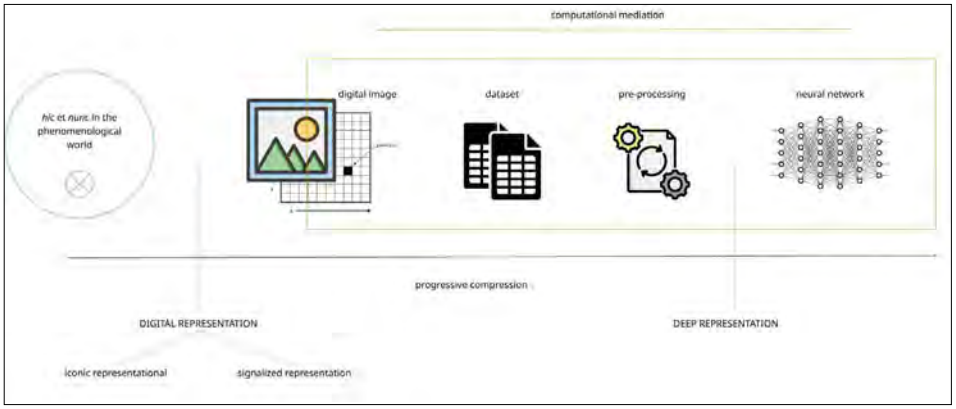

Fig. 2. Diagram visualizing the path from the world (on the left) to the model (on the right) and the layers of mediation involved.[2]

2 Representation from the World to the Model

Between the world and the neural network's internal representations lie multiple layers of mediation, each transforming how information is encoded and what is represented. We trace this path through four transformations: first, the constitution of data itself as structured observation rather than raw given; second, the digital image as a dual entity—both

1. In the remainder of the paper, we refer to the representations that are internal to deep learning models as 'deep representations', and not the more common 'neural representations', as the relation to neuroscience is beyond the scope of this work. For a nuanced discussion on this perspective see: Dhaliwal et al. 2024.
2. This diagram displays what appears as a simple and linear process; it is an attempt of making diagrammatically visible what is actually a complex and intricate phenomenon that involves an entangled and networked infrastructure of humans, devices, interfaces, submarine cables, data centers, energy resources, networks, software, hardware, algorithms - to name a few.

cultural object and numerical array; third, the dataset as an epistemic infrastructure that determines what knowledge enters the model; and fourth, preprocessing as the final architectural constraint that reshapes data into forms a specific model can process. Each layer is not merely preparatory but constitutive: it actively shapes what the model can learn to represent. We visualize a simplified version of this path in the diagram in Figure 2.

2.1 Data

The broad and general term "data" asks for a more nuanced and critical reflection when thinking about representation. Data are commonly defined as "a collection of facts, numbers, words, observations or other useful information" (IBM 2024) that constitute the resource for the processes of analysis behind decision-making. Data can be of different types and part of different operations. In this context, particularly, we refer to digital information, a representational modality founded on digits, which form bits, which form bytes.

However, as it has been argued against the mystification the term itself contains, of being "something given" (Drucker 2020; Gitelman 2013), data are neither "raw" nor "natural resources", but result from processes of generation, capture, gathering and curating all of each shaped by the epistemology structuring the "parameters of the search" (Drucker 2020, 49). They should rather be named *capta* as the term more clearly highlights how the model (both as technical and conceptual) shapes the information it necessitates according to its own parameters and measuring criteria.

This general critique of data as *capta* becomes concrete when we examine the specific form data takes in image generation models. Stable Diffusion consists of a layered computational architecture and, like digital computing technologies, it processes digital information. Stable Diffusion's training data consists of images with their corresponding textual captions, which means digital representations of pictorial and visual information and textual information. This representation constitutes what we commonly refer to as the digital image, a specific form of data that exists simultaneously as visual representation and computational substrate.

2.2 The Digital Image

Digital images are instantiated as specific raster data structures, also called *bitmaps*, characterised by pixel space. Pixel space is the modality through which historically the energy of photons forming light has been encoded within digital computational systems (Terras 2016): the pixel space results from processes of signalization (detecting energy, converting it into discrete, measurable electrical charges) and digitalization (converting the electrical charges into digital numbers described by bits, sequences of values between 0 and 1, the "language" of computers). In this way, pixels constitute digitized samples of a visual scene at a given

point, and those samples are bit sequences that a computational system can elaborate, process, and store in its RAM.

Pixel space is not only an informational structure and logical convention, but also the site of manifestation of the digital image, having the material skin of the screen. Composed of physical pixels emitting light, display screens function both as an interface and the terminal of what Berry calls *computational mediation* (2011) to refer to the double mediation enabled by software, which happens whenever computational technologies represent: first, the translation into bits, second, the translation of that representation into a human-readable display according to a specific format.

The digital image both represents and is represented. At a human scale, we experience and interpret it as a visual, continuous, and meaningful representational surface standing for what it depicts. As a cultural, social, and semiotic object, the digital image mediates the world, embodying a *way of seeing* (Berger 1972). At a computational scale, the image is represented as information through an invisual data structure,[3] existing as a numerical representation related to electrical voltage. The digital image is then the site where not only ways of seeing are embedded, but where those precipitate and get transcribed into arrays of numerical values open to processability.

Being specific about digital images (Rose 2023) means highlighting the materiality, configuration, and properties of a kind of representation as encoded, operated, and processed by digital computational technologies, software, and hardware. In their contemporary form, digital images need to be understood as objects "materially structured by the *logic of computation*" (Gaboury 2015, 45) and, for this reason, as the site where different conceptions and modalities of representation both collide and coalesce in weaving world and numbers together. Individual digital images, however, do not enter models in isolation. They are aggregated, ordered, and filtered through another representational structure: the dataset, which functions as the epistemic border determining what knowledge about the world the model inherits.

2.3 The Dataset

LAION 5B collects and gathers more than five billion digital images scraped from the web. The ensemble of these images with their ways of seeing and their bitmaps translations constitutes the *epistemic border* (Pasquinelli and Joler 2021) characterizing Stable Diffusion: what knowledge about the world it has inherited. In the context of machine learning, a dataset is an object of knowledge that takes the form of a table and hence works according to its representational function (Mackenzie 2017). It embodies a way of seeing as well, defined by the criteria behind the selection, collection, and ordering of its data. The ordering is

3. We draw this term from the conceptualization developed by Mackenzie and Munster (2019) and later taken up by Parikka (2023), in which "invisual" refers to the dimension to which images belong within digital computational technologies and platforms – a dimension that exceeds and outscales optical and visual paradigms of observation.

realized through the interplay of the verbal and the visual, which takes and varies shapes according to the purpose the dataset has been conceived for: it could be the assigning of categories' labels, as in the case of ImageNet or the scraping of images with their captions from the Web, as in the case of LAION5B.

In this way, a concept expressed through words is guaranteed a visual correspondence and, since the ruling criteria of this object of knowledge respond to statistical principles and logics of models, representation is twisted into *representativity* (Malevé and Sluis 2023). Qualitatively, images and text function as proxies for concepts;[4] quantitatively, more and more images, more and more bitmaps for the same label or caption help the model to learn a more accurate representation of that concept.

Materially speaking, in gathering huge quantities of data, the dataset has a specific *digital weight,* which requires a proportional computing power to process it. LAION5B counts 220 terabytes for images and approximately 2,65 terabytes for metadata. Neural networks in general and Stable Diffusion in particular are able to deal with this amount of information thanks to their representational logic and procedures: they, in fact, operate on and with different representations of the data they are fed. Yet even after images are gathered into datasets with their textual annotations, one final transformation occurs before they enter the model: preprocessing, which reshapes the digital image according to the structural expectations of specific neural network architectures. Here, the same image becomes a fundamentally different representational object depending on whether it enters a convolutional or transformer architecture.

2.4 Preprocessing

Digital images are dealt with by neural networks as *tensors*, a data structure consisting of spatial matrices of numbers (the grid) with an additional dimension for the color encoding (three dimensions for RGB) and, often, a batch size (inputs are processed in parallel in groups). Instead of thinking of images as two-dimensional areas, in neural networks, they are treated as four-dimensional volumes. Most versions of Stable Diffusion (1-2) process tensors of around 512x512x3 pixels with values ranging from -1 to 1.

Tensors, in order to be treated adequately by the networks, are pre-processed to achieve standardization and learn invariances, i.e., the aspects that the models should not rely on to make predictions. The preprocessing pipeline, however, diverges significantly depending on the architecture receiving the image. Convolutional neural networks and transformer-based models impose *different structural expectations* on their inputs, reflecting their fundamentally different approaches to spatial understanding. Convolutional architectures process images as continuous spatial fields, relying on explicit augmentation techniques,

4. Many investigations and analyses have shown the cultural and situated dimensions, which end up being discriminatory, of this apparently natural correspondence, see Prabhu and Birhane 2020, Crawford and Paglen 2021, Denton et al. 2021.

i.e., flipping, rotation, shifts, synthetic variations, and adjustments to hue, saturation, and intensity, to build invariance to transformations. The convolutional approach maintains locality.

Transformer architectures, by contrast, require explicit patching of the image into discrete segments that are processed separately. While convolutions handle spatial relationships implicitly, the patching process also needs to include position encodings, i.e., representations of where the patch is within the image. Different from convolutions, used by Rombach et al. (2021), in transformers, the flattening of patch embeddings fundamentally restructures how the image is represented. Where convolutional models emphasize *locality*, transformers prioritize *globality*, *relationality*, and *contextual understanding*. These architectural differences mean that the same digital image, depending on whether it enters a convolutional or transformer pipeline, becomes a fundamentally different representational object.

3 Representation Within the Deep Learning Model

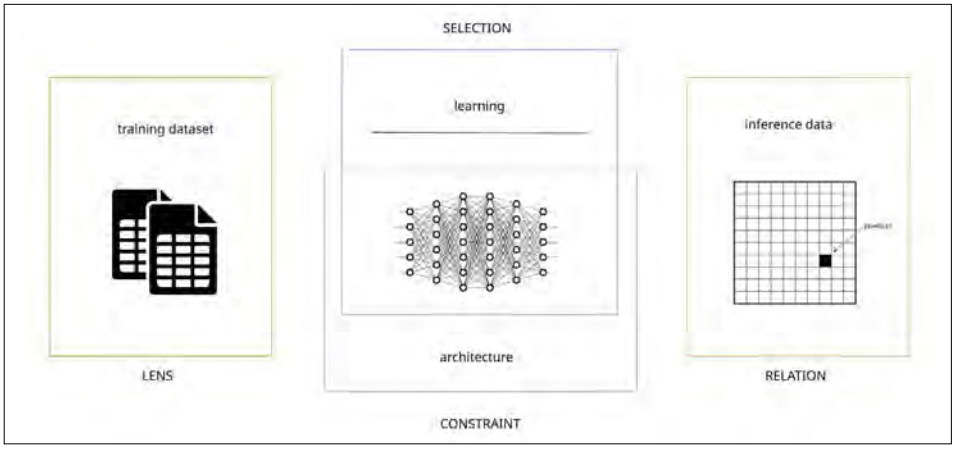

Fig. 3. Diagram schematizing the four layers influencing representation compression within the model. On the left, the training data acts as a lens to the model. In the center, the architecture and learning provide, respectively, a structure and functionality. On the right, the inference data is what determines the meaning: the relationships between representations.

If Section 2 traced the layers of mediation between world and training data, we now examine representations within the model: the transformations that constitute learning itself. Upon entering the network, the preprocessed tensor undergoes a sequence of non-linear transformations. Each layer produces what we might call *intermediate representations*: activation patterns that progressively abstract the input from pixel values into increasingly complex statistical descriptions of visual features. These transformations are not explicitly programmed but emerge from training, where the network learns which patterns to emphasize and which to suppress based on the dataset and the error function.

Stable Diffusion's architecture combines two neural networks: a Variational Autoencoder (VAE) (Kingma and Welling 2015) and a U-Net (Ronneberger et al. 2015). The VAE compresses image representations into a statistical representation, then decompresses them into the original space. What distinguishes latent diffusion from traditional diffusion models (DDPM) (Ho et al. 2020) is precisely this compression. Rather than operating on full pixel space, the entire diffusion process, the progressive denoising that generates the final image, happens within the

VAE's compressed latent space. The U-Net learns to predict and remove noise not from a 512x512x3 tensor but from its compressed equivalent of 64x64x4. This spatial reduction enables Stable Diffusion to generate high-resolution images *efficiently* while maintaining what its developers call "state-of-the-art synthesis results" (Rombach et al. 2021): images that plausibly represent their textual prompts. The Stable Diffusion paper explicitly identifies this compressed space as achieving "a near-optimal point between complexity reduction and detail preservation, greatly boosting visual fidelity" (Rombach et al. 2021). The term "representation" here refers specifically to this compressed form, a mediation shaped by the balance between what can be discarded and what must be preserved.

The technical details matter because they reveal a specific *economic logic*: the denoising in the compressed space denotes both a conceptual framework and a technical procedure that responds to, and is shaped by, economic criteria, namely, the management of material, energetic, informational, and temporal resources. Its main logic aims at maintaining the quality of the generation while "limit[ing] computational resources" (Rombach et al. 2021). It is in this respect that representation stands for both a specific instance and a typology of data representation shaped by specific purposes (as in the case of Stable Diffusion) and for a general historical problem of information representation within digital computational systems.

Representation is a profoundly material and economic term: it accounts for the historically situated practices of information processing and the materialities of information, "those properties of representations and formats that constrain, enable, limit, and shape the ways in which those representations can be created, transmitted, stored, manipulated and put to use" and how their arrangements "matter significantly for our experience of information and information systems" (Dourish 2017, 4-6). Bits, the smallest components digital information is made of, are both "*logical and material entities*" (Blanchette 2011) and these two dimensions exist within a dialectical tension: in this sense, on the one hand, Stable Diffusion and the representational techniques it involves can be understood both as the result of the material development and evolution of CPUs and GPUS and as the result of the correlated conceptual developments within computer science (neural networks, machine learning).

Concurrently, *compression* has accompanied representation throughout the history of digital computing; it is 'the process that renders a mode of representation adequate to its infrastructures' (Sterne 2015, 35). But not all compression operates according to the same logic. The difference between JPEG compression and VAE compression illuminates what makes deep representations distinctive. JPEG compression operates through a standardized algorithmic procedure applied uniformly to any image, discarding information that is imperceptible to the human eye. It compresses the channel in Shannonian terms:[5] a

5. For a detailed discussion of this point see: Impett and Offert, 2026. taken up by Parikka (2023).

task-agnostic, isolated transformation. It is a compressed form of data that allows for transmission, circulation (a format that can be read by hardware and software, that can exist on the web), and storage.

Within the neural network of a VAE, JPEG files of grid pixels (already in their way compressed) are processed through the layers, but the way those get encoded follows another kind of compression. The VAE, in fact, compresses representation through four interrelated forces rather than universal perceptual principles: First, the training dataset functions similarly to what Pasquinelli and Joler (2021) term a 'Nooscope', a *lens* that orients the model's representations according to the statistical regularities, biases, and cultural patterns embedded within it. The VAE learns to compress based on what recurs, what varies, and what correlates in its training data. Second, the architectural choices *constrain* what kinds of representations are possible, privileging certain forms of compression over others through the specific structure of layers, dimensions, and operations. Third, the loss function and learning procedure define what constitutes error, *directing* what the model optimizes for and therefore what features are preserved versus discarded in the compressed representation. Fourth, and most fundamentally, the compressed representation is *relational* rather than absolute. Each image is encoded not according to its intrinsic properties but in relation to all other images the model encounters. A representation is meaningful only through its position relative to other representations: its distances, similarities, and structural relations within the latent space. The four layers are schematized in Figure 3.

These four dimensions make the VAE's compression a fundamentally different kind of mediation than JPEG. Where JPEG applies a standardized transformation, the VAE produces representations that are layered (through sequential transformations), interconnected (defined relationally), and polysemic (simultaneously functional, technical, and relational). This is what we mean by calling them deep.

Return to Kosuth's fourth chair: the latent vector visualized in Figure 1. This representation differs from the photograph not merely in being numerical rather than optical, but in how it came to be. The photograph captures light reflected from a material chair at a specific moment. The latent vector, by contrast, is shaped by its statistical relations to millions of other images the model has encountered: it represents 'chair' not through direct capture but through learned patterns of what correlates, recurs, and varies across LAION5B. It differs from the dictionary definition not in offering a verbal rather than visual description, but in substituting relational position for content. The embedding means 'chair' not through definition but through its distances from 'table,' 'stool,' 'furniture,' and countless other positions in the latent space. The fourth chair thus reveals that representation is always already multiple, and deep learning makes this multiplicity *operational*.

4 Discussion

The deep representations traced in this paper encompass three concurrent definitions. *Functionally*, representation is optimisation: transforming data into a summary that omits unnecessary details and preserves useful ones. *Technically*, it is encoding: computational procedures that transform data from one form to another, from tensor to vector, modifying the information load. *Relationally*, it is re-presentation: the mediated output bears a structured, if complex, relation to all other elements. These definitions do not contradict each other but reveal a constitutive tension: the first two treat representation as *substrate* – material-informational infrastructure shaped by economic constraints – while the third treats it as *form*, i.e., a mediation that stands in some relation to what it re-presents. Deep representations are irreducibly both, and this is what makes them difficult to locate within existing frameworks.

This tension between substrate and form suggests that our inherited vocabulary of representation may be inadequate to what deep learning systems actually do. Agre (1997) already identified representation as AI's central unexamined assumption: classical AI treats it as internal symbols mirroring an external world, inheriting a Cartesian split that makes the relation between representation and world permanently mysterious. His alternative of deictic, situated representations that track affordances rather than model worlds, were designed for agents acting in environments. But deep representations fit neither category. A latent vector is not a symbolic model *of* the world, nor is it an indexical-functional coupling *with* an environment. It is a generative topology: a structured space of possible outputs shaped by statistical regularities and material constraints, whose relation to what it re-presents is economic and relational before it is semantic.

Continental philosophy has long distinguished between two fundamentally different operations that the English word "representation" collapses together: representation as *reference* – the subject placing the world before itself as object,[6] and representation as *production* – the making-actual of sensible form.[7, 8] Deep representations do not point back to a given world; they produce a structured space of possible sensible outputs through a sequence of transformations. They are generative rather than mimetic, constitutive rather than adequate. What deep learning calls "representation" may be neither the transparent sign mapping its referent nor the modern subject-grounded mediation that replaced it, but an *operational mode of presentation* – production of the sensible organised by statistical regularity and material constraint rather than by reference or consciousness.

6. What Heidegger identified as the defining gesture of modernity.
7. What Benjamin called the essence of philosophical method.
8. This distinction finds one origin in Kant's separation of *Vorstellung* (mental representation presupposing a subject-object structure) from *Darstellung* (sensible presentation through mediating operations) (Kant 1998).

For our technical understanding, the implication is that the inherited vocabulary of 'standing for' systematically misdescribes what deep representations do. They do not abbreviate a world; they produce a representation that is defined not by what it refers to but by layers of mediation. Recognising deep representations as productive presentation rather than referential representation is not merely a philosophical refinement. It shifts the questions we ask of these systems: from representational accuracy – does the model stand for the world? – to the material and epistemic conditions that determine what forms of presentation become possible, and what remains unrepresentable.

We question whether the inherited vocabulary of representation, standing for, mediating, encoding, adequately describes what deep representations do, or whether it projects onto them a referential logic they have structurally moved beyond.

References

Agre, Philip E.
1997. *Computation and Human Experience*. Cambridge University Press.

Berger, John.
1972. *Ways of Seeing*. Penguin Books.

Berry, David M.
2011. *The Philosophy of Software: Code and Mediation in the Digital Age*. Palgrave Macmillan.

Blanchette, Jean-François.
2011. "A Material History of Bits." *Journal of the American Society for Information Science and Technology* 62 (6): 1042–1057. https://doi.org/10.1002/asi.21542

Crawford, Kate, and Trevor Paglen.
2021. "Excavating AI: The Politics of Images in Machine Learning Training Sets." *AI & SOCIETY* 36 (4): 1399–1399. https://doi.org/10.1007/s00146-021-01301-1

Denton, Emily, Alex Hanna, Razvan Amironesei, Andrew Smart, and Hilary Nicole.
2021. "On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet." *Big Data & Society* 8 (2). https://doi.org/10.1177/20539517211035955

Dhaliwal, Ranjodh S., Théo Lepage-Richer, and Lucy Suchman, eds.
2024. *Neural Networks*. University of Minnesota Press.

Dourish, Paul.
2017. *The Stuff of Bits: An Essay on the Materialities of Information*. The MIT Press.

Drucker, Johanna.
2020. *Visualization and Interpretation: Humanistic Approaches to Display*. The MIT Press.

Gaboury, Jacob.
2015. "Hidden Surface Problems: On the Digital Image as Material Object." *Journal of Visual Culture* 14 (1): 40–60. https://doi.org/10.1177/1470412914562270

Gitelman, Lisa, ed.
2013. *"Raw Data" Is an Oxymoron*. The MIT Press. https://doi.org/10.7551/mitpress/9302.001.0001

Ho, Jonathan, Ajay Jain, and Pieter Abbeel.
2020. "Denoising diffusion probabilistic models." *Advances in Neural Information Processing Systems* 33: 6840-6851.

IBM.
2024. "What Is Data?" Webpage. https://www.ibm.com/think/topics/data

Impett, Leonardo, and Fabian Offert.
2026. *Vector Media*. University of Minnesota Press.

Kant, Immanuel.
1998. *Critique of Pure Reason*. Translated by Paul Guyer and Allen W. Wood. Cambridge University Press.

Kingma, Diederik P., and Max Welling.
2013. "Auto-encoding variational Bayes". Preprint, arXiv. https://arxiv.org/abs/1312.6114

Mackenzie, Adrian.
2017. *Machine Learners: Archaeology of a Data Practice*. The MIT Press.

MacKenzie, Adrian, and Anna Munster.
2019. "Platform seeing: Image ensembles and their invisualities." *Theory, Culture & Society* 36 (5): 3-22.

Malevé, Nicolas, and Katrina Sluis.
2023. "The Photographic Pipeline of Machine Vision; or, Machine Vision's Latent Photographic Theory." *Critical AI* 1 (1–2). https://doi.org/10.1215/2834703X-10734066

Parikka, Jussi.
2023. *Operational Images: From the Visual to the Invisual*. Minnesota University Press.

Pasquinelli, Matteo, and Vladan Joler.
2021. "The Nooscope Manifested: AI as Instrument of Knowledge Extractivism." *AI & SOCIETY* 36 (4): 1263–80. https://doi.org/10.1007/s00146-020-01097-6

Prabhu, Vinay Uday, and Abeba Birhane.
2020. "Large Image Datasets: A Pyrrhic Win for Computer Vision?" Version 2. Preprint, arXiv. https://doi.org/10.48550/ARXIV.2006.16923

Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer.
2021. "High-Resolution Image Synthesis with Latent Diffusion Models." Preprint, arXiv. https://doi.org/10.48550/ARXIV.2112.10752

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox.
2015. "U-net: Convolutional networks for biomedical image segmentation." In *International Conference on Medical image computing and computer-assisted intervention (MICCAI 2015) Proceedings Part III*, edited by Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

Rose, Gillian.
2023. "Being Specific about 'Digital Images'." *Visual Studies* 38 (2): 181–82. https://doi.org/10.1080/1472586X.2023.2198338

Sterne, Jonhatan.
2015. "Compression. A Loose History." In *Signal Traffic: Critical Studies of Media Infrastructures*, edited by Lisa Parks and Nicole Starosielski, 31–52. University of Illinois Press.

Terras, Melissa.
2016. *Digital Images for the Information Professional*. Routledge.