mj

Published by: © Membrana Institute
Publisher's Homepage: www.membrana.org

membrana
JOURNAL

The Mystics and Magic of Latent Space: Becoming the Unseen

John S. Seberger, R. Aubrey Slaughter

Author(s): John S. Seberger, R. Aubrey Slaughter
Title: The Mystics and Magic of Latent Space: Becoming the Unseen
Published Online: June 10, 2020
Published in Print: 2020
Doi: 10.47659/m8.088.art
Referencing Style: Chicago
Section: Article (peer-reviewed)
How to Cite the Article: Seberger, John S., and Richard Aubrey Slaughter. 2020. "The Mystics and Magic of Latent Space: Becoming the Unseen." *Membrana – Journal of Photography, Theory and Visual Culture* 5 (1): 88–93. https://doi.org/10.47659/m8.088.art.

Published in: *Membrana – Journal of Photography, Theory and Visual Culture*, ISSN: 2463-8501(Print), 2712-4894 (Online)
Journal Homepage: https://www.journals.membrana.org/index.php/membrana/index
Submit to the Journal: https://www.journals.membrana.org/index.php/membrana/about/submissions

The Mystics and Magic of Latent Space

Becoming the Unseen

John S. Seberger and R. Aubrey Slaughter

“The latest advances in cognitive computation are a move inexorably towards a shamanism of the machine, a magical phenomenology based on fanciful but effective latent structures that we lack either the capacity or the sensorium to interrogate.”

Bridging concerns from human-computer interaction (HCI) and media studies, this essay theorizes deepfake images in terms of their phenomenological implications: the extent to which they enfold the human viewer in a world of the otherwise unseen. Drawing on comparative phenomenology of Vilém Flusser and Louis Bec, we focus on variational autoencoders (VAEs). We contend that the processes underlying deepfake image construction, as much as deepfake images themselves, evidence a parallel, prosthetic, and computational phenomenology: a study of “that which appears” to a computer, and which appears secondarily to a user-human as image. We use the example of VAEs to argue for the emergence of a second-order, received phenomenology of the augmented human as we reside in an increasingly computational world.

Keywords: deepfake, computer vision, augmented reality, computer phenomenology, magic phenomenology, machine shamanism

Two roads diverged in a yellow wood and I,
I split the difference.

(With apologies to Robert Frost)

Introduction

The vampire squid – or any type squid for that matter – has no spine. This is an old divergence in the evolutionary road: to have or not to have, to be or not to be (in one form or another). Even in its age, the dichotomy of the vertebrate and the invertebrate mirrors a more contemporarily troubled dichotomy: that of the subject and the object (Gibson 1977; Johnson/Latour 1988). How do computational forms of vision allow us to see the unseen of human subjectivity and the lifeworld that emerges through and by means of such subjectivity? What remains obfuscated – unseen – beyond the traditional subject/object dichotomy?

With the continued advancement and proliferation of computational technologies, the historically dichotomous categories of the subject and the object are less mutually exclusive than they appear to be in the realm of grammar and written/spoken communication. Computational objects demonstrate agency in prosthetic forms of sight, audition, and cognition in the form of data processing; subjects (in the reductionist form of “the user”) *are configured* (Woolgar 1990), acted upon by their devices: they are seen, heard, analyzed. In an inversion, objects sense and subjects *are sensed*; computational objects interpolate and we, as users, *are interpolated*.

Scaling Up

But computational interpolation is not limited to the relatively small scale of the individual user. Along with the user – as more and more of us fall into that category, having computational devices in our bags and pockets, automobiles, homes, workplaces, and neighborhoods – the world that constructs the user is interpolated, refreshed. Even now, different generations of humans effectively live in different worlds (Serres 2015), despite these worlds’ foundations in the same conditions of embodiment. The negative space that emerges from the known, experienced lifeworld of the human – the unseen – becomes seeable: the mystical, the magical. A fundamental question of our proximal futures then emerges: How is the human lifeworld seen through the eyes of computation? And how will humans, as users, see our worlds through the lens of that which is already computationally seen?

In the spirit of Vilém Flusser and Louis Bec (Flusser and Bec 2012), the mention of the squid and its spinelessness that began this essay hints at a form of phenomenology that is increasingly relevant as computers and their visions both inherit and interpolate the lived world of the human, or the point at which the biological umwelt of

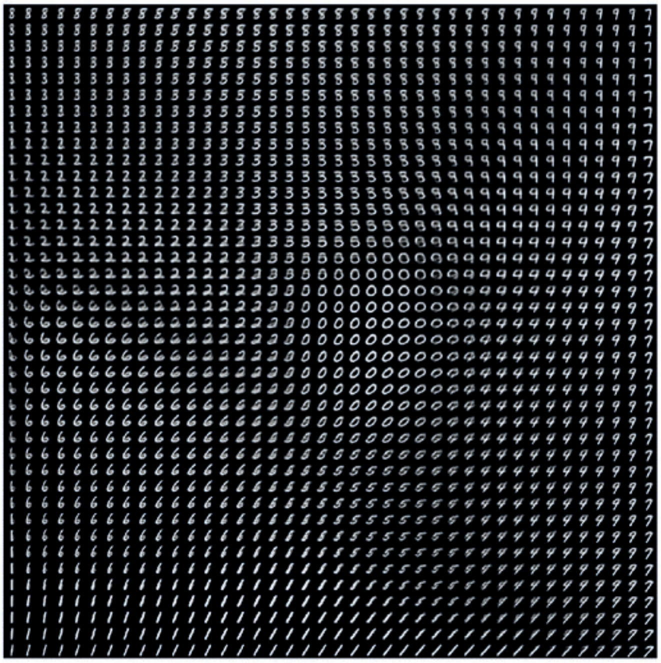

Figure 1. A visualization of handwritten digits, 0 – 9, generated by a variational autoencoder. https://www.jeremyjordan.me/variational-autoencoders/

the human merges with the social, cultural, historical, and archival. This form is that of comparative phenomenology: a comparison of the ways in which the world appears to various actants of different embodiment/materiality. Just as a squid has no spine, a computer has no inherent subjectivity – computers are machines of objectification. And yet, humans insert their subjectivity into computational outputs: if the computer "sees", then users see what is already seen and was heretofore un-seeable to the naked human eye. The human comes to see itself through that which has already been seen by the computer.

This paper explores how the human, as biological and cultural-historical actant, might see itself anew in an increasingly computerized world: how we might be recreated in the image of that which has been phenomenologically inaccessible, unseen. In particular, we (the authors) begin our exploration through the lens of deepfake technology and the neural networks – variational autoencoders (VAEs) in particular – that underlie them.

Neural Networks and VAEs

Neural networks are computer systems modeled on the function of neurons in the brain. Where, historically, we have used metaphors of the most advanced technologies to describe the brain – hydraulic systems, chemical soups, circuits – we have now reached a point of Ouroboric feedback: the technologically deciphered brain becomes, itself, the foundation for novel technologies. These novel technologies create separate and separable distortions of the human world of perception.

Broadly, a neural pathway that gets used more gets stronger, whereas one that does not get much use becomes weaker. If, for example, you lived in a world that was exclusively different shades of blue, your neural pathways for detecting, differentiating, and discerning between different types of blue would be enhanced, whereas your ability to tell puce from mauve would be diminished. Neural networks work in a similar way: they "learn" to detect patterns in data and, having done so, get better at detecting the patterns that they "know".

This sort of pattern detection is useful when you want to store a lot of data in a very small space, such as in image compression. Sending or storing large image files is costly, so these files are often compressed for transmission and storage. Many forms of image compression are lossy, in that they lose some nuance of the image in the compression process. One can think of a monochrome copy of a full-color image as a particularly lossy form of image compression; though color has been sacrificed, the monochrome image now takes up less than a third of the space than the full-spectrum original.

Of course, this kind of lossiness is high price to pay for a little extra storage space. Most image compression tools use algorithmic processes to ensure that compression is as lossless as possible – that the human eye won't detect loss. Although continuous pigments are rendered a discontinuous pixel through this process, these image compression algorithms are designed so as to render the difference between the two nearly imperceptible, already flirting with the interstices of the seen and unseen.

Imagine I have handed you a picture of a human face. Asked, "What's in the image?" you are likely to say, "a face". You are less likely to tell me that the picture consists of, say, 1024 pixels and subsequently list the hex color codes of each one. Humans and computers communicate what is seen differently. ("A face" is more human-readable than a description of a thousand or so pixels.) Like forests and trees, what is communicated signifies what is seen and unseen.

In this example, the list of each pixel's hex color code represents the (computer-oriented) lossless compression of the image. On the other hand, the (human-oriented) "face" description is, to a computer, highly lossy. If you tell me that the picture is of a face, I and my computer are unlikely to perfectly reproduce the picture. But, because I have seen faces before, I know something of how they are structured – eyes, ears, nose, chin, cheeks, jaw, forehead – and can draw upon what I know of their structure when I imagine the image you describe. While I am unlikely to perfectly reproduce the face in the image you are looking at, I will be able to quickly and effectively approximate it.

This is roughly how autoencoding neural networks function – what is "known" (e.g., the color blue, the structure of faces) informs the output. First, the autoencoder reads a set of images. Then, it attempts to find patterns in the images, regularized co-occurrences of structure in the data. These patterns are referred to as *latent structures*. (You can think of latent structures as rules or heuristics that the neural network has learned in the process of encoding the data. For illustration, the grammatical axiom "I before E, except after C" is an example of the latent structure of the English language. Like many of the rules of spelling, this structure is not to be found in the data itself, but is rather latent, and produced in the process of learning to "encode" thoughts into English.)

As the autoencoder reads images, it learns to recognize the latent structures. The more latent structures it can recognize, the more rules that it has for understanding how images are structured, the more accurately the autoencoding neural network can "draw" the image. But, just as with humans, neural networks do not have an infinite capacity for storing rules. You can perfectly describe a data set of *n* size using *n* rules, but this offers nothing in terms of compression; the set of rules could be larger than the data itself, and you would be better off memorizing images pixel by pixel. So, autoencoders need to use as few rules as possible to effectively compress an image, but need as many rules as possible in order to effectively decompress the image again. These rules, or *latent* *structures*, are representations of the data stored in what is referred to as *latent space*. The bigger the latent space is, dimensionally speaking, the more representations of the latent structure of the data we can store, and the more accurately we can decompress, decode, or "draw" the initial data.

Each of these latent structures describes a dimension in which it is structurally possible for an image to emerge. Through a recombination of latent structures, the original image can be effectively decoded from its encoded representations, and "redrawn". As the rules about the data take up less space than the data itself, we are also able to store the data in a more compressed form, without excess loss.

These sorts of autoencoding neural networks are designed for image compression, and are excellent at effectively encoding an image, storing its latent structure, and then decoding these structures back into an image. What they cannot do, however, is generate new images based upon these latent structures, other configurations of lines and color that are not in the original image, but are structurally possible given what the system has learned of the latent structure. For that, a slight variation on the autoencoder principle is required.

Known as a variational autoencoders (VAEs), these neural networks work in a similar manner to autoencoding neural networks, in that they both decompose and encode input data into a set of lower-dimensional rules than it then uses to decode output data. The twist is that variational autoencoders do not encode data as latent structures, but rather as latent *variables*; clouds of probability rather than points of precision. While the latent spaces of autoencoders are striated and discontinuous, the latent spaces of variational autoencoders are smooth and continuous; the spaces between points of structuration are filled with potentially actualizable, structurally possible alternates which *could* have been encoded using the same rules as before.

If given a series of digits of ten handwritten digits from zero to nine, an autoencoding neural network will return a series of ten handwritten digits, from zero to nine. A *variational* autoencoding neural network, on the other hand, will generate every structurally possible variation of these ten handwritten digits, including digit-like glyphs that were not included in the original data (**fig. 1**).

The unseen emerges from the seen. Variational autoencoding neural networks use the latent structures within images in order to produce images that are latent within these structures – possible but not perceptible. In other words, variational autoencoders generate the otherwise unseen, with reference to the otherwise unstructured; they regularize and structure, rendering visible the seeable unseen and normalizing the otherwise discontinuous. As computers come to see the world on behalf of their users, the world of the seen merges with the world of the unseen: the overlapping umwelts of the computer and the human – one prosthetic and the other organic – merge.

Deepfake Images

Marvin Gaye (and anyone who's ever sung "Heard it Through the Grapevine") once sang the line, "Believe half of what you see and none of what you hear."' It's a more cautious approach to the truth of vision than appears in an old saying: "seeing is believing". Seeing might still be believing, but that in which we believe through sight is changing. To see the outputs of computational vision systems is to believe in the world as computational: to believe in a world of near-magic that subtends or surrounds the world of the humanly perceptible.

Deepfake images are uncanny. They are computationally derived creations of the visual world grounded in the modes of image processing described above. And they are increasingly common. These images are used in politics, pornography, art, and advertising. They call into question the limits and finality of the world that is versus the world that might be or might have been. Perhaps as is already evident in their name, deepfake images are instruments of deception: they are fake, fugazi. Despite the deception central to the use of deepfake imagery, they must have a basis in truth; it would hardly be a good deception otherwise. Deepfakes do not simply create dubious doubles out of whole cloth, but rather by discovering hidden truths about how a person moves, acts, and appears; the latent structures of our visual identities.

It is possible to be deeply faked only if you are deeply, computationally known. Otherwise, the deception is superficial, flawed; it neither represents accurately nor misrepresents effectively. Deepfake technologies and their autoencoding cousins can be understood as technologies for seeing the unseen, whether it be in the hidden truths within the latent structures of our data and our *selves*, or the factually false but structurally sensible deepfakes it produces. In either case, these unseens are occulted from an anthropic view not because they are small, or hidden, or otherwise secreted. We can only ever approach these unseens vicariously, at a remove, through the machine.

Becoming the Unseen

Within the processes that underlie deepfake image production, we find evidence of a computational umwelt (Von Uexküll 2009). Just as for Flusser and Bec (2012), the world of the squid is fundamentally different than the world of the human, the world of the computer is, too, different. Nevertheless, it is a world wherein sensation and perception – or prosthetic, computational analogs thereto – emerge. But the computer's is not an umwelt in the traditional sense of the term.

Computers cannot reasonably possess a lifeworld because they possess no life. They do not breathe, eat, or procreate. (To paraphrase Haugland, they just don't give a damn.) But they do exist *within* a biological umwelt: they exist within that of the human, each one a rabbit hole, a looking glass to go through into the realm of the otherwise unseen. The prosthetic umwelt of the machine overlaps with the human umwelt. In this overlap, we find evidence of the magical.

In the case of VAEs, the computer constructs a visual world based on the inputs it receives. The overt construction of this world appears at the point of output – that which is seeable to the human eye – but it is not a construction solely intended for computational use: computer outputs become human inputs when the computer is placed in relation to a user. There is magic in the transformation from computer-output to human-input: in the form of deepfake images, predicated as they are on phenomenologically alien and inaccessible modes of prosthetic vision, we encounter the unseen: a world beyond the reach of human sensation and perception, but which is ultimately fed through the process of computational output into the human umwelt. The visual realm of the human umwelt becomes somewhat more magical through the seeing of the unseen (particularly if the seer does not understand how deepfakes are produced).

Magic is not a very well-defined concept (Glucklich 1997), and is "used for distinctly different purposes and to denote distinctly different phenomena" (Lee and Schmidt 2018). What is defined as magic is contingent on both where and when we are (Kieckhefer 1989) and "all definitions of magic are relative to the culture and sub-culture under discussion" (Segal 1981). As with similarly relational concepts such as infrastructure (Lee and Schmidt, 2018) privacy (Thomson 1975) and data (Gitelman 2013), magic may be considered a "generalizable category" (Lehrich 2009), but one that is neither stable (Bailey 2006) nor absolute (Bernd-Christian and Stausberg 2014). A comprehensive definition of magic remains elusive, but a comprehensive definition of magic is not a necessary prerequisite to the study of magic. It is critical to examine the ways in which "Modernity has depended upon the surreptitious – and magical – power it so denies, and the study of magic provides a vivid window onto the cultural logics upon which the modern world has been structured" (Steyers 2013). Defining what magic is in a general sense is difficult, and this analysis does not intend to offer a definition. Instead, we focus on the relations between the occult and the anthropic, and recognize that such relations have historically been the purview of magic and magical practice.

Magic can be understood as the art of occult aspection, a series of tools and techniques for discovering hidden interconnections, structures latent to the otherwise imperceptible. Though it could be argued that many of the more esoteric latent structures posited in magical traditions do not actually exist, this is less of a problem than it initially appears; "In general, we don't need to worry about ensuring that the latent structure exists. If such latent structure helps the model ... then the network will learn that structure at some layer" (Doersch 2016). That is to say, it does not matter whether or not the latent structures that one has learned to recognize actually exist; what matters is whether or not learning these latent structures enables you to accurately and effectively "encode" or understand the original data. Though these latent structures only exist insofar as they are useful, they are real in the sense that they inform and inflect our relations to the otherwise insensible. This is as true for people as is it for neural networks; to offer a computational twist on the Thomas theorem, "If [computers] define situations as real, they are real in their consequences" (Hildebrandt 2011; Thomas and Thomas 1938).

As a practice of seeing the unseen, magic is predicated on hidden connections, occulted patterns, latent structures; seeing the unseen requires an unseen. In regard to deepfakes and neural networks, we have two forms of the unseen at play: first, in the otherwise unseen of the latent structure; second, the otherwise unseeable variants of said latent structures. The same might also be said of magic: that magical practices are oriented towards learning the latent structure of the otherwise occult in order to make effective use of said structures. Understanding how magic does what it does is no prerequisite in either case; "Comparisons between machine learning and magic are common even amongst experts and practitioners [….]. They work, but it is difficult to explain why or how" (Browne and Swift 2018). In this sense, the latest advances in cognitive computation are a move inexorably towards a shamanism of the machine, a magical phenomenology based on fanciful but effective latent structures that we lack either the capacity or the sensorium to interrogate.

When we, as users, interact with this unseen realm, we become less residents of the seeable world than residents of the unseen. We abandon, however reluctantly or unknowingly, the world of the physical subject – the world of direct human experience, wherein "that which appear"' appears because it does so *to us*. In such abandonment, we become objectified subjects: subjects living in a world known through the lens of the computational object. This is a dark magic, indeed.

In a future undergirded by the ubiquity of computation – and the presumed ubiquity of computational visual mediation – the human user comes to reside doubly in the world of the seen and the unseen. That which is biologically un-seeable to the human becomes seeable to the computer, and therefore seeable to the user. When we live in this world, we are ontologically chimerical: the human becomes both perceiving subject and object *because of* what they are perceiving: images that are objectified so as to be computer-readable and then deconstructed into latent space and reassembled into some form of output. The human-as-user becomes – along with their cultural, societal, and historical conditions – objectified through enrollment into the vision of the computer. We become that which we were previously unable to see.

References

Bailey, Michael D. 2006. "The meanings of magic." *Magic, Ritual, and Witchcraft* 1 (1): 1–23.

Browne, Kieran, and Ben Swift. 2018. "The Other Side: Algorithm as Ritual in Artificial Intelligence." In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–9. CHI EA '18. Montreal QC, Canada: Association for Computing Machinery. https://doi.org/10.1145/3170427.3188404.

Doersch, Carl. 2016. "Tutorial on Variational Autoencoders." *ArXiv:1606.05908 [Cs, Stat]*, August. http://arxiv.org/abs/1606.05908.

Flusser, Vilém, and Louis Bec. 2012. *Vampyroteuthis Infernalis: A Treatise, with a Report by the Institut Scientifique de Recherche Paranaturaliste*. University of Minnesota Press.

Gibson, James J. 1977. "The Theory of Affordances." In *Perceiving, Acting, and Knowing*, edited by Robert Shaw and John Bransford, 67–82. New York: Lawrence Erlbaum Associates.

Gitelman, Lisa, ed. 2013. *Raw data is an oxymoron*. MIT press.

Glucklich, Ariel. 1997. *The end of magic*. Oxford University Press

Hildebrandt, Mireille. 2011. "The Rule of Law in Cyberspace?" Inagural Lecture, Radboud University Nijmegen, December 22. http://works.bepress.com/cgi/viewcontent.cgi?article=1060&context=mireille_hildebrandt.

Johnson, Jim, and Bruno Latour. 1988. "Mixing Humans and Nonhumans Together: The Sociology of a Door-Closer." *Social Problems* 35 (3): 298–310. https://doi.org/10.2307/800624.

Kieckhefer, Richard. 2000. *Magic in the middle ages*. Cambridge University Press.

Lehrich, Christopher I. 2007. *The Occult Mind: Magic in Theory and Practice*. Cornell University Press.

Segal, Alan F. 2015 "Hellenistic magic: Some questions of definition." *Studies in Gnosticism and Hellenistic Religions presented to Gilles Quispel on the Occasion of his 65th Birthday*, 349–375. Brill.

Serres, Michel. 2015. *Thumbelina: The Culture and Technology of Millennials*. Rowman & Littlefield International.

Sørensen, Jesper. 2013. "Magic reconsidered: Towards a scientifically valid concept of magic." In *Defining magic: A reader*, edited by Bernd-Christian Otto and Michael Stausberg, 229–242. Routledge.

Styers, Randall. 2013. "Magic and the Play of Power." *Defining Magic: A Reader*, edited by Bernd-Christian Otto and Michael Stausberg, 255–262. Routledge.

Thomas, William I., and Dorothy Swaine Thomas. 1938. *The Child in America*. New York: Alfred A. Knopf.

Thomson, Judith Jarvis. 1975 "The right to privacy." *Philosophy & Public Affairs*, 295–314.

Von Uexküll, Jakob. 2009. "A Stroll through the Worlds of Animals and Men: A Picture Book of Invisible Worlds." *Semiotica* 89 (4): 319–391. https://doi.org/10.1515/semi.1992.89.4.319.

Woolgar, Steve. 1990. "Configuring the User: The Case of Usability Trials." *The Sociological Review* 38 (S1): 58–99. https://doi.org/10.1111/j.1467-954X.1990.tb03349.x.

Short biographies

John S. Seberger is currently a Postdoctoral Fellow in the Department of Informatics in the Luddy School of Informatics, Computing, and Engineering at Indiana University, Bloomington. In 2019, he received his PhD in Information and Computer Science from the University of California, Irvine. Additionally, John holds a BA in Psychology from Kenyon College, an MSc in Research Methods in the Psychology of Music from Keele University, and an MLIS from University of Pittsburgh. He is the recipient of numerous awards and fellowships, including a Graduate Research Fellowship from the Intel Science and Technology Center for Social Computing, research funding from the Internet of Things Group at Intel, and GAANN fellowships from the United States Department of Education. Focusing largely on the philosophical impacts of the Internet of Things (IoT), John specializes in humanistic and interdisciplinary approaches to human-computer interaction (HCI), the relationship between emerging technologies and theories of the archive, and infrastructure studies.

Aubrey Slaughter is currently a doctoral candidate in the Informatics department at the Donald Bren School of Information and Computer Science at the University of California, Irvine. His MA in Social Science was conducted at the University of Chicago, and his BA in Anthropology at the University of Texas, Austin. He is a recipient of the 2016 DTEI Pedagogical Fellowship, and has received funding from the National Science Foundation for the studies of sameness in scientific sampling. Publication of his latest work is forthcoming in the 2021 *Routledge Handbook of Digital Media and Communication.* Principally concerned with the intersection between the seen and the unseen in human/infrastructural relations, he specializes in examining non-standard uses of informational technology, such as automated blackmail, ad hoc hacktivism, and magical infrastructures.