The Return of the Uncanny: Artificial Intelligence and Estranged Futures

Anthony Downey

Abstract

Although often considered to be a fault or a glitch in the system, the event of hallucination is central to the generative models of image processing employed by artificial intelligence (AI). Evoking an inhuman logic, the works in Trevor Paglen's Adversarially Evolved Hallucinations series (2017–ongoing) mine this hallucinatory space. Employing a generative adversarial network (GAN), the resulting images in the series depict the uncanny domains of automated image production and, this essay will argue, effectively question the efficacy and purpose of deploying such technologies in image processing tasks. Through training neural networks—such as those employed in a GAN—to see the world for us, this essay also investigates whether we are priming and instructing ourselves to effectively see like machines, despite their questionable value and apparent convenience as image processing platforms. What will happen when these inhuman, hallucinatory models of seeing wholly supersede human vision, not least in the adjacent fields of surveillance and automated models of warfare? To pose such questions is to address a potentially more radical component in generative AI: if automated models of image production, complete with their hallucinatory inclinations, replace ocular-centric ways of seeing, will the affordances of such technologies further estrange us from the present and, indeed, the future?

Keywords: Artificial Intelligence (AI), Visual Culture, Digital Methodologies, Generative Adversarial Networks (GAN), Trevor Palgen, The Uncanny

If an image could be described as baleful, or as having an ominous appearance, Rainbow would certainly fit the bill (fig. 1). Apart from the toxic-looking "sky," parts of it appear to have mutated into the fiery trace of munitions or, more cryptically, a series of glitches. Suggesting the collation of natural elements and a physical, possibly dead body (corpse/corps), the full title of the work—Rainbow (Corpus: Omens and Portents)— further bolsters the overall impression of trepidation and estrangement. Complete with the apparition of deceptively unseeing eyes, this sense of apprehension is equally evident in Human Eyes (Corpus: The Humans) where features that resemble—or, more likely, re- assemble—the components of a face mutate into a monstrous visage (fig. 2). Appearing both present and yet disconnected, in Vampire (Corpus: Monsters of Capitalism) the eyes return but this time with a more cartoonish, disembodied countenance (fig. 3).Something is awry in these images which, for the most part, appear to be almost "right" but not quite.

Evoking an inhuman logic, Rainbow, alongside other works in Trevor Paglen's Adversarially Evolved Hallucinations series (2017–ongoing), mines the latent spaces of automated image production. Produced by a generative adversarial network (GAN), an artificial intelligence (AI) model that trains itself on a dataset of images in order to recognize, classify, and generate new ones, these unnerving visions suggest an uncanny realm where the familiar and unfamiliar are fused into an embryonic space of image production.1 Given that AI image-processing models do not experience the world in phenomenological, embodied terms but replicate a once-removed and askew version of it, the peculiar characteristics of images such as Rainbow reveal how algorithms computationally generate disquieting allegories of our world.2 The outcome of algorithmic processes, the works we thereafter encounter offer a view into the "subconscious," often concealed machinations of AI.3 Throughout this series, a recognizable order of being in the world—be it physical, rational, or otherwise—is dis-placed, or usurped, by a potentially more alien and disturbingly mechanized order. Something

comes to light in this spectral realm: an apparition, or a nightmare, that is indebted to the hidden and invariably recursive logic of algorithmic apparatuses.4

The inhuman processes at work in AI models of image classification and production prompt a series of questions concerning the degree to which machinic ontologies of perception—powered by algorithmic ratiocinations—are disrupting the ocular-centric field of human vision. How, we could subsequently ask, do machines see the world? To this we could add another, perhaps more pertinent, question: How do machinic methods of seeing determine, if not overdetermine, how we perceive and experience the world? Through training neural networks—such as those employed in a GAN—to see, are we priming and instructing ourselves to see like machines? In time, moreover, will inhuman models of seeing supersede human vision in certain areas, not least in the adjacent fields of surveillance and automated models of warfare? To pose such questions is to address a potentially more radical component in prototypes of technologically-induced sight: If AI models of image production replace ocular- centric ways of seeing, do these models have the capacity to further estrange us from the world? It is through addressing these and other questions that we can explore the far-from-abstract impact of delegating the ocular-centric regimen of human perception to a preprogrammed regime of machine vision.5 Although presented as an objective "view from nowhere," AI models of image recognition, classification, and production are designed to identify images according to input (datasets) and instructions (algorithmic weightings). Inasmuch as these apparatuses produce models that are riven with political, racial, and gender- based bias, the hermeneutic ambition underwriting systems of AI image-processing—the impetus to interpret and categorize—can be understood in epistemological terms: they produce meaning and endeavour to make sense, in often outlandish but nonetheless narrow terms, of our world.6 AI can, as a result, reductively encode our perception of the world through machinic frames of reference. It is with these and other concerns in mind that the Adversarially Evolved Hallucinations series examines how machine learning—working from datasets (images)—functions as a computational means to produce knowledge (epistemologies) and, throughout that process, promote AI as a heuristic device: capable, that is, of making sense of, if not predefining, how we perceive the world. As we will see, AI image-processing techniques perform on varying scales of error, so much so that, as Paglen notes, "the apparent correspondence between what a model is classifying and how it relates its systems of classification to referents 'out there in the world' is not only misleading but hallucinatory."7 The epistemological affect of automated image generation, however hallucinatory the latter may turn out to be, can veer from the merely misleading to the coercive. To fully explore the ramifications of machinic perception, we need to engage in a diagnostic form of reverse engineering: working backwards from the manifest, final iteration of an image such as Rainbow, we can explore how the systematic training of a neural network—through the use of datasets—produces images. Through examining Rainbow, alongside other works in the series, we can not only better understand how AI models are systematically trained on datasets, we can also address their epistemological impact—or, more specifically, how they function to simultaneously adumbrate and yet overdetermine our world.8 When considering the systematic training of a GAN model of image-processing, we likewise need to investigate how neural networks are systemically calibrated by algorithms. It is from within this latent, methodically obscured, space of algorithmic reasoning—involving as it does the machinic calibration of neural networks—that AI produces questionable surrogates and hallucinatory visions of our world.
Through foregrounding the operations involved in compiling, labelling, and algorithmically rationalizing the input data that powers neural networks, Paglen invites us to deconstruct the technological components involved in training a machine to see. This approach makes

known, to begin with, how the categorical definitions attributed to datasets imply the potential for epistemological violence—the degree to which, that is to observe, machinic interpretations reduce our world to normative and non-normative, or proscriptive, categories. In undermining the reliability of the image-recognition tasks so readily undertaken by AI, alongside their disputed capacity to fully, if ever, make sense of the complex realities of our world, Adversarially Evolved Hallucinations encourages us to see through the systematic, systemic, and epistemological dysfunctions of such apparatuses and question their present-day and, indeed, future impact on our perceptions of the world.

Categorical Dissonance and Epistemological Affect

In order to train a GAN, Paglen established a series of taxonomies with titles that included SPHERES OF PURGATORY, EYE MACHINE, THE INTERPRETATION OF DREAMS, and OMENS AND PORTENTS. Referencing sources from literature, visual culture, psychoanalysis and folklore, these idiosyncratic taxonomies tended to be both broadly obscure and yet reasonably identifiable. In part, this allusiveness addresses the expansiveness and complexity of knowledge systems, especially as they relate to the world. “Humans,” Paglen says, “have all sorts of weird taxonomies that we use to try to makes sense of the world: taxonomies for dreams, tarot cards, historical events, ideas about some things being ‘lucky’ or ‘unlucky,’ and even taxonomies of allegories.”9 For Adversarially Evolved Hallucinations, Paglen developed some of his conceptual taxonomies into corpuses that contained datasets.10 To return to Rainbow (Corpus: Omens and Portents), the corpus OMENS AND PORTENTS consisted of a dataset comprized of individual image categories such as “rainbows,” “comets,” “eclipses,” and “black cats.” In these examples, the choice of each image category for a dataset is far from restrictive or regulative. In fact, the cumulative effect of such diverse image categories would appear to contest the overarching restrictions of more functional datasets or, indeed, the supposed practicality and the validity, more generally, of generic classification systems.11

Through training an AI model on the OMENS AND PORTENTS corpus/dataset, the GAN began to recognize patterns and features associated with each image category and, in time, classify them. This process of classification is constantly grounded in the original corpus/dataset, so much so that the GAN can only ever classify images that the model has been already trained upon. This entire process of training may appear relatively straightforward if not a tad recursive: you instruct an AI model, using datasets, to classify and produce images similar to those it has been trained upon. However, and depending on the system in use, the process is never totally predictable, nor is it reliable. Among other factors, the training is contingent on biases in the datasets (whereby certain images are over- or under-represented), discrepancies in procedures and, notably, variables in the less-than-transparent adjustments involved in the iterative process of applying algorithmic weightings to input data. The operative logic of a GAN is, in addition, specifically geared towards generating new, as-yet- unseen images, which further renders the entire process subject to a significant degree of computational fortuitousness. Despite the sense of technological determinism often associated with algorithmic devices— the notion that the programmatic identification of patterns in datasets and the application of appropriate weightings to the values associated with such patterns will, in time, give correct predictions—the procedures involved do not automatically yield predictable outcomes.

Unlike an ocular-centric field of vision, GANs learn through the statistical analysis of data to capture patterns and features that exist within datasets. These patterns are the basis of the successive predictions, or classifications, that AI models formulate as outputs. Ultimately, these predictions are designed to calculate or recognize future patterns. When we look again

at Rainbow, it is obvious that it is an uncanny vision, or projected hallucination, of a rainbow, inasmuch as it possesses a passing but far-from-unqualified resemblance to one. There is a clear distinction to be had here between machinic and ocular-centric models of classification; however, for the GAN model the image it produced is categorically a "rainbow" insofar as it can only ever produce images associated with the dataset upon which the model has been trained.

This latter point remains central to our discussion, as the images we encounter throughout the Adversarially Evolved Hallucinations series have all been apportioned, with high levels of certainty, a definitive category—rainbow, comet, eyes—by a GAN. These frequently bizarre, if not uncanny, classifications convey the degree to which AI image-processing models are commonly, if not ubiquitously, involved in producing, to use Paglen's phrase, a form of "machine realism": "creating a training set involves the categorization and classification, by human operators, of thousands of images. There is an assumption that those categories, alongside the images contained in them, correspond to things out there in the world […] I refer to these assumptions as 'machine realism.'"12 The question we are left with is what happens when systems identify, or produce, a "rainbow" that is patently not a rainbow, at least not in the conventional sense. Or, similarly, what happens when a generative AI model—such as a GAN—creates, or hallucinates, an image that it announces to be, despite evidence to the contrary, a given thing that is patently not the entity in question. We return here to our earlier point: given the omnipresence of AI apparatuses (as evidenced in facial recognition technologies, for example) and their impact on how we see the world, how do we gauge the validity, or effect, of understanding the world through the affordances of machinic models of perception—"machine realism"—and computational frames of reference?

The apparently abstract event of (mis)classification, or hallucination, discloses the deterministic reasoning implied in AI models of image production—this is a rainbow; this is an apple; this is a face—and how it imposes meaning upon the world. The legacy of this imposition, its epistemological affect, is far from inconsequential: when deployed in facial recognition technologies, for instance, such systems assign a classification to a particular object or entity—say, a face— and assigns a name or, more ominously, a level of threat to it. Respectively, there often exists a concomitant tendency to take these categorizations for granted and act accordingly.13 In programmatically presenting the world through the computational inferences of neural networks, AI models of image-processing would appear to be increasingly programming us to accept machinic conjectures as the "truth" of our world rather than, as they in fact are, conditional projections and probabilistic predictions. Generative AI models, such as GANs, are statistical systems of rationalization that classify, with varying levels of efficacy, images and other data. Despite the inherently biased nature of machine learning (not to mention the tendency to hallucinate), we increasingly appear to be delegating responsibility for, and our responsiveness towards, the epistemological impact and affect of such technologies.14 Through the statistical analysis of patterns, conducted in order to calculate or recognize future patterns, mechanistic predictions of people's identities, shopping preferences, credit ratings, career prospects, health status, political affiliations, and supposed susceptibility to radicalization, become the norm rather than the exception. The innately machinic process of classifying an image can, in turn, provoke or bring about an action, or event, wherein which the calculus of an algorithmic "aperture" is procedurally focused on "distill[ing] something for action."15 The predictive inclination of AI technologies, their projective functioning and distillation of realities into precepts for action (which are usually disciplinary in nature), become self-fulfilling and unaccountable—if not unfathomable—principles in the determination of a subject's suitability across a range of situations, positions, and tasks. We confront here the implications involved in the machinic calculation of normative and—perhaps more troublingly for those caught in the capacious

ambit of automated models of image-processing and classification—non-normative activities, behaviors and subjectivities.

If Rainbow is a machinic analogy of a rainbow, summoned forth by algorithmic reasoning, how then do we understand the processes through which neural networks arrive at such images? How, that is to ask, do we think from within these systems rather than merely reflect upon their potential impact? To these inquiries, we could ask what happens when image-processing models are computationally deluded in their projections. Instances of hallucination in neural networks are, to be clear, neither rare nor unaccounted for; on the contrary, they are indelibly associated with a "counterintuitive and unexpected form of brittleness [that is] replicated across most deep neural networks currently used for object recognition."16 We could note here a particularly germane study involving an InceptionV3 image classifier that consistently classified an image of a turtle as a "rifle."17 The authors of the paper noted that as "an example of an adversarial object constructed using our approach," a 3D-printed turtle was "consistently classified as a rifle (a target class that was selected at random) by an ImageNet classifier."18 This occasionally dry technical detail reveals a profound reality that remains intrinsic to the neural networks and deep-learning models involved in training machines to see: they are not only systematically prone to category errors, they are also systemically susceptible to inventing (or hallucinating) objects that do not exist.

Image-processing models can also add interpretive context that is grievously biased, not least when we consider the widespread use of such apparatuses in policing. In an investigation undertaken by AlgorithmWatch in 2020, to take a particularly apt example of interpretive and epistemological affect, it was demonstrated that Google's Vision Cloud labelled "an image of a dark-skinned individual holding a thermometer [as a] 'gun' while a similar image with a light-skinned individual was labeled [as an] 'electronic device.'" Even though Google, once alerted to the bias, fixed it, the investigation by AlgorithmWatch went on to conclude that "the problem is likely much broader."20

We will return to the subject of "brittleness" below but, for now, I want to observe that the inclination to hallucinate or misclassify a specific class of image is not a one-off fault or glitch in the system; rather, the event of hallucination is central to the functioning of neural networks and their generative modelling of the world. Established through the statistical rationalization of data, AI produces hermeneutic structures that are often the outcome of distortion—hallucination—and opaque methods of algorithmic calibration.21 It is this element of distortion in the latent space of machine learning that Paglen activates when he explores, in conjunction with his inquiry into the systematic training of a neural network, the systemic, iterative contexts of machine learning. In focusing on the hallucinatory, latent, and systemic domain of algorithmic computation, we can see how conventional, and increasingly instrumentalized, applications of machine learning systems can be provisionally uncoupled from their utilitarian applications.

Machinic Hallucinations: How to See through Generative Adversarial Networks

Consisting of interconnected nodes or neurons, neural networks employ layers that mimic the function of biological neurons in the human brain. This is true of a GAN system where there is a layer for input data, one or more hidden layers where algorithmic convolutions occur, and an output layer for prediction or image classification. There are two operative neural networks in a GAN, both working in tangent with one another. Obstinate competition in the task of image classification (the responsibility of the discriminator) and image production (the function of the generator), ensures that these neural networks are, as the name suggests,

profoundly adversarial.22 Although mindful not to anthropomorphize neural networks (inasmuch as they are, technically, machinic methods of computation), we could consider the relationship between the discriminator and the generator as similar to that which exists between, respectively, a law-maker and a law-breaker.23 The discriminator (law-maker) is consistently preoccupied, in this reciprocal alliance, with discerning the difference between real images and "fake" images, whereas the generator (law-breaker) is absorbed with trying to "fool" the discriminator with synthetic, as-yet-unseen images.

Over the course of this competitive relationship, the discriminator is effectively encouraging (training) the generator to deceive it: the more convincing the generated image, the more likelihood it will be ascribed a specific class by the discriminator. A GAN can be therefore understood, in part, as an autodidactic, if not autopoetic, mechanism: it teaches itself to learn and make distinctions.24 This apparently dexterous process, based on an iterative procedure that involves looped models of feedback, has nevertheless proved to be a fertile ground for the generation of delusions—or hallucinations—and figments of the algorithmic "imagination." It is from within this occluded zone (often referred to as a "black box") that we can further locate the evolving imagistic logic that is central to the Adversarially Evolved Hallucinations series and how images such as Rainbow encourage the viewer to see through neural networks.

Although there are differences between how a GAN and other neural networks operate, the methods of rendering digital images or video data ready for processing is similar across most image-processing and image-classifying tasks. Images, digitally rendered and compiled into datasets, are commonly but not exclusively submitted in the form of a legible vector or raster-based model of representation, the latter being a rectangular matrix or grid of square pixels. When magnified, in the case of a raster-based image, a pixel appears as a square of sorts—or, more precisely, a "blob."25 Image-processing algorithms, used to train neural networks to see, assign a numerical value to these colored blobs. This value is based on intensities of colors, which are often represented by three-or four-color models such as red, green, and blue (RGB), or cyan, magenta, yellow, and black (CMYK). The numeric values attached to these intensities of color (or blobs/pixels) are thereafter weighted through the application of algorithms. Each of these weights, or biases, are repeatedly adjusted and attuned until a desired conclusion is realized; until, that is, the neural network classifies a certain image as being a "real" or identifiable image.

This is, in albeit simple terms, the basis of machine vision: images, rendered as pixels, are assigned a numerical intensity value that can be subsequently weighted by algorithms or, as is often the case, groups of algorithms. These weights, when calculated alongside other weights, can produce an estimation (output) as to what the input image represents. When a neural network has been calibrated to recognize and identify known images (inputs), images can be uploaded (again as numeric code) to train (test) its capacity for predicting a class of images. To this end, a neural network does not see an image as such; rather, it scans a series of numerical values that add up to, or stand in for, an image. Neural networks are trained, in sum, not on an image but on images converted into numbers—and it is from within this multidimensional, latent space of numeric manipulation that they begin to hallucinate.

While both the discriminator and the generator are engaged a zero-sum game of optimization, the overarching purpose of the former is to correctly classify both real and generated (fake or synthetic) data.26 In the latent spaces of computation, where the generator seeks to "fool" the discriminator, the counterfeited images that pass muster assume the distinction of being categorically "real." This is regardless of their frequently bizarre or, as we see in Rainbow, estranged and uncanny appearance. To fully appreciate how this occurs, we need to stress that in the initial stage of training the generator produces random noise that is relayed to the discriminator. The discriminator subsequently supplies the generator with automated

feedback as to how closely the generated data resembles the images that the model was trained upon. To the machine eye, which is in a recurrent state of algorithmic calibration and recalibration, some images will look more "real," as opposed to fake or synthetic, than others.27 To begin with, this resemblance might be as low as 0.00001%. However, through multiple iterations of the process (and allowing for the vast computational power that can be now brought to bear upon data inputs), this figure will eventually shorten.

The ultimate goal in a GAN is to therefore reach a stage where the generated samples are indistinguishable from the real samples. There exists, in consequence, a distinctly delusional basis to the entire generative process involved in training a GAN: the generation of synthetic images is designed to delude the sentinel- like structure of the discriminator. It is these so-called "fooling" images, an image type that the generator has produced to deceive the discriminator, that eventually become indistinguishable from real data. The sheer power of recursion, the looped subjection of data to countless iterations and weightings, seems to invite a spiralling sense of mechanical delirium, or hallucination. It is in this mise en abyme, where images mutate and transmogrify, that we can see how the brute "force[s] of computation" can give way to computational delusions.28

When we look again at the images in Adversarially Evolved Hallucinations, which often appear to be

in a state of perpetual evolution or collapse, we can recognize certain components in them, including shapes, edges, and forms. These subcomponents are called "primitives" (fig. 3) and Paglen understands them as being akin to a brushstroke or a pencil mark, whereby each of the primitives represents a basic shape or line that could constitute a bigger picture: "a banana is likely to have two arcs ranging from the top to the bottom of the fruit; it could have yellow color gradients, some brown spots, a stem at the bottom, and so on. Each of these subcomponents will be represented in the primitives of the image—arcs, nonparallel lines, brown spots, stems, and so on."29 In a neural network, primitives, the subcomponents of an image, exist in the latent space of the AI model and provide the amorphic foundations for machine learning to produce more and more complex images. The arcs, gradients, lines, colors, shapes, and other subcomponents will, over time and through iterative looped procedures, evolve into manifest images. Crucially, it is the intervention into the evolutionary stage of image production that reveals the systemic operative logic involved in algorithmically calibrating a neural network. As Paglen notes, it is at this stage of image evolution that he can instruct the neural network to "generate an image of 'neuron 7382,' or any other place (neuron) in the latent space. The generator then evolves an image in the direction of criteria dictated by the specificities of what is in the latent space."30

In this scenario, "neuron 7382" has been produced by the generator (law-breaker/counterfeiter) and, as Paglen acknowledges, he can intervene to instruct it to develop this abstract image or "neuron" toward ever more fantastical ends. Through intervening into the systemic process involved in training a neural network, Paglen can effectively harness the computational forces to develop—from a given point in the latent space of the iterative process—a given output (image), however bizarre, that can nevertheless "fool" the discriminator into believing it is a class of image that the model has been trained upon. In this context, Rainbow is indeed an image of a rainbow insofar as its training data (input) has been weighted to the degree that a synthetic image can pass for real, at least in the eyes of a GAN. Insofar as image-processing systems do not return exact replicas or accurate classifications of the world, they can hallucinate realities into being.31 It is this predisposition that Paglen heightens when he intervenes into the systemic, latent sphere of algorithmic reasoning. It is here, where images return to us in uncanny variations on a theme, that the common applications of image-processing algorithms can be critically disconnected from their utilitarian function and revealed for what they are: statistical approximations and mechanical

allegories of reality. Given the relative opacity in the systemic functioning of neural networks, the abiding concern is that the algorithmic rationalization of data—which employs a range of weights and biases to support machine learning processes to better recognize images—can pick up on patterns in data that simply do not exist except, that is, within the preserve of a computational illusion or in the pathologies of a mechanically induced delirium. From the outset of our discussion about GANs it is apparent that the seemingly delirious resolve to produce ever more accurate ("real") and yet "counterfeit" images can and does give rise to hallucinatory realms. This tenaciousness, an integral element in the operative logic of a GAN, is crucial and yet it divulges a seemingly pathological impulse toward generating ever more fantastical, if not phantasmal, images. Revealing inherent forms of "brittleness," these flashes of computational delirium contradict the frequently inflated claims made in relation to the effectiveness of neural networks in image-classification tasks. We return here to the uncanny affect of such systems, and the degree to which, regardless of their intrinsic failings, they are widely used to produce paradigms—or epistemological frameworks—for understanding the world. We could note here, in the context of datasets, AI, and the deterministic logic of such apparatuses, Taina Bucher's insights into how the algorithms that render neural networks viable are resolutely "political in the sense that they help to make the world appear in certain ways rather than others. Speaking of algorithmic politics in this sense, then, refers to the idea that realities are never given but brought into being and actualized in and through algorithmic systems."32 Algorithmically defined outputs, systemically calibrated from input data and optimized—modulated—by weightings, are always already political inasmuch as they summon forth computational, routinely normative, models of our world.

Notwithstanding the misplaced degree of confidence in AI, alongside the proven shortcomings, or should that be excesses, of neural networks, computational projections are frequently presented as categorically deterministic—this is a rainbow; this is a face; this is a threat—rather than, as they are in reality, approximate estimates of a given reality based on statistical inferences garnered from patterns in a dataset. Given the accumulative and ascendant influence of AI on our lives and how we live, there is a strong argument here for developing research methods—such as those deployed throughout the Adversarially Evolved Hallucinations series—that are designed to encourage a critical range of thinking from within these structures rather than merely reflecting upon their impact. Through developing such research, we can ensure that the systematic methods (involved in labelling and inputting data, for example) and systemic (latent and algorithmic) spaces of computation are more readily understood for what they actually are: statistical calculations of probability that seek to define our realities and, in so doing, further estrange us from the world and our futures.

List of Images:

Fig. 1 Rainbow (Corpus: Omens and Portents

Fig 2. Human Eyes (Corpus: The Humans)

Fig 3. Vampire (Corpus: Monsters of Capitalism)

1. The concept of the uncanny as an unhomely or frightening apparition is key to Sigmund Freud's seminal essay "The Uncanny," first published in 1919. It is here that Freud observes how the German word unheimlich is "obviously the opposite of 'heimlich'

[homely], ‘heimisch’ [native]” and therefore the inverse of the familiar or not known. Sigmund Freud, “The Uncanny,” in Art and Literature, vol. 14, The Pelican Freud Library (London: Penguin Books, 1988), 341.

2. There are numerous critiques of AI in relation to its hermetic, non-embodied prefigurations of the world. Among the more enduring are Hubert Dreyfus’s 1965 paper for the RAND corporation, “Alchemy and AI,” and his later volume What Computers Can’t Do: The Limits of Artificial Reason (New York: Harper and Row, 1979). The question of the relationship between mind, knowing, experience, and embodiment is, needless to say, a perennial philosophical concern, and the question of disembodied intelligence is but one issue raised in relation to the proficiencies of machine learning.

3. In “The Uncanny,” Freud describes an event where, having been confronted by his own reflection in a mirror and the prospect of someone mistakenly entering his private train compartment, he recalls not only being aghast at the sight of this “intruder” but repulsed by the “vestigial trace of the archaic reaction which feels the ‘double’ to be something uncanny.” For Freud, the effect, or affect, of doubling is consistent with a visual specter—a phantasmal presence that is accelerated by processes of mechanization and automation. Freud, “The Uncanny,” 371.

4. Quoting German philosopher F. W. J. Schelling’s Philosophy of Mythology (1857), Freud notes that the uncanny can be understood as “the name for everything that ought to have remained (…) secret and hidden but has come to light.” Freud, “The Uncanny,” 345.

5. The process of delegating sight and perception to machines (cameras, in particular) was addressed by, among others, Vilém Flusser (1920–1991). Considering how the concept of human agency was being increasingly regulated, Flusser argued that the “technical image”—which he viewed as a nascent regime of image production— was effectively produced by apparatuses rather than humans. The outcome of sequenced, recursive computations, the technical image heralded the demise of human-centric activities in models of image production. Through the technical image, the “original terms human and apparatus are reversed, and human beings operate as a function of the apparatus.” Vilém Flusser, Into the Universe of Technical Images (1985; repr., Minneapolis, MN: University of Minnesota Press, 2011), 74. Emphasis in original.

6. For an in-depth discussion of how AI image-processing models, alongside the ImageNet database, reify political, racial, and gender-based bias, see “Algorithmic Anxieties: Trevor Paglen in conversation with Anthony Downey,” Digital War 1 (2020): 18–28. http://www.anthonydowney.com/wp-content/uploads/2023/03/00_Algorithmic-Anxieties_Paglen-Downey-OnlinePDF-copy.pdf.

7. See page 122.

8. I am drawing here upon the Latin root of the term “adumbrate”—namely, umbra or shadow—and the way in which it describes the event of giving an outline or form to an object, through foreshadowing, and also the fact of casting a shadow upon it.

9. See page 120.

10. Throughout the Adversarially Evolved Hallucinations series, Paglen used "corpus" as another term for dataset, the latter being a collection of image categories. Capitalized throughout for the sake of clarity, the corpuses/datasets that make up the individual taxonomies for the series include OMENS AND PORTENTS; THE INTERPRETATION OF DREAMS; AMERICAN PREDATORS; EYE MACHINE; THE AFTERMATH OF THE FIRST SMART WAR; MONSTERS OF CAPITALISM; THE HUMANS; THINGS THAT EXIST NEGATIVELY; FROM THE DEPTHS; KNIGHT, DEATH, AND THE DEVIL; SPHERES OF HEAVEN; SPHERES OF PURGATORY; and SPHERES OF HELL.

11. According to Paglen, his choice of image categories for each dataset/ taxonomy are a direct critique of the "Linnaean taxonomies used in machine learning models" (See p. 120). In his discerning analysis of Adversarially Evolved Hallucinations, Luke Skrebowksi has proposed that the artist "deliberately scrambles taxonomic categories and in so doing denaturalizes and destabilizes the notion of taxonomy as a framework within Enlightenment modernity as carried across into computer science and computer vision." Luke Skrebowski, "Trevor Paglen's Adversarially Evolved Hallucinations: Computer Vision, GAN Photomontage, and the Displacement of Photography's Liquid Intelligence," in Marcel Finke and Kassandra Nakas, eds., Materials in Motion (Berlin: Dietrich Reimer Verlag, 2022), 145–64 (151).

12. See page 118.

13. For an engaging discussion of Adversarially Evolved Hallucinations and facial recognition technologies, see Lila Lee-Morrison, Portraits of Automated Facial Recognition: On Machinic Ways of Seeing the Face (Bielefeld: transcript Verlag, 2019), 159–75.

14. Observing the automation of perception through the use of artificial neural networks, Matteo Pasquinelli and Vladan Joler have argued that AI models of image-processing represents a new "cultural technique": "What a neural network computes is not an exact pattern but the statistical distribution of a pattern. Just scraping the surface of the anthropomorphic marketing of AI, one finds another technical and cultural object that needs examination: the statistical model. What is the statistical model in machine learning? How is it calculated? What is the relationship between a statistical model and human cognition?" See Matteo Pasquinelli and Vladan Joler, "The Nooscope Manifested: AI as Instrument of Knowledge Extractivism," AI & Society 36 (2020): 1263–80.

15. Louise Amoore, Cloud Ethics: Algorithms and the Attributes of Ourselves and Others (Durham, NC: Duke University Press, 2020), 16.

16. Paul Scharre, Army of None: Autonomous Weapons and the Future of War (New York: W. W. Norton & Company, 2018), 182. In one particularly germane example of this "brittleness," a state-of-the-art image-recognition neural network was shown images that to the human eye resembled nothing more than "white noise" or static, but were nonetheless identified as an "armadillo" and a "cheetah" with a 99.6 percent certainty by the AI image-processing model. See Scharre, 180–88.

17. The Inceptionv3 image- identification model is a pretrained convolutional neural network, or CNN, that is forty-eight layers deep. Although it operates differently from a

GAN, both use neural networks that display a recurrent "brittleness" that regularly generates levels of hallucination.

18. Anish Athalye et al., "Synthesizing Robust Adversarial Examples" (paper, Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, June 7, 2018), 19. https://arxiv.org/pdf/1707.07397.pdf

19. The issue of hallucinations in AI models such as ChatGPT has become a key source of concern for both programmers and users alike.Generative pre-trained transformers (GPTs) are instructed on large language models (LLMs), which use neural networks—specifically deep neural networks—to predict and project outcomes (in this case, text-based outputs). Although an AI system such as ChatGPT is understood to have hallucinated when it produces inaccurate information or specious answers, such outputs can appear all too plausible.

20. Nicolas Kayser-Bril, "Google Apologizes after Its Vision AI Produced Racist Results," Algorithm Watch (April 7, 2020). https://algorithmwatch.org/en/google-vision-racism/.

21. Dan McQuillan has argued that the statistical rationalization of data reveals a system susceptible to generating probabilistic simulations (forecasts) based on the transformation (modulation) of data: "Let's say we are dealing with a video: each pixel in a frame is represented by a value for red, green, and blue and the video is really a stack of these frames. So, when representing the video as numbers, the input into the algorithm is a huge, multidimensional block of data. As the input is passed through a deep learning network, the successive layers enact statistically driven distortions and transformations of the data, as the model tries to distil the latent information into output predictions." Dan McQuillan, Resisting AI: An Anti-fascist Approach to Artificial Intelligence (Bristol: Bristol University Press, 2022), 19.

22. Ian J. Goodfellow et al., "Generative Adversarial Networks," Communications of the ACM 63, vol. 11 (November 2020): 139–44.

23. Writing in the 2014 paper that announced the discovery of GANs, the authors outlined this adversarial relationship in precisely such terms: "The generative model [generator] can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model [discriminator] is analogous to the police, trying to detect the counterfeit currency. Competition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles." See Goodfellow et al, "Generative Adversarial Networks," 1–2.

24. Autopoiesis is a term used to describe a self-sustaining and self- replicating system that can produce and organize its own components, allowing it to subsist and sustainably adapt to its environment. The concept was introduced by biologists Humberto Maturana and Francisco Varela to explain the self-maintenance and self-reproduction observed in living organisms. Their theory was subsequently adopted in the context of cybernetics and, in turn, the development of AI and machine learning. See Humberto Maturana and Francisco Varela, Autopoiesis and Cognition: The Realization of the Living, Boston Studies in the Philosophy of Science 42 (1972; repr., Boston: D. Reidel Publishing, 1980).

25. For a fuller discussion of what a pixel looks like, see Alvy Ray Smith, A Biography of the Pixel (Cambridge, MA: MIT Press, 2021). Smith affirms that pixels are not square as such but rather more "blob-like" and have no shape until they are "spread" by a so-called pixel spreader—the latter being a formal filtering mode of reconstructing waves of light for the purpose of computational display.

26. Neural networks are, as one paper would have it, easy to fool insofar as "it is easy to produce images that are completely unrecognizable to humans, but that state-of-the-art DNNs [deep neural networks] believe to be recognizable objects with over 99% confidence (e.g. labeling with certainty that TV static is a motorcycle)." See Anh Nguyen, Jason Yosinski, and Jeff Clune, "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images" (paper, Conference on Computer Vision and Pattern Recognition, 2015), 427–36.

27. It is common in contemporary models of machine learning and computer vision for multiple algorithms to be used, and the combination is usually defined by the task in hand. The event of machinic seeing could involve the use of filtering algorithms (that smooth or sharpen images), edge-detection algorithms (for the identification of edges and boundaries), color-processing algorithms (for adjusting color balance, saturation, and contrast), or segmentation algorithms (for the division of an image into zones based on pixel intensity or for grouping adjacent pixels with comparable properties).

28. For an insightful account of algorithmic violence as a "force of computation," see Rocco Bellanova et al., "Toward a Critique of Algorithmic Violence," International Political Sociology 15, vol. 1 (March 2021): 123.

29. See page 125.

30. See page 127.

31. Somewhat worryingly, as Scharre has noted, this "bizarre [vulnerability] that humans lack" raises considerable doubts about the "wisdom of using the current class of visual-object-recognition AIs for military applications." Scharre, Army of None, 182. For a fuller discussion of these vulnerabilities in computer vision and the implications for aerial warfare, drone technologies, and lethal autonomous weapon systems, see Anthony Downey, "Algorithmic Predictions and Pre- emptive Violence: Artificial Intelligence and the Future of Unmanned Aerial Systems," Digital War 5, vols. 1–2 (2024). https://link.springer.com/article/10.1057/s42984-023-00068-7.

32. Taina Bucher, If … Then: Algorithmic Power and Politics (Oxford: Oxford University Press, 2018), 3. Emphasis added.