THE
MIT PRESS
READER

Uncanny Returns: Trevor Paglen and the Hallucinatory Domain of Generative AI

Do AI-generated images have the capacity to further estrange, if not profoundly alienate, us from the world?
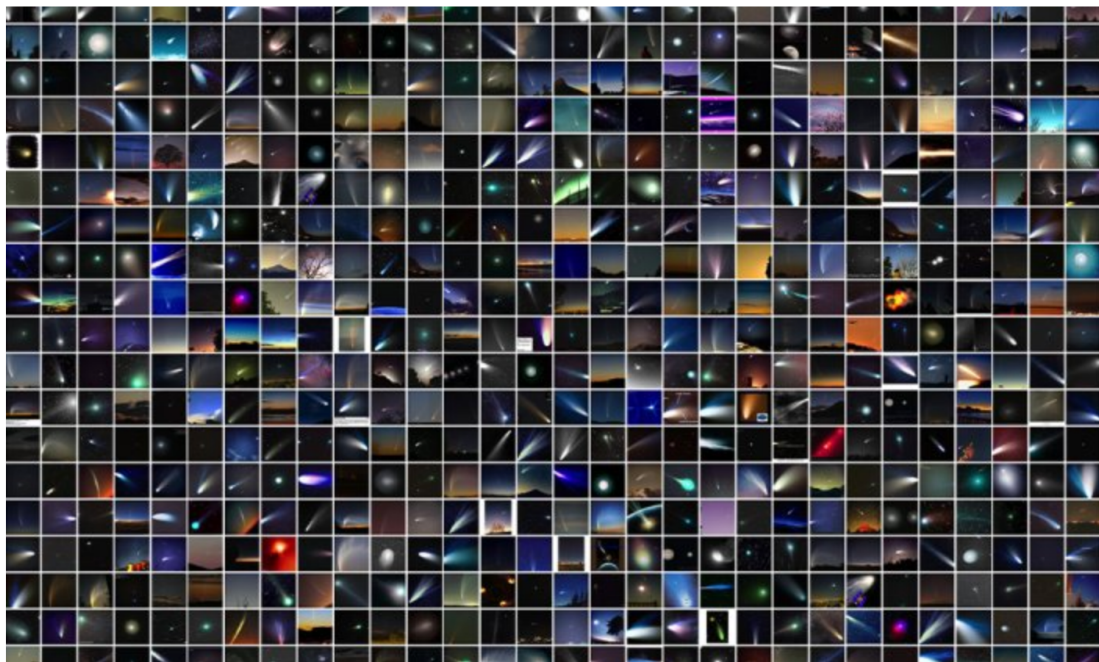
*Dataset relating to "Corpus: Omens and Portents" (Trevor Paglen)*

I f an image could be described as baleful, Trevor Paglen's "Rainbow" would fit the bill. Apart from the toxic-looking "sky," parts of it appear to have mutated into the fiery trace of munitions or, more cryptically, a sequence of glitches. The full title of the work, "Rainbow (Corpus: Omens and Portents)," suggests the collation of natural elements and a physical, possibly dead body (*corpus/corpse*), further bolstering the overall impression of estrangement and trepidation.

Anthony Downey is the editor of "Trevor Paglen: Adversarially Evolved Hallucinations" in Sternberg Press's Research/Practice series.
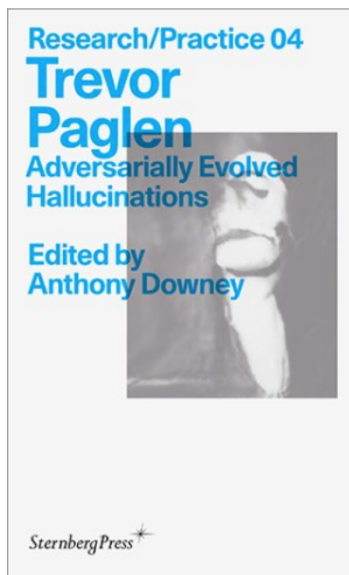
Along with other works in Paglen's *Adversarially Evolved Hallucinations* series (2017–ongoing), including the monstrous "Vampire (Corpus: Monsters of Capitalism)" and "Human Eyes (Corpus: The Humans)" — the latter complete with the apparition of deceptively unseeing eyes — "Rainbow" was produced by a generative adversarial network (GAN), an AI model that trains neural networks to recognize, classify and, crucially, generate new images. Given that AI image-processing models do not experience the world as we do, but rather replicate a once-removed and askew version of it, the images they produce reveal the degree to which AI computationally generates disquieting allegories of our world. Emerging from an embryonic space of automated image production, images such as "Rainbow" disclose that which is usually hidden or otherwise obscured — an apparition, or a nightmare, that is indebted to the hallucinatory, often erroneous logic of the algorithms that power AI. Often dismissed as a fault or glitch, this logic is nevertheless a fundamental aspect of AI, not merely a side effect. All of which leads us to an over-arching question: Do these hallucinatory models of AI-powered image production have the capacity to further estrange, if not profoundly alienate, us from the world?

"Rainbow (Corpus: Omens and Portents)" / courtesy of Trevor Paglen

An acclaimed artist, and a recipient of a MacArthur Fellowship, Trevor Paglen has consistently engaged with the invisible and occluded in our world, including, in his words, "the invisible visual culture" of machine-made images that remain indecipherable to the human eye.

In his 2019 exhibition "From 'Apple' to 'Anomaly,'" for example, he highlighted how algorithms — as coded sets of instructions — habitually give rise to biased models of AI. Throughout the project, he presented audiences and researchers alike with an opportunity to understand the extent to which contemporary subjects are being predefined and established through the biases — racial and otherwise — of repeatedly opaque models of machine learning, despite their questionable assumptions and design. And in his *Adversarially Evolved Hallucinations* series, which is likewise concerned with latent and clandestine elements, Paglen engages in a method of diagnostic reverse engineering: Working backward from the manifest, final iteration of an image such as "Rainbow," he explores how — through compiling datasets — the systematic training of a GAN is designed to produce new, as-yet-unseen, images. In so doing, the series demonstrates how these procedures are influenced by biases in the datasets (whereby certain images are over- or under-represented), discrepancies in overall training processes, and, notably, the often-opaque adjustments made during the iterative application of algorithmic weights to input data.

Despite the limitations and inherently biased nature of machine learning (not to mention the tendency to hallucinate), we are increasingly relying on these technologies to predict and influence our lives. Through the statistical analysis of past patterns, conducted in order to calculate and recognize future patterns, mechanistic predictions of people's identities, shopping preferences, credit ratings, career prospects, day-to-day behavior, health status, political affiliations, and supposed susceptibility to radicalization, become the norm rather than the exception. This predictive inclination of AI technologies, to this end, runs the danger of becoming in time both self-fulfilling and unaccountable, if not unfathomable.

"Vampire (Corpus: Monsters of Capitalism)" / courtesy of Trevor Paglen

The implications of machine-driven calculations are profound, especially when they attempt to predict not only everyday behaviors but, perhaps more disturbingly, those that fall outside the norm — behaviors, activities, and identities that resist predefined patterns. By exposing the deterministic reasoning behind AI models of image production — this is a rainbow; this is an apple; this is a face; this is a threat — and how they impose meaning upon the world, the seemingly abstract act of (mis)classification, or hallucination, has an all-too-real bearing on how we live our lives. The legacy of this imposition, its epistemological *affect*, could not be more significant: When deployed in facial recognition technologies, for instance, such systems assign a classification to a particular object or entity — a face — and allocate a name or, more ominously, a level of threat to it.

"Human Eyes (Corpus: The Humans)" / courtesy of Trevor Paglen

Consequently, there often exists a concomitant tendency to take these categorizations for granted and proceed — that is, flag up certain patterns of behavior — in accordance with their forecasts and automated recommendations. In programmatically presenting the world through the computational inferences of neural networks, AI models of image-processing, such as a GAN, are programming, or regulating, people and communities to accept machinic conjecture. We are, in short, being primed to consider computational supposition as *the* "truth" of our world rather than, as it in fact is, a conditional projection based on probabilistic predictions and the opaque workings of an algorithm.

The operative logic of a GAN is, importantly, specifically geared toward generating new, as-yet-unseen images, which further subjects the procedure to a significant degree of computational fortuitousness. Despite the sense of technological determinism often associated with algorithmic devices (the supposedly accurate identification of people and debatable predictions of future events), the procedures involved do not *automatically* yield predictable or, indeed, correct outcomes.

We are, in short, being primed to consider computational supposition as the "truth" of our world.

Having already been assigned a definitive category — rainbow, eyes, vampire — by a GAN, this point is central to the images we encounter throughout the *Adversarially Evolved Hallucinations* series, where 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 suspect form of "machine realism." Creating a training set involves, he has observed, the categorization and classification, by human operators, of thousands of images. Thereafter, Paglen continues, there "is an assumption that those categories, alongside the images contained in them, correspond to things out there in the world."

Inherent within this machinic reality we find a disturbing surrogate of our world, one in which we need to confront the fact that the neural networks and deep-learning models involved in training machines to see are not only systematically prone to error, they are also systemically susceptible to hallucinating objects that do not exist. Insofar as image-processing systems do not return exact replicas or accurate classifications of the world, they regularly hallucinate realities into being. It is this predisposition that Paglen amplifies when he intervenes in the latent spheres of algorithmic reasoning — those hidden layers within neural networks where AI makes its classifications and associations. It is here, where images return to us in uncanny variations on a theme, that algorithms can be tweaked and discretely weighted by a programmer — or, as is the case here, a programmer-cum-artist such as Paglen. In heightening the potential for hallucination in a GAN, the common applications of algorithms (image identification) can be critically disconnected from their utilitarian function and revealed for what they are: statistical approximations and mechanical allegories of reality.
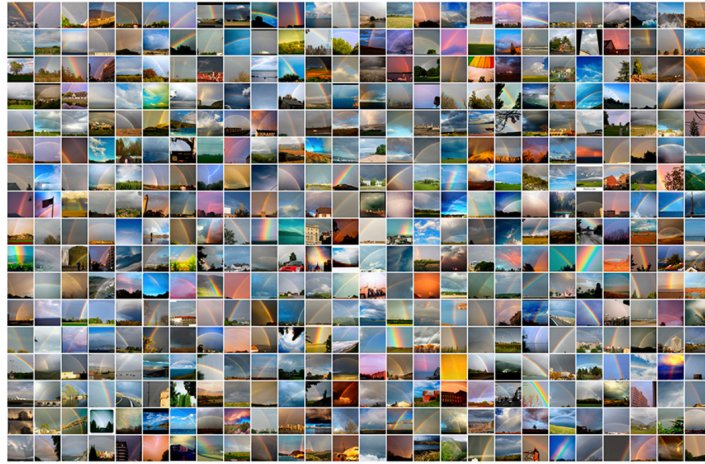
Selection of images from datasets relating to "Rainbow (Corpus: Omens and Portents)"

In illuminating the computational delirium that drives generative AI, Paglen's practice mines, exploits, and contradicts the frequently inflated claims made in relation to the effectiveness of neural networks in image classification tasks. If "Rainbow" is indeed an uncanny analogy of a rainbow, summoned forth by algorithmic reasoning, how then do we understand the procedures through which neural networks arrive at such hallucinatory images? How, Paglen asks, do we think from *within* these systems rather than merely reflect upon their potential import?

The accumulative and ascendant influence of AI on our lives and how we live presents a strong argument for developing research methods, such as those deployed by Paglen, designed to encourage a critical range of thinking from *within* the apparatus of AI. Through developing such methodologies, we can ensure that the systematic training (involved in labeling and inputting data, for example) and systemic (latent and algorithmic) spaces of computation are more readily understood for what they are: statistical calculations of probability, rather than certainty, that have nonetheless come to define significant aspects of how we perceive and live in the world today.

__Anthony Downey__ is _Professor of Visual Culture_ in the Middle East and North Africa (Birmingham City University). He sits on the editorial boards of _Third Text_, _Digital War_, and _Memory, Mind & Media_, and recently edited "_Trevor Paglen: Adversarially Evolved Hallucinations_" (Sternberg Press, 2024) for the _Research/Practice_ series.