CHAPTER
TWO

witnessing algorithms

LAUNCHED IN AUGUST 2020, the latest edition of the venerable Microsoft Flight Simulator video game series offered an open-ended experience of a world made suddenly inaccessible by COVID-19. Unlike its predecessors, MS Flight Simulator 2020 makes the entire planet its gameplay environment. In the hyperbole typical of much of the media coverage, *New York Times* tech columnist Farhad Manjoo proclaimed that Microsoft had "created a virtual representation of Earth so realistic that nearly all sense of abstraction falls away."[1] As a technical achievement, Flight Simulator is certainly impressive. Combining data from OpenStreetMaps and Bing Maps via the Azure artificial intelligence cloud computing platform, Microsoft created an algorithmic system to assign and render photorealistic 3D imagery of skyscrapers, homes, trees, oceans, mountains, and so on. This imaging of the world is not, however, photographic but datalogical: generated algorithmically by a machine learning model fed vast amounts of map, satellite, photogrammetric, and other data. It is a machinic imagining of the textures of the world. Like Google Earth, it is a datalogical attempt at solving the fundamental problem that plagues the unusable map from Borges's short story "On Exactitude in Science," which in the effort to precisely represent every detail of an empire grows to the same size as the territory. Rather than indexing its cartography to the world perceived by human mapmakers, Flight Simulator generates what its algorithms believe the world to be. Players quickly found numerous strange glitches: a corporate office tower in place of the Washington Monument, a mashup of vegetation and buildings in the Norwegian town of Bergen, obelisks in place of palm trees. Far from a utopian rendering of a world made beautiful yet knowable, Kyle Chayka writes in *Slate* that Flight Simulator reminds us that "an automated, unchecked process is warping the (virtual) world around us, leading to these weird errors and aberrations."[2] Even as Flight Simulator seemed to offer a new algorithmic means of witnessing in wonder at the world, its glitches remind us of the necessity of witnessing those same algorithmic systems. If algorithms are themselves witnessing, making knowledge, and forging worlds of their own design, what might it mean to witness their workings?

The world-making capacity of the algorithm is not readily apparent in its more common definitions: a step-by-step instruction of how to solve a task; a recipe; a form of programmed logic; an automated filtering mechanism. These commonplace accounts fail to get to the heart of things, the operative processing made possible by the "if . . . then" procedure of the algorithm and its potentially harmful outcomes.[3] In principle, algorithms are abstract processes, which means they are not dependent on a specific computer language for their validity. But in practice, algorithms are typically encoded in distinct computer languages and ecosystems. More than this, though, they are also inescapably *codes* in the sense that they unlock certain translations, operations, or transformations of data.[4] We might even think of them as *magic* in the sense that the incantation of the algorithm by the software within which it is packaged enables action to be performed. Like codes and magic, algorithms conceal their own operations: they remain mysterious, including to their makers. This inscrutability is particularly the case with the machine learning algorithms that have become the principal means by which power is now enacted, maintained, and reproduced in the digital domain.

Machine learning is a technique for the statistical analysis of huge quantities of data. A machine learner is an algorithmic system in which computer code learns from data to produce a model that can be deployed on more data. Machine learning produces models by using algorithmic techniques to look for patterns in huge amounts of data, then applying those patterns to the data to become increasingly discerning: able to recognize, differentiate, and discriminate between elements within the database. Machine learning powers everything from inbox filtering to Netflix recommendations and it feeds on the data produced through our interactions with those systems. Machine learning systems and the companies that promote them almost always seek to obscure both the “free labor” of user interactions and the low-paid labor of digital pieceworkers on platforms such as Mechanical Turk in an effort to sell the technical prowess of their “AI” inventions. Machine learning uses layers of neural networks—arrays of computational nodes that work collaboratively to build relations between bits of data—to make predictions about the data. With OpenAI’s ChatGPT, this manifests as the statistical production of text based on what the model anticipates to be the desired answer to a query. In military operations, it might mean identifying and prioritizing distinct threats in a crowded conflict zone. Rather than following a defined sequence of steps, machine learning models act recursively to build relational functions that can be applied ever-more accurately and efficiently, provided the learner is trained and optimized appropriately.[5]

But this technicity is not purely technical. As Adrian Mackenzie points out, there are no machine learning systems without human coders and humans are also needed to tag objects in the datasets for the supervised training by which many machines learn.[6] In so-called unsupervised learning, algorithms develop their own data tags, but human effort is still constantly required to tweak, select, optimize, and monitor training. Jathan Sadowski calls this “Potemkin AI,” or artificial intelligence that is actually only thinly computational and largely driven by human labor.[7] On top of the obscured human labor, Sy Taffel shows how computational systems also elide massive ecological costs of powering and cooling data centers, not to mention mining rare and common metals or shipping equipment across the globe.[8] To bring machine learning into the language of this book, its models and algorithms are not alien, purely technical agents wholly separate from the human, but rather enmeshed with the human and with the more-than-human world. How machine learners make knowledge matters because they are increasingly pivotal to contemporary finance, logistics, science, governance, national security, and culture, yet they remain hard to scrutinize, building blocks in what Frank Pasquale calls the “black box society.”[9]

Despite their technical veneer, algorithms are shaped and bound by assumptions and values about the world, drawn from the datasets upon which they are built, the biases of their architects, and the instrumental objectives of the institutions that use them. These assumptions and values might be as straightforward as whether to order library books by alphabet or catalog number, or as outrageously discriminatory as Facebook allowing housing advertisers to exclude users from target audiences using zip codes and other proxies for race, class, and religion. Given the colonial entanglements of modern science, regimes of classification, and the statistical techniques that underpin contemporary data science and machine learning, the constitutive violence of many such systems should come as no surprise. Algorithmic violence, whether in the form of digital redlining or autonomous weapons, is an ethicopolitical problem much more than a technical one.[10] As Ed Finn points out, the algorithm is a crucial site of critical inquiry because it is "the object at the intersection of computational space, cultural systems and human cognition."[11] Traceable back to the cybernetic era of computational research that followed World War II, algorithms were at the center of a radical transformation that substituted rationality for reason. Within two decades of the war, as Orit Halpern argues, "the centrality of reason as a tool to model human behavior, subjectivity, and society had been replaced with a new set of discourses and methods that made 'algorithm' and 'love' speakable in the same sentence and that explicitly correlated psychotic perspective with analytic logic."[12]

Now deployed across almost every field of human endeavor and inquiry, algorithms bridge the gap between computation, culture, and thought—but they are not reducible to any of those domains. According to Taina Bucher, algorithms are "entangled, multiple, and eventful and, therefore, things that cannot be understood as being powerful in one way only."[13] Consequently, "algorithmic systems embody an ensemble of strategies, where power is immanent in the field of action and situation in question."[14] Research by Safiya Noble and others into the oppressive biases of Google and Facebook shows how supposedly objective systems are inseparable from racism, sexism, and other socially produced and reproduced structures of domination.[15] Generative AI tools such as Dall-E 2 or Midjourney are no exception, evidenced by the efforts of their architects to engineer user inquiries rather than resolve the impossible problem of the underlying data.[16] As Nick Seaver argues, "algorithms are not singular technical objects that enter into many different cultural interactions, but are rather unstable objects, culturally enacted by the practices people use to engage with them."[17] Much like a poem, algorithms are tricky objects to know and often cannot even reveal their own workings.[18] Critical research thus attends less to what an algorithm *is* and more to what it *does*.[19]

In pursuing nonhuman witnessing of, by, and through algorithms, my focus is on their operative, extractive, and generative qualities, rather than their computational mechanics. Through a series of investigations into distinct machine learning systems, I argue that algorithms can engage in a perceptual process that constitutes nonhuman witnessing, elevating mere observation into an ethicopolitical plane. In drone warfare, an algorithm might "see" certain activity, "decide" it is threatening, and "recommend" the prosecution of violence. My contention is that such algorithmic registering and translating of worldly phenomena constitutes witnessing because it does so to violent ends and caries the most immediate traces of that violence. Facial recognition software is a tool for producing evidence through machinic witnessing, yet both the data that feeds such systems and the unknowable neural dynamics that animate them make it so dangerous that facial recognition has been described as akin to plutonium.[20] Algorithmic witnessing, then, often takes place through the enactment of violence, with the algorithm as both witness and perpetrator. At the same time, such algorithms are themselves entities that must be witnessed—yet by their entangled nature they resist being broken into consistent elements that can then be rendered knowable.

This chapter grapples with the competing dynamics of the doubled meaning of its title: algorithms that do witnessing and the witnessing of algorithms (and what they do). Or, to put this differently, this chapter asks both how algorithms might be agents *of witnessing* and how algorithms might *be witnessed*? Rather than look for machinic relations to events that might be analogous to human witnessing, this chapter seeks out intensive sites within human-nonhuman assemblages where machinic affect—technical yet contingent, potential rather than predetermined—enables forms of encounter that generate a relation of responsibility between event and algorithm. Doing so requires the bracketing of any ethical imperative to witnessing: algorithmic witnessing can only ever be grasped within the milieu of the algorithm, an agency that can only be ascribed ethics or morality through anthropomorphic gymnastics. Delving into the machinic affects of witnessing algorithms will require us to depart further still from the narrow humanistic conception of witnessing and to insist on the separation of witnessing from testimony. If algorithmic technologies are now crucial knowledge machines, yoked to power, and the infliction of state violence, then asking how witnessing reckons with them and takes place through them requires attending to how computational processes generate their machinic relations, and how those relations sustain the power of those systems.

Even as the image increasingly overwhelms the word as the dominant form of communication, the expansion of technologies that identify and organize images means that a new form of aggregated, relational perception is taking hold. Writing on the aggregation of huge numbers of images into datasets analyzed by machine learning systems, Adrian Mackenzie and Anna

Munster understand these relational processes as "generative technical forces of experience."[21] They propose the concept of "platform seeing" to describe an operative mode of perception "produced through the distributed events and technocultural processes performed by, on and as image collections are engaged by deep learning assemblages."[22] In their account, "seeing" is not the act of a singular entity but rather something that takes place across a great many human, material, and computational agents. Images become subject to a host of functions: precisely formatted for input to models; labeled, processed, and used to configure small neural networks onboard smartphones; moved from the devices of consumers to platforms and their data centers and back. Through these and other functions, images transformed from bearers of indexical relations to elements within operational (image) collections.[23] Consequently, the relations between images within the dataset, including the relations of elements within images to elements within others, become more important than the images themselves.

Platform seeing is thus the "making operative of the visual by platforms themselves."[24] This *invisual* mode of perception takes place outside the domain of representation: images no longer take their meaning from things in the world but rather in relation to the elements and edges of other images. Crucially, this "operativity cannot be seen by an observing 'subject' but rather is enacted via observation events distributed throughout and across devices, hardware, human agents and artificial networked architectures such as deep learning networks."[25] Despite the absence of a human subject, these processes still constitute something called "seeing" precisely because they remain within the perceptual domain of recognizing and differentiating images. In this chapter, I make a parallel argument about witnessing: that even without a witnessing "subject" in the unitary humanist sense, witnessing occurs within and through algorithmic systems. Such witnessing necessarily exists on a continuum with perceiving and cannot be neatly distinguished from it. Different contexts, media technics, and human entanglements produce distinct intensive fields of relation that shift perceiving into the modality of witnessing. Not all human perception entails witnessing, and so neither does all perception by the nonhuman agencies of algorithms.

While witnessing rarely figures in discussions of algorithms and artificial intelligence, terms that appear in witnessing discourses abound: truth, recognition, memory, transparency, ethics. This is not to suggest an inherent synchronicity between witnessing and the algorithmic, but rather to point out that the perception required for both to operate possesses a purposive dimension. As Amoore writes, "A defining ethical problem of the algorithm concerns not primarily the power to see, to collect, or to survey a vast data landscape, but the power to perceive and distill something for action."[26] In much the same way, witnessing is not reducible to seeing, but is a perceptual encounter that produces an injunction to action through its configuring of a particular scene and its coalescing of that scene's relational dynamics. Like algorithms, witnessing makes truth claims about the world as well, and is also prone to oversight, misapprehension, and misstatement.[27] Like algorithms, witnessing is prone to falsity, whether deliberate or accidental. Their distributed, multiple, contingent, and operative existence means that algorithms cannot be known or accounted, and yet neither can *the human*. It is only ever humans, plural, who can give account, and doing so is always incomplete. This is why, for Amoore, "algorithms do not bring new problems of blackboxed opacity and partiality, but they illuminate the already present problem of locating a clearsighted account in a knowable human subject."[28] Neither human nor algorithm can give an account of itself that is complete or transparent. An ethics for algorithmic worlds cannot "seek the grounds of a unified *I*" but must instead "dwell uncertainly with the difficulty of a distributed and composite form of being."[29] This chapter pursues the question of what distributed, opaque, and decentered witnessing might look like within technics of the algorithm—and how such a contingent and multiple domain might itself be witnessed.

Crucial to that task is tracing what I call *machinic affect*, or the intensive relations that bind technical systems to one another and humans to technical systems. By machinic affect, I mean the capacity to affect and be affected that occurs within, through, and in contact with nonhuman technics. In keeping with Félix Guattari's expansive conception, my own use of "machinic" is not restricted to the mechanistic but rather refers to the processual assemblage of elements, objects, concepts, imaginaries, materialities, and so on that form "machines" through their distinct yet transversal relations. Guattari's machines are organic and inorganic, technical and social, material and abstract.[30] Machinic affect is not so much indifferent to the flesh as it is promiscuous in its adhesive and intensifying properties, such that the corporeality of the human does not default to center stage.[31] Excavating machinic affect from technical assemblages requires attending to the distinctiveness of individual technical objects as they assemble, attenuate, modulate, amplify, and terminate technical and nontechnical relations. In the context of witnessing, machinic affect can be applied to understand the relations forged between witness and event when mediated through screens. But more importantly and generatively, machinic affect offers an analytic for making visible the otherwise obscured machinic relations of complex technical systems and especially learning algorithms.

Machinic affect describes the dynamic intensities of technical systems. As such, machinic affect is autonomous intensity: owned neither by one body nor another, but constituting and constituted by them, whether human or non. Pursuing machinic affect within the media-specificities of algorithmic systems, I am interested in how the processual empiricism of what Massumi calls the "virtual" illuminates the relational dynamics of machine learning. For Massumi, the virtual describes the immanence of potentiality, its passage from futurity through experience and into pastness. The virtual is what might arise and what might have been. It is not opposed to the actual, but its underside. Affect is "precisely this two-sidedness, the simultaneous participation of the virtual in the actual and the actual in the virtual, as one arises from and returns to the other."[32] This chapter is about the necessity of witnessing how algorithms, particularly machine learning ones, oscillate between actualizing the virtual and virtualizing the actual.

If algorithmic systems are about taming potential into probability in the name of emergent ordering of worldly phenomena, we can understand them in Massumi's terms as ontopowerful: as technological processes for the mastery of becoming.[33] Machine learning systems are constituted by unreason—even madness—through looping recursivities.[34] This nonlinearity, too, finds much in common with Massumi's recognition that "intensity would seem to be associated with nonlinear processes: resonation and feedback that momentarily suspend the linear progress of the narrative present from past to future."[35] As well as disassembling and distributing the subject, witnessing algorithms requires dismantling and dispersing the event in time as it is taken up and worked upon by algorithmic agencies. This chapter thus excavates the distinctive dynamics of nonhuman witnessing across four instances of algorithmic world-making: the false witnessing of deepfakes; the animating of evidence in Forensic Architecture's machine learning investigations; military imaginings of archival and real-time processing of full motion video imagery from loitering drones; and the witnessing of machine learning processes in aesthetic interventions into algorithmic systems. Operating with different learning models and data sources and within very varied contexts, these examples show the dangers of algorithmic witnessing and the necessity of witnessing algorithms, but they also suggest the potential of such systems to work against state and corporate violence.

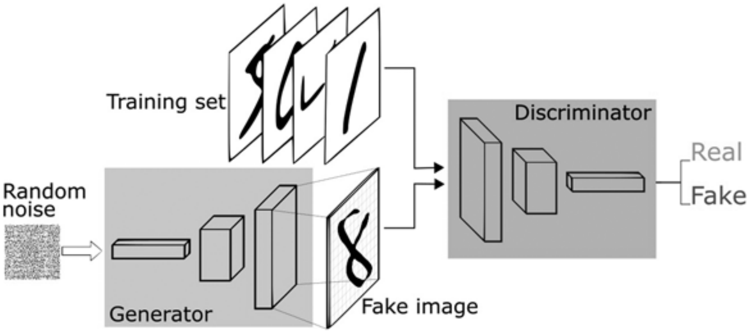
The synthetic media that would become known as "deepfakes" first surfaced to mainstream attention with a December 2017 article by Samantha Cole in Vice Media's tech site Motherboard about a pornographic video that appeared to feature the actress Gal Gadot having sex with her stepbrother (figure 2.1). As Cole reported on the tech site, the video was a fake, the clever but decidedly imperfect creation of a Reddit user with some basic machine learning skills and open-access tools downloaded from the code repository GitHub.[36] Fake and face-swapped pornography are not new phenomena: CGI porn is widely available, while photoshopped porn images have been around as long as the internet and altered nude photographs since the early twentieth century at least. The difference in the Gadot video was the application of deep learning techniques to automatically swap one face with another. That technique gave the Redditor his handle and the new genre a name: deepfakes. "With hundreds of face images, I can easily generate millions of distorted images to train the network," deepfakes told Cole. "After that if I feed the network someone else's face, the network will think it's just another distorted image and try to make it look like the training face."[37] With so many high-quality images on which to train the system available online, celebrities like Gadot are easy targets. But that same ease could readily apply to politicians, and to voice as well as video. Arriving amid a rising tide of distrust in systems and institutions, deepfakes seemed to herald a new threat, undermining democratic processes and cybersecurity and facilitating misinformation and revenge porn. A cottage industry of deepfake creation and detection sprung up in response. Deepfakes seemed to enable the bearing of algorithmic false witness—a problem only complicated by the arrival of more user-friendly artificial image and video generation tools in the years since.

While there are several techniques that can be used to generate deepfakes, the most effective are produced through a form of deep learning neural network called "generative adversarial networks," or GANS. While image recognition algorithms are typically trained using convolutional neural networks (CNNS) that slide filters across images to learn their spatial properties, GANS work by pitting two algorithms against each other in a game of true and false (figure 2.2). First proposed by AI researchers from Google Brain in a 2014 paper, the premise of GANS is simple enough: one neural network (the generator) learns to create images that it then feeds to another network (the discriminator), which decides if the image is "fake" or "real" compared to its own training dataset.[38] Those results are then fed back into the generator, so

FIGURE 2.1. Still from Gal Gadot deepfake porn

that it can learn from the assessment of the discriminator. What makes the technique powerful is that both networks are learning at the same time, with the discriminator learning just enough to get ahead of the generator each time the quality of its fakes catches up. To produce the Gal Gadot deepfake with a GAN, the generator would be fed the pornographic video while the discriminator learned from real photos of Gadot. As the generator modified its video using several image-blurring and blending techniques, proximity to what the discriminator was learning about Gadot would yield better

FIGURE 2.2. Diagram of general adversarial network structure



and better results until the discriminator could no longer identify the fake images as fake at all. If the GAN was then trained on other video and image sets, it would get even better at its task. In this way, GANs can become highly accomplished at swapping any face for any other. Versions of this technique can be applied to specific parts of the face, too, such as the lips, or to audio, enabling the falsification of someone's voice to match a script, as in the widely reported Obama lip-sync demonstration.[39]

While computer science papers have focused on deepfake creation and detection, the humanities and social sciences has begun to address a wider set of questions.[40] The most attentive examinations of deepfakes have occurred within porn studies, where the gendered nature of the technology in practice—more than 99 percent of documented deepfakes feature women face swapped into pornographic videos—has been documented and examined in a range of contexts, from revenge porn to communities of practice to the emergence of "designer" porn.[41] Legal scholarship within the United States has addressed how deepfakes create tension between rights to free speech and privacy, as well as how they present a potential crisis for the verification of evidence presented to court.[42] Possible impacts for cybersecurity and information warfare are articulated in more apocalyptic terms.[43] But deepfakes also point to the vibrancy of everyday data cultures, and the experimental, open-source approach to AI and automation literacies taking place on GitHub, YouTube, and Reddit.[44] "Deepfakes are complex epistemic things," observes Rebecca Uliasz, which "testify to the ongoing socio-technical value we place on visual accuracy which manifests in our continued investment in imagistic realism as truthful."[45] As such, deepfakes can be understood as yet another technological blow to shared epistemic frameworks, further undermining certainty in image authenticity for both journalists and publics.[46]

For WITNESS, a New York–based human rights organization that equips citizens and activists with video tools and resources, deepfakes and related forms of synthetic media are an urgent danger because they can amplify or microtarget the kinds of media disinformation and incitement that spark massacres, assaults, and political instability. In a report on synthetic media, WITNESS distills dozens of scenarios into five key problem areas: reality edits, credible doppelgängers, news remixing, floods of falsehood, and plausible deniability, in which claims of deepfaking allow bad faith actors to deny having said or done what a video shows.[47] Deepfakes lead to two interrelated epistemic challenges: "the inability to show that something real is real, and then the ability to fake something as if it was real."[48] For WITNESS, the inability to prove that something real is real presents the more serious dilemma because it suggests an existential peril for the mediated processes that are so essential to contemporary shared realities. This algorithmic false witness risks placing all mediated witnessing into question. Deepfake false witnessing cuts and intensifies preexisting risks to individuals and communities by catalyzing uncertainties within contemporary media ecologies. Over the last few years, WITNESS has worked with Partnership on AI to develop guidelines for appropriate use of synthetic media in human rights contexts. These are necessary and important practical steps, but the epistemic challenges of deepfakes and related media forms persist.

By threatening the legitimacy of the image, deepfakes destabilize the very foundations of media witnessing as a shared means of producing an agreed reality. Deepfakes emerged in a media witnessing ecology in which power has shifted from the authority of legacy media to the immediacy of smartphone and other user-generated content.[49] As the necessity of grounding truth claims becomes more urgent, deepfakes heighten the fallibility of witnessing in, by, and through media. These are fake images that make truth claims, even as they undermine the possibility of common epistemic ground.[50] In places with declining trust in government or with significant government instability and insecurity, deepfakes have the potential to incite violence and violate human rights. Weaponized deepfakes assembled on the fly from social media records are one nasty possibility for the future of what Tom Sear calls "xenowar."[51] If neither still nor moving images can be trusted to bear the indexical relationship to the world that their authority depends upon, the potential for any mediated witness to be false threatens to pry open the fractures already running through any sense of shared reality. With their emphasis on altering or swapping faces, deepfakes are affect machines even more than cognitive deceptions. Machinic affect here takes a very recognizable form in the micromovements of faciality described by Deleuze in his account of the affection image in cinema and by Silvan Tomkins in his theory of nine discrete relational affects manifested on the face.[52] Face, voice, and gesture are among the most crucial embodied qualities of bearing witness: deepfakes seek to synthesize both fake and real to affect the viewer. Created through the intensive interplay of machinic relations, deepfakes are also affect engines when loose in the wild. As false witnessing algorithms, deepfakes exemplify the inextricability of human and non in witnessing assemblages within today's deeply computational world.

Deepfakes are among the most unsettling instances of the shifting relationship between image and data. “An image that is computational is no longer strictly concerned with mimesis, nor even with signification,” writes Steven F. Anderson. “Computational images may serve as interfaces, carriers, or surface renderings, the real importance of which are their underlying processes or data structures.”[53] Deepfakes are syntheses of recorded and generated images made possible by the encoded nature of both. At the level of code itself, neither bears any more material relation to the world beyond computation than the other. Even as the image reaches its zenith in visual culture, the transfiguration into code that made its domination possible contains within it the collapse of the authority granted to the image by its seemingly indexical relation to the world. Ironically enough given their origins in DIY porn communities, deepfakes speak to how “the once voyeuristic gaze of cinema has given way to power relations articulated through computational systems rather than through ocular regimes predicated on reflected light and bodies in space.”[54] The false witnessing of deepfakes suggests that contestations over the meaning of images is moving away from *signification* and into *generation*. For deepfakes and imagery produced by Stable Diffusion or Dall-E 2, contestation ceases to be about what the video image *means* and comes to be about the *process* of its generation.

This movement from semiosis to process means that the false witnessing of deepfakes must be contested at an ontological level. While early iterations had a tendency for glitching and an unsettling uncanny valley-like quality, advances in the deep learning processes of GANs now mean that humans can typically detect deepfakes about half the time, or at the same rate as random chance. Deepfake detection tools that draw on the same kind of deep learning neural networks have become increasingly important, but their emergence and growing accuracy has led to an arms race between forgers and detectors.[55] This formation of a new adversarial, nonhuman, and machinic relationship between witness and interrogator points to yet another site in which critical debates about culture, politics, ethics, and knowledge play out without the human in the driver’s seat. A potentially endless game of deception and unmasking awaits in which witnessing itself becomes the ground of contestation between adversarial machine learning systems and where social and political life become the field upon which the consequences of that struggle play out. Ironically enough, algorithmic false witnessing heightens computation’s claim as both figure and ground for how knowledge is produced and contested.

BETWEEN EVIDENCE AND WITNESSING: FORENSIC ARCHITECTURE AND OPEN-SOURCE MACHINE LEARNING

Synthetic media are everywhere, not just in deepfakes. Digital images and objects that appear to index something in the world but do nothing of the sort have their roots in video games and online worlds like Second Life. With the growing appetite for niche machine learning training sets and artificial environments for testing autonomous machines, synthetic media are increasingly central to the development of algorithmic systems that make meaningful decisions or undertake actions in physical environments. Synthetic media are swift to produce and can be tagged as part of the production process, which reduces costs, delays, and inaccuracies from using people to tag images or other data.

Microsoft AirSim is a prime example, an environment created in Epic's Unreal Engine that can be used to test autonomous vehicles, drones, and other devices that depend on computer vision for navigation.[56] Artificial environments are useful testing grounds because they are so precisely manipulable: trees can be bent to a specific wind factor, light adjusted, surface resistance altered. They are also faster and cheaper places to test and refine navigation software prior to expensive material prototyping and real-world testing. In machine learning, building synthetic training sets is now an established practice, particularly in instances of limited data availability or lack of data diversity. For example, the company Synthesis.ai produces synthetic images of nonwhite people to train various kinds of recognition algorithms. Synthetic media are valuable in contexts such as armed conflict, where images might be too few to produce a large enough corpus and too classified to be released to either digital pieceworkers for tagging or private sector developers to train algorithms.

But what happens when synthetic media are marshaled to do the activist work of witnessing state and corporate violence? What are we to make of the proposition that truths about the world might be produced via algorithms trained almost exclusively on synthetic data? This section sketches answers to these questions through an engagement with *Triple Chaser*, an investigative aesthetic project from the UK-based research agency Forensic Architecture. Founded in 2010 by architect and academic Eyal Weizman and located at Goldsmiths, University of London, Forensic Architecture invents investigative techniques using spatial, architectural, and situated methods. Using aesthetic practice to produce actionable forensic evidence, their work appears in galleries, courtrooms, and communities. In recent years, they have begun to use machine learning and synthetic media to overcome a lack of publicly available images on which to train their machine learning models and to multiply by several orders of magnitude the effectiveness of images collected by activists. In contrast to the false witnessing of deepfakes, these techniques show how algorithms can do the work of a more resistant and generative witnessing, translated into open-source tools for activists via well-documented GitHub pages.

Presented at the 2019 Whitney Biennial in New York, *Triple Chaser* combines photographic images and video with synthetic media to develop a dataset for a deep learning neural network able to recognize tear gas canisters used against civilians around the world. It responds to the controversy that engulfed the biennial following revelations that tear gas manufactured by Safariland, a company owned by Whitney trustee Warren B. Kanders, was used against protestors at the US-Mexican border. Public demonstrations and artist protests erupted, leading to significant negative press coverage across 2018 and 2019. Rather than withdraw, Forensic Architecture submitted an investigative piece that sought to demonstrate the potential for machine learning to function as an activist tool. Produced in concert with artist and filmmaker Laura Poitras, *Triple Chaser* was presented as an eleven-minute video installation. Framed by a placard explaining the controversy and Forensic Architecture's decision to remain in the exhibition, viewers entered a severe, dark room to watch a tightly focused account of Safariland, the problem of identifying tear gas manufacturers, the technical processes employed by the research agency, and its further applications. Despite initial intransigence, the withdrawal of eight artists in July 2019 pushed Kanders to resign as vice chairman of the museum and, later, announce that Safariland would sell off its chemicals division that produces tear gas and other antidissent weapons. Meanwhile, Forensic Architecture began to make its codes and image sets available for open-source download and began applying the same techniques to other cases, uploading its Mtriage tool and Model Zoo synthetic media database to the code repository GitHub. A truth-seeking tool trained on synthetic data, *Triple Chaser* reveals how machinic affects oscillate between witnessing and evidence.

In keeping with the established ethos of Forensic Architecture, *Triple Chaser* demonstrates how forensics—a practice heavily associated with both policing and the law—can be turned against the very state agencies that typically deploy its gaze. As Pugliese points out, "Embedded in the concept of forensic is a combination of rhetorical, performative, and narratological techniques" that can be deployed outside courts of law.[57] For Weizman, the fora of forensics is critical: it brings evidence into the domain of contestation in which politics happens. In his agency's counterforensic investigation into Safariland, tear gas deployed by police and security agencies becomes the subject of interrogation and re-presentation to the public.[58] In this making public, distinctions and overlaps can be traced between different modes of knowledge-making and address: the production of evidence, the speaking of testimony, and the witnessing of the audience. But how might we understand the role of the machine learning algorithm itself? And how are we to conceptualize this synthetic evidence?

Weizman describes the practice of forensic architecture as composing "evidence assemblages" from "different structures, infrastructures, objects, environments, actors and incidents."[59] There is an inherent tension between testimony and evidence that counterforensics as a resistant and activist practice seeks to harness by making the material speak in its own terms. As method, forensic architecture seeks a kind of "synthesis between testimony and evidence" that takes up the lessons of the forensic turn in human rights investigations to see testimony itself as a material practice as well as a linguistic one.[60] Barely detectable traces of violence can be marshaled through the forensic process to become material witnesses, or evidentiary entities. But evidence cannot speak for itself: it depends on the human witness. Evidence and testimony are closely linked notions, not least because both demarcate an object: speech spoken, matter marked. Testimony can, of course, be entered into evidence. But something more fundamental is at work in *Triple Chaser*. Its machine learning model doesn't simply register or represent. It is operative, generating relations between objects in the world and the parameters of its data. Its technical assemblage *precedes* both evidence and testimony. It engages in nonhuman witnessing. *Triple Chaser* brings the registering of violations of human rights into an agential domain in which the work of witnessing is necessarily inseparable from the nonhuman, whether in the form of code, data, or computation.

As development commenced, *Triple Chaser* faced a challenge: Forensic Architecture was only able to source a small percentage of the thousands of images needed to train a machine learning algorithm to recognize the tear gas canister produced by Safariland. They were, however, able to source detailed video footage of depleted canisters from activists and even obtained some material fragments. Borrowing from strategies used by Microsoft, Nvidia, and others, this video data could be modeled in environments built in the Unreal gaming engine, and then scripted to output thousands of canister images against backgrounds ranging from abstract patterns to simulated real-world contexts (figure 2.3). Tagging of these natively digital objects also sidestepped the labor and error of manual tagging, allowing a training set to be swiftly built from images created with their metadata attached (figure 2.4). Using several different machine learning techniques (including transfer learning, combining synthetic and real images, and reverse discriminators), investigators were able to train a neural network to identify Safariland tear gas canisters from a partial image with a high degree of accuracy and with weighted probabilities. These synthetic evidence assemblages then taught the algorithm to witness.

Like most image recognition systems, *Triple Chaser* deploys a convolutional neural network, or CNN, which learns how to spatially analyze the pixels of an image. Trained on tagged datasets, CNNs slide—convolve—a series of filters across the surface of an image to produce activation maps that allow the algorithm to iteratively learn about the spatial arrangements of pixels, which can be repeated across large sets of images. These activation maps are passed from one convolution layer to the next, with various techniques applied to increase accuracy and prevent the spatial scale of the system from growing out of control. Exactly what happens within each convolutional layer remains in the algorithmic unknown: it cannot be distilled into representational form but rather eludes cognition.[61] Machine learning processes thus exhibit a kind of autonomic, affective capacity to form relations between objects and build schemas for action from the modulation and mapping of those relations: machinic affect. Relations between elements vary in intensity, with the process of learning both producing and identifying intensities that are autonomous from the elements themselves. It is precisely this that cannot be "visualized" or "cognized." Intensive relations assemble elements into new aggregations; bodies affect and are affected by other bodies. Amoore writes that algorithms must be understood as "entities whose particular form of experimental and adventurous rationality incorporates unreason in an intractable and productive knot."[62] Reflecting on economic self-interest and the false grounds of rational choice, Massumi points out that "rationalities are apparatuses of capture of affectivity."[63] Machine learning works in concomitant ways. There is an autonomic quality to such algorithmic knowledge-making, more affective than cognitive. This machinic registering of relations accumulates to make legible otherwise unknown connections between sensory data, and it does so with the potential (if not intention) for that registering to make political claims: to function as a kind of witnessing of what might otherwise go undetected.

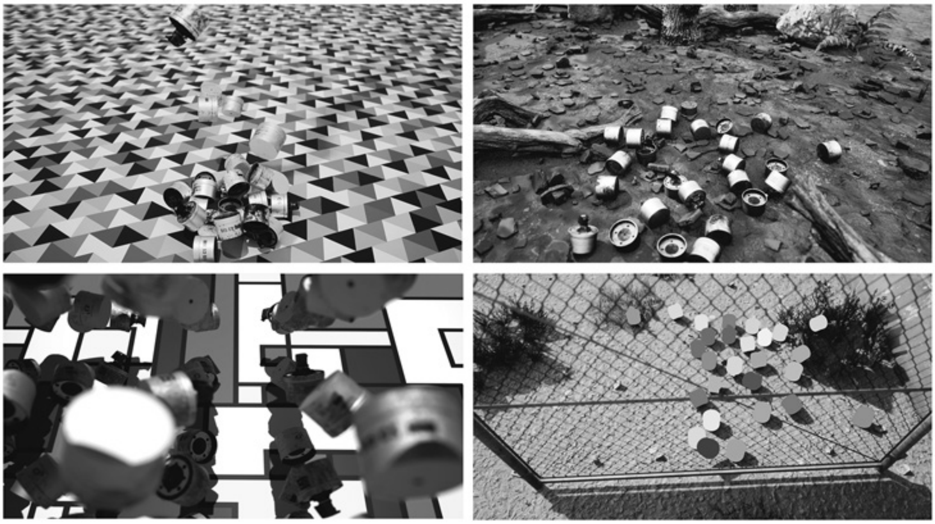


FIGURE 2.3. Four variations of synthetic media from *Triple Chaser*, Forensic Architecture, 2019. Courtesy of Forensic Architecture.

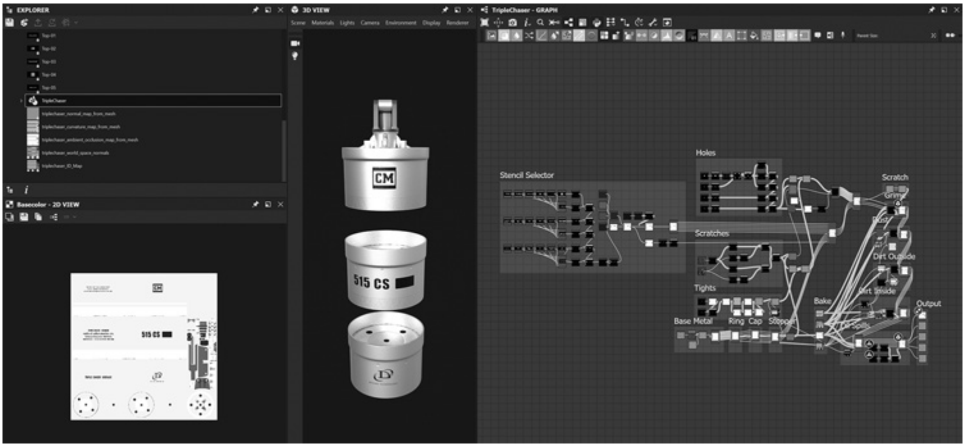


FIGURE 2.4. Applying weathering and wear effects to synthetic cannisters, Forensic Architecture, 2021. Courtesy of Forensic Architecture.

Underpinning the project is the proposition that social media and other image platforms contain within them markers of violence that can and should be revealed. For the machine learning algorithm of *Triple Chaser*, the events to which it becomes responsible are themselves computational: machinic encounters with the imaged mediation of tear gas canisters launched at protesters, refugees, and migrants. But their computational nature does not exclude them from witnessing. With so much of the world now either emergent within or subject to computational systems, the reverse holds true: the domain of computation and the events that compose it must be brought within the frame of witnessing. While the standing of such counterforensic algorithms in the courtroom might—for now—demand an expert human witness to vouch for their accuracy and explain their processes, witnessing itself has already taken place long before testimony occurs before the law. Comparisons can be drawn to the analog photograph, which gradually became a vital mode of witnessing and testimony, not least in contexts of war and violence.[64] Yet, despite its solidity, the photograph is an imperfect witness. Much that matters resides in what it obscures, or what fails to enter the frame, as in the nonhuman witnessing of Aleppo's aftermaths that I examined in the last chapter. With the photograph giving way to the digital image and the digital image to the generative algorithm, the ambit of witnessing must expand. As power is increasingly exercised through and even produced by algorithmic systems, modes of knowledge-making and contestation predicated on an ocular era must be updated for an age of more overt and complex machinic affect-ability. Forensic Architecture's work is also a potent reminder that nonhuman witnessing is a matter for galleries and activist politics as much as the courts, providing the aesthetic means for the human to comprehend its constitutive entanglement with the non. Even if the law resists the displacement of the human, art does not.

As *Triple Chaser* demonstrates, algorithmic witnessing troubles both relations between witness and evidence and those between witnessing and event. This machine learning system trained to witness via synthetic datasets suggests that the linear temporal relation in which evidence—the photograph, the fragment of tear gas canister—is interpreted by the human witness cannot or need not hold. Through their capacities for recognition and discrimination, nonhuman agencies of the machinic system enact the witnessing that turns the trace of events into evidence. Witnessing is, in this sense, a relational diagram that makes possible the composition of relations that in turn assemble into objects that can be experienced. If witnessing precedes both evidence and witness, then witnessing forges the witness rather than the figure of the witness granting witnessing its legitimacy and standing.

While this processual refiguring of witnessing has ramifications for nonhuman agencies and contexts beyond the algorithmic, Forensic Architecture's movement into this space suggests the strategic potential for an alternative politics of machine learning. In the four years since the release of *Triple Chaser*, Forensic Architecture has extended their use of machine learning to deal with identifying Russian tanks in Ukraine and other investigations. While I firmly believe that skepticism toward the emancipatory and resistant potential for machine learning and algorithmic systems more generally is warranted, there is also a strategic imperative to do more to ask how such systems can work for people rather than against them. With its tools, techniques, and synthetic media databases all made open source, Forensic Architecture aims to democratize the production of evidence through the proliferation of algorithmic witnessing that works on behalf of NGOs, activists, and oppressed peoples, and against the technopolitical state. This investigative commons becomes an intensive field for nonhuman witnessing, in which the entangled agencies of machines and humans work to register and make addressable otherwise elusive violence.

UNWITNESSED: PROJECT MAVEN AND LIMITLESS DATA

In June 2018, word spread inside Google that the company was partnering with the US Department of Defense (DoD) to apply its artificial intelligence expertise to the identification of objects in drone footage. A week later, the same news broke on the tech site Gizmodo. Within days, Google had withdrawn its engagement and released a set of principles for AI development that precluded working on weapons systems, although with plenty of wiggle room for other defense and national security applications.[65] The controversy marked a new notoriety for Project Maven, the code name for the Algorithmic Warfare Cross-Functional Team (AWCFT) created in April 2017 by order of the Deputy Defense Secretary Robert Work. Its stated aim was to "turn the enormous volume of data available to DoD into actionable intelligence" with an initial focus on providing "computer vision algorithms for object detection, classification, and alerts" in full-motion video from drone systems.[66] The AWCFT had a mandate to "consolidate existing algorithm-based technology initiatives related to mission areas of the Defense Intelligence

Enterprise, including all initiatives that develop, employ, or field artificial intelligence, automation, machine learning, deep learning, and computer vision algorithms."[67] Not only would the team seek partnerships with Silicon Valley, it would also adopt tech industry development techniques, such as iterative and parallel prototyping, data labeling, end-user testing, and algorithm training.[68] In a reversal of the Pentagon's typical hierarchical, drawn out, and multiyear technological development processes, Project Maven would be agile. It would fail often and learn quickly; move fast and break things—but with weapons systems.

Military secrecy makes even an approximation of the scale of data requiring analysis impossible to determine. Media reports suggest that the proportion of drone sensor data currently analyzed by humans represents a tiny fraction. An article in *Wired* cites DoD officials claiming that 99 percent of all drone video has not been reviewed.[69] Project Maven boss General John Shanahan is quoted as saying that twenty analysts working twenty-four hours a day are able to successful analyze—exploit, in military lingo—around 6 to 12 percent of imagery from wide-area motion sensors such as the ARGUS-IS persistent surveillance platform discussed in chapter 1. Project Maven aimed to bring AI analysis to the full-motion video data from the drone platforms doing much of the surveillance work against ISIS in Iraq and Syria: the MQ-1C Gray Eagle and the MQ-9 Reaper. By February 2017, DoD had decided that deep learning algorithms should ultimately be able to perform at near human levels but recognized that to do so meant working at scale. In its initial scoping, Project Maven was intended to enable several autonomous functions, including identifying thirty-eight different classes of objects, reverse image search, counts within bounded boxes and over time, and selective object tracking. It would integrate with Google Earth, ArcGIS, and other geographic information systems (GIS). Building datasets able to train machine learning systems would require human tagging of huge amounts of data. According to media reports, Project Maven outsourced much of this to the piecework platform Figure Eight (formerly CrowdFlower), providing unclassified and nonviolent images with instructions to draw and label boxes around various objects. Combined with classified imagery tagged by internal analysts, this data could train the convolutional neural network algorithms to identify and classify objects within video feeds, using iterative training and testing techniques honed in the tech industry.

Stored as ones and zeroes demarcating the position and color of pixels and accompanied by crucial metadata that makes them legible to the computational system, these images are optical only in potential. Unless called up by a human analyst for display on screen, the optical, communicative, and representational modality of such images remain potential only. Full-motion video (FMV) of the December 2013 drone strike on a wedding procession in Yemen carries no connotative or denotative meaning for the algorithm, despite its horrifying toll on the families and communities that lost a dozen lives.[70] The event's significance is obtained purely through its relations of similarity and difference to the sets of attributes invisually perceived by the learning algorithm. All such images are simultaneously virtual and actual along several dimensions at once: virtual code carrying the seeds of actual optical imagery; actual correlation in the unidentified scatter of virtual arrangements of pixels; virtual events crowding into the actualizing tendencies of the learning algorithm. Flagged as significant—a truck moving too swiftly; a cluster of bodies on a roof—virtual and actual coalesce to pull the sequence to prominence. We might name this *recognition*: the algorithm doing its job of *observing* and *discriminating*. But the algorithm does more than recognize, and we know from Oliver that recognition alone is not sufficient to produce witnessing. Such algorithms forge a relation of responsibility, rendering a set of relational attributes actionable within the field of possibilities produced by the rules of engagement and other framing structures of war. By producing claims to know the world that demand response, even if that response is to pull a trigger that kills, algorithmic witnessing within the drone apparatus does something more than mere observation. Or, rather, within a certain confluence of machinic affective dynamics, the drone video algorithm generates a field of human-nonhuman relations that becomes witnessing.

Applying these same principles to activities more complex than object identification—assembling machine learning tools that can determine that a particular confluence of objects and attributes constitute a target—both heightens the stakes of nonhuman witnessing and introduces new problems into the technical processes themselves. In the signature strikes undertaken by the US military, a narrow set of data points—most of them drawn from cellphone signal interceptions—provides the basis for algorithmic analysis. As Amoore writes, "When a random forest algorithm sentences someone to death by drone strike, the infinite (gestures, connections, potentials) makes itself finite (optimal output, selector, score), and the horizon of potentials is reduced to one condensed output signal."[71] This violent mediation is produced by and through machinic affects: "A random forest algorithm will never know a terrorist in the sense of acting with clearsighted knowledge, but it mobilizes proxies, attaches clusters of attributes, and infers behaviors to target and to act regardless."[72] Intensities of relation spark the algorithm into response: action necessitated by clusters of machinic intensities coming together to stir a determination that carries with it an ethical weight.

From the perspective of the machine learning system, shifting the emphasis of analysis from cellphone metadata to video imagery is largely a matter of complexity and access to large arrays of the GPUs necessary for computing imagery at scale, which is not to say that such systems will work accurately, limit violence, or reduce civilian harm. In much the same way that deepfake tools can be trained on audio as well as video, the distinctions are largely a matter of input data. Yet while the AWCFT has access to countless hours of mundane footage that can provide a base dataset for tasks like identifying vehicles and buildings, FMV of threats that might warrant lethal action or even tracking by drone systems seemingly remain too scarce, too ambiguous, or too like other imagery. There simply isn't enough video for the machines to learn effectively. Using similar techniques to Forensic Architecture, the AWCFT reportedly generated artificial environments to produce training data for threat detection systems.[73] While details of that training data remain classified and inaccessible, the introduction of synthetic data into a target-detection system has a rather different valence from its use by Weizman and his collective. We know that both war and policing, its domestic corollary, depend upon and reproduce existing sociocultural codings, particularly those around race, gender, and class. In a country like Afghanistan, where men are frequently armed, the baseline designation of "military-aged males" predisposes the system to see activities such as the jirga or council as incipient threats. Just such a prefiguring contributed to the unconscionable drone strikes against just such a gathering at Datta Khel, a village in North Waziristan, that killed forty-four people in 2011. If training data is synthesized within existing frames of war, what structures of domination and their attended biases, misconceptions, and fantasies are coded into such training materials? How might predictive tools gear toward identifying certain peoples and activities as threats? Nonhuman witnessing within the algorithmic systems might bear false witness in far more subtle and ingrained ways than deepfakes can manage.

Drone warfare and drone policing alike are necessarily racializing: they encode and produce racialized subjects as threats, with threat and race intimately bound up with each other.[74] Race is coded into the drone system all too readily. This is because, as Ruha Benjamin writes, "race itself is a kind of technology—one designed to separate, stratify, and sanctify the many forms of injustice experienced by members of racialized groups, but one that people routinely reimagine and redeploy to their own ends."[75] In the algorithmic shift from visual to nonvisual regimes of classification, Thao Phan and Scott Wark argue that "race emerges as an epiphenomenon of automated algorithmic processes of classifying and sorting operating through proxies and abstractions," refiguring "racial formations as data formation."[76] According to Lauren Wilcox, drone warfare, primarily deployed in the Islamic world, "simultaneously produces bodies in order to destroy them, while insisting on the legitimacy of this violence through gendered and racialized assumptions about who is a threat."[77] Identifying specific bodies as threats necessitates preconditioning them as threats within the system, which means determining which bodies should be coded for exposure, to borrow a phrase from Benjamin. Generating training sets from synthetic events staged in 3D environments begins with a set of decisions about who and what to include, what people and places should look like, how people will act, and what constitutes and defines the relations between places, people, and actions. Both the production of data and the iterative development procedures used by Project Maven mean that the nonhuman witnessing of its algorithms is meshed with the human. Far from removing human partiality, such processes embed the discursive, affective, and fantastic logics of war in all their racializing and gendering dimensions into the algorithm at every stage of its design, training, and operation. Nonhuman witnessing in the context of drone warfare is thus not a move toward impartiality or the diminishment of the human, but rather the technical concretization in code of predetermined meanings that are inescapably colonial and racist.

Project Maven and its AI ilk train the martial gaze on the unwitnessed events of life in the age of drone warfare. Yet this witnessing is not analogous to the human: it is fractured and distributed within the system, a techno-affective witnessing composed of machinic intensities. Signature strike, threat detection, and targeting algorithms are not witnessing subjects in the humanist sense, but witnessing assemblages distributed across the nonhuman, invisual perception of machine learning systems. This invisual perception necessitates the exclusion of much that is captured by the drone's optical and multispectral sensors: an infinitude of moments both major and minor necessarily go unwitnessed. Or, rather, nonhuman witnessing within drone algorithms always entails unreasoned and psychotically rational judgments about what matters. Nonhuman witnessing in its algorithmic, war fighting form must necessarily fail to note forms of violence (martial, environmental, interspecies, interpersonal) that do not figure in the criteria for tagging imagery and reinforcing machine learning. Such imagery does not go "unseen" as such, but rather its seeing fails to register. Algorithmic witnessing of this kind is necessarily narrow, not more efficacious or richer than the human but stunted and strange.

WITNESSING THE ALGORITHM: BEYOND THE BLACK BOX

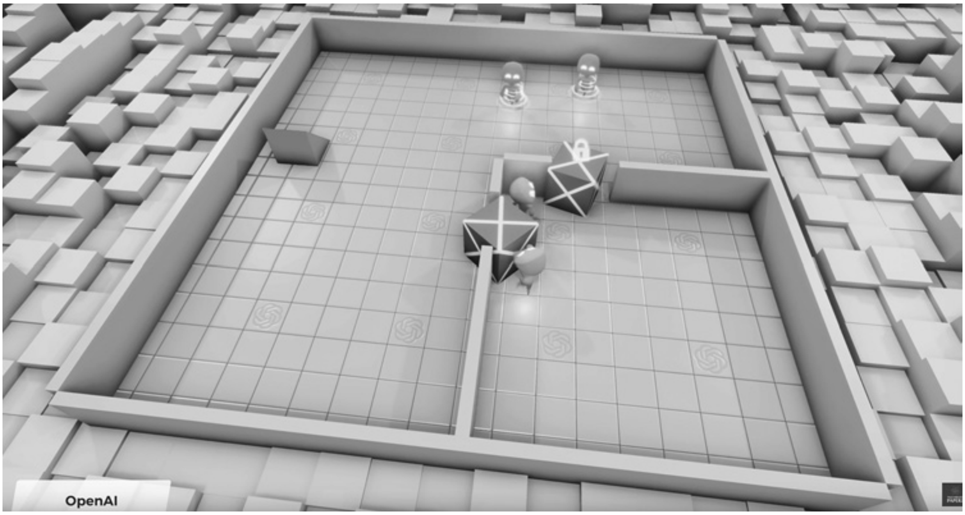
Safiya Noble's *Algorithms of Oppression* opens with a now famous anecdote about searching for "black girls" on Google and finding porn sites, then searching for "white girls" and finding young girls at play. Throughout her book, Noble shows what has since become well understood: algorithmic systems repeat, entrench, and even sharpen the racism, misogyny, homophobia, and other normative biases already present in the cultures from which they arise. Faced with the technical, legal, and commercial black boxing of algorithms, critical scholars such as Noble have rightly focused on the institutions, structures, and applications through which data is collected, computed, and instrumentalized by government and corporate entities. Noble's work is part of a growing body of critical practices (scholarly, artistic, and activist) concerned with the reproduction of inequality in algorithmic systems that has had a significant impact on public debate. In the United States, Joy Boulminwi's activist research and poetry exposed the biases of facial recognition, while Virginia Eubanks's ethnographic investigation revealed the inequalities exacerbated by the infiltration of algorithms into social welfare systems. In Australia, collective advocacy spearheaded by journalist Asher Wolf and others forced the federal government to abandon its automated "RoboDebt" welfare debt collection tool. These, and many other interventions, have increased public understanding of the existence and effects of coded bias, and forced the tech industry to take steps to redress its harms: creating ethics boards, inviting critical research, appointing bias engineers, and seeking to diversify their workforces. But such gestures are often mere fig leaves, swiftly sidelined or rolled back when their presence becomes uncomfortable, evidenced most prominently by the dismissal of Timnit Gebru from Google's Ethical AI team for her refusal to withdraw an academic paper from publication that raised ethical and ecological concerns about large language learning models.[78] Achieving what Lina Dencik calls "data justice" requires more radical change and tech companies and their clients have been far less inclined to ask whether certain computational systems should be built at all.[79] Nor have they been willing to peel away the legal and commercial wrapping on their algorithmic black boxes, which are tightly guarded as data becomes a contemporary form of capital.[80] And even if those boxes were more open, the question of whether deep learning algorithms would reveal themselves in a comprehensible way remains fraught. How, then, to witness the algorithm?

One answer to this question is in the witnessing of their effects: the increasingly common irruptions of outrage in response to injustices perpetrated by algorithmic systems. Or we might think of glitches in the algorithmic that make visible their computational construction, such as the strange fault lines and blurs that can be found on Google Earth, the misrenderings of the world-making, machine learning network responsible for Microsoft Flight Simulator, or the hallucinations of Microsoft's chatbot version of its Bing search engine. Another strategy has been to make infrastructures themselves visible. If we cannot see into the algorithm itself in a meaningful way, then might its infrastructural assemblages be worthy of attention? Trevor Paglen's eerie photographs of the National Security Agency Building and of militarized data centers across the United States are one such project. Another is John Gerrard's "The Farm" (2015), which used aerial photography of a Google data center in Oklahoma as the reference for a finely detailed simulation of the same center. Exhibited as a high definition projection, Gerrard's work stays on the outside of the data center itself but recreates its digital world through the construction of a computational simulation using the Unigine gaming engine.[81] Venturing speculatively within the data center itself, Kynan Tan's "Polymorphism" (2016) digitally recreates its interior materialities, such as the nonhuman movements of automated tape back-up systems as they robotically traverse arrays of server racks. When first exhibited, Tan used powerful subwoofers to sonically simulate the noise of the data center, a sensory engagement with the imposing infrastructures, resource intensity, scale, and speed of computation without making specific operations legible. In these and other such works, the algorithmic is witnessed not through its code but through its infrastructures: the hard, imposing materialities that undergird and make possible the purportedly ephemeral clouds of global computation.

Within the AI research community, this problem of invisibility and inaccessibility is well recognized. For example, the computer science association ACM now runs FAccT, an annual interdisciplinary conference on fairness, accountability, and transparency in AI.[82] San Francisco company OpenAI, now widely known for its ChatGPT platform, provides one response to the issue of AI black boxing via its Two-Minute Papers channel on YouTube, which presents the learning undertaken by neural networks in two-minute animations that show both what and how learning occurs over time, using newly published papers as the primary source. With more than 1.1 million subscribers and many videos viewed several million times, the channel represents a remarkably effective approach to making visible the learning processes of AI systems. In one video, an OpenAI neural network learns to play hide and seek, using boxes, ramps, walls, and game rules to succeed (figure 2.5). As the environments change and the algorithm learns, strange happenings occur: agents run up ramps to jump on boxes, use glitches in the simulation's physics to fly through the air, and throw objects off the screen. Both the limits and possibilities of machine learning are immediately evident. Algorithms become agencies that seem comprehensible because their workings can be broken down into episodic form and cutely animated. Despite their propagandistic intent, these videos hint at possibilities for witnessing through making machine learning visible—even if OpenAI has become increasingly secretive about how its ChatGPT uses the GPT series of large language models.

Placing this dynamic within contexts of labor, logistics, and warfare, Kynan Tan's *Computer Learns Automation* (2020) slows down the machine learning process within three simulated environments and in doing so allows a human audience to become cowitness to the nonhuman witnessing of AI training. *Computer Learns Automation* is composed of three separate training environments and agents: "RideShare," in which a vehicle learns to navigate an urban environment to pickup and drop-off fares; "Robot Arm," in which an automated device learns to pick up boxes from one conveyor

FIGURE 2.5. Still from "OpenAI Plays Hide and Seek . . . and Breaks the Game!," OpenAI, 2019

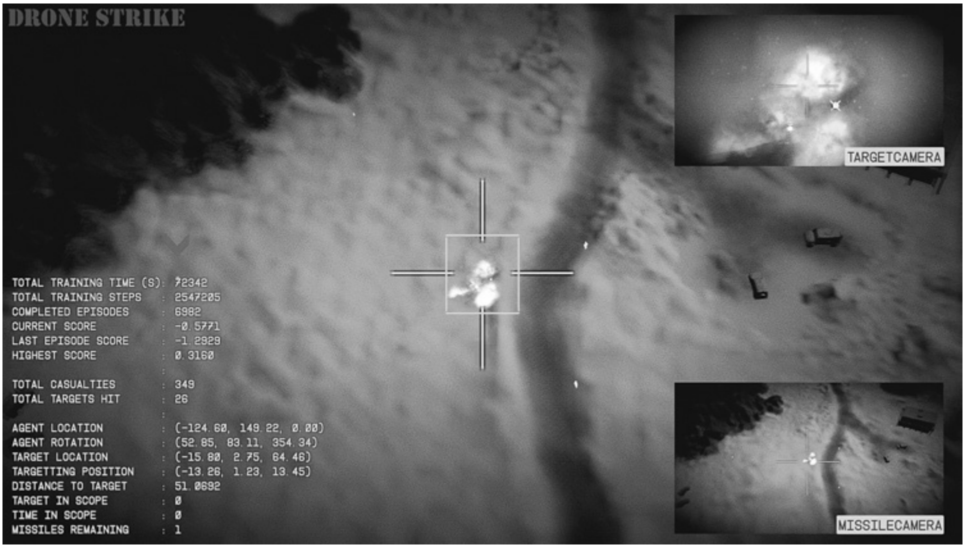


belt and place them on another; and "Drone Strike," in which a targeting reticule learns to move across terrain, identify human targets, and launch accurate missile strikes against them. Exhibited at the Adelaide Biennial 2020 at the Art Gallery of South Australia, Tan's work slowed the neural network's learning process down with the intention that all three agents develop reliable capability in their tasks during one hundred days on display. On three high-definition screens, visitors to the gallery could watch thousands upon thousands of learning cycles as the car, robot arm, and crosshairs moved across the screen. On the lower left of each screen, data readouts list training time, training steps, total episodes, current and highest scores, and so on, as well as information more specific to each agent, such as boxes moved or casualties inflicted. This set-up allows viewers to watch in real time as the machine learns to navigate and act on its environment, seeing what it sees as it learns. Situated in identified contexts rather than playing hide and seek or undertaking an abstract task, relations between labor, death, and value production are visually tied to the learning of the machine.

Computer Learns Automation was built in the Unity 3D engine using its native ML-Agents package, which provides a set of trainable algorithms that can be linked to the simulated environment so that it can send and receive data. This allows the algorithm to learn and act in the environment simultaneously, which in turn makes it possible to watch the process of the machine's learning unfold. Slowed down, each learning cycle of the system can be a viewed in a legible, computational real time. On display in Tan's work is an algorithm known as "proximal policy optimization," a reinforcement learning neural network technique developed by OpenAI. Earlier "policy" deep learning algorithms used analysis of an environment to select between possible options for an agent, but had a tendency to be overly influenced by choices around how much or little the policy should adjust in response to stimuli: "Too small, and progress is hopelessly slow," notes OpenAI, "too large and the signal is overwhelmed by the noise, or one might see catastrophic drops in performance."[83] Proximal policy optimization, or PPO, corrects this tendency by feeding updates back to the policy at each step, aiming to produce just enough reward to ensure the network learns quickly but doesn't rush down a false path. Tan's "Rideshare" vehicle learns to find, collect, and deliver fares by attempting an action, receiving a reward—or not—and then updating its policy to reflect that information. Driving forward, avoiding a building, not hitting pedestrians—actions such as this accumulate through cycles of training to teach the algorithm to achieve an objective. PPO requires a lot of sample cycles to be effective, but it balances the network's dueling desires to explore options and to exploit what it knows will reap rewards. *Computer Learns Automation* reveals the stuttering, iterative, inhuman, and imperceptible graduations through which this mode of deep learning takes place.

In contrast to the crisp aesthetics that characterize much of Tan's digital work, *Computer Learns Automation* has an operative, processual quality. Its simulated environments are just real enough: decodable by the viewer in the domain of signification while appropriate to context for the machine. Yet while the environments, agents, and actions are recognizable to the visitor to the gallery, their semantic significance within the machinic network eludes comprehension. Just what is happening in each of the three learning simulations is uncomfortable, even unsettling. Machinic affects course through the visual field yet are themselves what Mackenzie and Munster call "invisual": concerned with the composition of relations that are not themselves visual at all but rather the perceptual upwell of operations deep within the hidden layers of the PPO network as it explores and exploits according to technical logics that retain a certain unassailable incomprehension. As the targeting reticule of "Drone Strike" inches its way uncertainly across the mountainous terrain, learning to find, fix, and finish the small collection of pixels that indicates a human body, its movements are not those of the intentional human operator but rather of a machine motivated by an autonomic system for which the notion of a meaningful goal is itself without meaning—or, rather, without correlative meaning for the human viewer (figure 2.6). The machinic resides in its emerging relational awareness, its becoming operative. Tan invites us into the disjunctive space where unthinking yet agential, operative, and transformative machine learning systems intersect with bodily violence, labor exploitation, and automated logistics.

The human audience can only ever become cowitness, a status reinforced by the practical impossibility of watching all three environments learn over one hundred days. Yet this partial cowitnessing is crucial to pulling the nonhuman witnessing already at work in the visualization of the invisual learning process. With casualty counts, founder wealth, and shareholder value ticking up as the machine learns in its slowed down real time, *Computer Learns Automation* reminds us of the necessity of nonhuman witnessing bridging the seeming divide between technical systems, material conditions, and the politics of technocultures. Tan insists on pursuing the always incomplete task of witnessing algorithms in their elemental states, in their becoming increasingly, brutally, and efficiently operative through the intensification and accumulation of machinic affects within the invisual domain of computer vision.



```
TOTAL TRAINING TIME (S): 72342
TOTAL TRAINING STEPS    : 2547205
COMPLETED EPISODES      : 6982
CURRENT SCORE           : -0.5771
LAST EPISODE SCORE      : -1.2929
HIGHEST SCORE           : 0.3160
                        :
TOTAL CASUALTIES        : 349
TOTAL TARGETS HIT       : 26
                        :
AGENT LOCATION          : (-124.60, 149.22, 0.00)
AGENT ROTATION          : (52.85, 83.11, 354.34)
TARGET LOCATION         : (-15.80, 2.75, 64.46)
TARGETTING POSITION     : (-13.26, 1.23, 13.45)
DISTANCE TO TARGET      : 51.0692
TARGET IN SCOPE         : 0
TIME IN SCOPE           : 0
MISSILES REMAINING      : 1
```

FIGURE 2.6. Still from *Computer Learns Automation*, Kynan Tan, 2020. Courtesy of the artist.

WITNESSING ALGORITHMS

Across this chapter, witnessing algorithms has emerged as a polysemic concept. In the narrowest sense, witnessing algorithms can be understood quite straightforwardly as algorithms that enable witnessing. In this vein we might think of social media algorithms that bring certain kinds of news reportage to the fore or algorithmic systems such as MS Flight Simulator's that enable an entirely simulated witnessing of the world. Yet witnessing algorithms are also engaged in witnessing on their own behalf: a registering of happenings in material and machinic worlds to which the algorithm obtains its own version of responsibility. Algorithmic responsibility is not identical to that of its (multiple, contestable) human equivalent, but rather describes the emergence of relations that cohere and produce a necessity for action, however that might come to be within the technicity of the algorithm. Consider how images of the *Triple Chaser* tear gas canister come to matter in the machine learning algorithms of Forensic Architecture, or the particular movements of bodies and vehicles in computer vision designed for drone applications under the aegis of the Algorithmic Warfare Cross-Functional Team. Exactly how those neural networks raise certain phenomena to significance but not others—how worldly happenings trigger machinic interest or not—eludes complete knowing, yet as the algorithm learns it is shaped by its (un)witnessing. At the same time, algorithms are themselves entities that must be witnessed, both in their effects and in their operations. As Kynan Tan's *Computer Learns Automation* teaches us, at issue here is the unknowable machinic affects of the learning networks themselves. Witnessing the algorithmic is not another demand to open the proverbial black box, but rather a call for attending to algorithmic agencies in their emergence and on their own terms.

With algorithms swiftly becoming preeminent knowledge instruments of governance, commerce, culture, science, and social life, how we reckon with the technopolitics of their identification and formation of relations between worldly phenomena is an urgent question. As the hype of generative AI reaches fever pitch, the political stakes of this task only heighten. Within critical algorithm studies and data justice movements and activism, crucial new lines of inquiry continue to multiply, not least in conjunction with the enduring political projects of resisting settler colonialism, struggling for racial justice, and fighting for a meaningful response to the climate crisis. This chapter has sought to think in sympathy with those inquiries, projects, and goals, asking how algorithmic technologies witness the world and how they in turn might be witnessed. My approach here follows Amoore's call for a cloud ethics that "does not belong to an episteme of accountability, transparency, and legibility, but on the contrary begins with the opacity, partiality, and illegibility of all forms of giving an account, human and algorithmic."[84] Following Édouard Glissant, such an opacity provides the ground for understanding knowledge and politics as emerging from irreducible difference rather than being founded upon its erasure.[85] It is the condition that underpins relation. Pluriversal politics necessitates just such an opacity, as well as its accompanying partiality and illegibility, because a world of many worlds requires the impossibility of transparency, otherwise any world becomes legible to any other, and in doing so ceases to possess vitality on its own terms. In the domain of algorithms, analytically separating witnessing within the event from the testimony that takes place after is urgent because so much of witnessing algorithms is about registering events, machinic or otherwise, as processes of emergent relations (of knowledge, power, violence, control and aesthetic and political potential)—rather than explaining, narrating, or communicating them.

Perhaps more than any other domain addressed in this book, the algorithmic cuts across world-ending catastrophes old and new. Algorithms are infrastructural, institutional, and embedded in technocultural milieus that are, in turn, inseparable from the production of race, gender, and class. Rooted in settler-colonial practices of categorization and control, financialization, climate modeling, and war, as much as in social media and internet search, algorithms are yet another site of deep entanglement between human and nonhuman. Nonhuman witnessing of, in, and through algorithmic processes is about finding those links and recognizing that it no longer, if it ever did, makes sense to think of the human witness outside the nonhuman technicities of our media, our archives, and their agential materialities. Nonhuman witnessing arises from a field of relations between the human and nonhuman, relations that are as ecological as they are technical—and it is to the ecological that this book now turns.