2

The Learning Machines

Neural Networks and Regimes of Recognition

> The peculiarity of men and animals is that they have the power of adjusting themselves to almost all the features [of their environment]. The feature to which adjustment is made on a particular occasion is the one the man is attending to and he attends to what he is interested in. His interests are determined by his appetites, desires, drives, instincts—all the things that together make up his "springs of action." If we want to construct a machine which will vary its attention to things in its environment so that it will sometimes adjust itself to one and sometimes to another, it would seem to be necessary to equip the machine with something corresponding to a set of appetites.
>
> —Richard Braithwaite, "Can Automatic Calculating Machines Be Said to Think?"

Springs of Action

In January 1952 the BBC recorded what was to be the world's first public debate on the mathematics and ethics of machine learning. Participating in the discussion of the question "Can Automatic Calculating Machines Be Said to Think?" were Manchester mathematicians Alan Turing and Max Newman; their neurosurgeon colleague Geoffrey Jefferson; and Cambridge moral philosopher Richard Braithwaite. Turing explains that he has "made some experiments in teaching a machine to do some simple operation," but that "the machine learnt so slowly that it needed a great deal of teaching." Jefferson is skeptical of the use of the verb "to learn" in relation to the machine, and he interjects to ask Turing a question: "But who was learning, you or the machine?"[1]

This distinction, between humans and machines as the locus of learning is of great significance to the neurosurgeon, for whom the electronic circuits of computing machines were not analogous to the "fragments of the nervous system" he encountered in the fleshy materiality of the human brain.[2] Yet, Jefferson's question prompts Turing to reflect on his own embodied experience of experimenting with his machines. In the audio recording of the debate, Turing can be heard to pause for a moment's reflection before responding: "I am inclined to believe that when one has taught it [the machine] to do certain things[,] one will find that some other things one had planned to teach it are happening without any special teaching being required." In response to Jefferson's question of who was learning, the mathematician or the machine, Turing responds, "I suppose *we both* were."[3]

The entangled "we both" of mathematician and machine, learning together, expresses Turing's belief that intuition is a mathematical faculty that "consists of making spontaneous judgements which are not the result of conscious trains of reasoning." For Turing, the intuitive faculty is entangled with what he calls the "ingenuity" of the building of rules as arrangements of propositions.[4] The iterative relationship between intuition and ingenuity in mathematical reasoning necessarily entangles the mathematician's affective and haptic relations to a puzzle with the making of a formal axiom or logic. The human and machinic elements of mathematical learning, then, are not so readily disaggregated for Turing. Though for the philosopher in the discussion, Richard Braithwaite, it is the unique "peculiarity of men and animals" that they are able to learn intuitively by "adjusting themselves to almost all the features of their environment," his notion of a "spring of action" afforded by appetites nonetheless calls to mind today's capacities for machine learning algorithms to learn and to generate things in excess of their taught rules.[5] The 1950s radio discussion of the character of machine learning did, in some respects, envisage a future world in which machines would exceed the rules-based decision procedure and extend to the affective pull of intuitions and appetites for data.

Thus, even in the mid-twentieth century, the mathematics and philosophy of machine learning was centered on the entangled relations of humans and machines. The question, as articulated in the debate on automatic calculating machines, was "Who was learning, you or the machine?," and Turing's reply was "We both were." In this chapter, I focus on the *we* invoked by Turing in this public debate precisely because it runs against the grain of contemporary moral panics amid machine autonomy and algorithmic decisions that appear to be beyond the control of the human. On the contrary, the *we* of machine learning is a composite figure in which humans learn collaboratively with algorithms, and algorithms with other algorithms, so that no meaningful outside to the algorithm, no meaningfully unified locus of control, can be found. In contemporary machine learning, humans are lodged within algorithms, and algorithms within humans, so that the ethicopolitical questions are concerned less with asserting human control over algorithms and more with how features are extracted and recognized from a teeming data environment.[6] In short, the ethicopolitics of machine learning algorithms is located within the figure of the *we*—in the very relations to ourselves and to others implied in the *we* who have a spring of action.

In our contemporary moment, the "we both were" extends the already multiple body via the sinewy and invasive techniques of deep learning and neural network (neural net) algorithms.[7] The extended *we* of the multiplicity of data to which the learning algorithm is exposed heralds an intimate communion of the learning machines with a vast and incalculable *we*: all of us, all our data points, all the patterns and attributes that are not quite possessed by us.

In the pages that follow, I begin by taking up the theme of intuitive learning via the extracted features of a data environment in the context of the twenty-first-century advent of surgical robotics. At the level of this specific type of deep neural network algorithm, there is no technical distinction between learning actions for robot surgery and learning actions for robot weaponry. Across different domains of life, these algorithms are concerned with translating the input data from their environment into a "feature space," mapping the features into clusters of significance, and extracting the object of interest for the action.[8] Thus, algorithms designed to save lives, via robot surgery, or to end lives, via robot warfare, share the same arrangements of propositions. In following the machine learning of surgical robotics, I am concerned to capture the impossibility of establishing definitive boundaries of good and evil in relation to algorithms. The machine learning algorithms deployed in robot surgery do save lives through the lower infection rates of noninvasive methods, but they also endanger life through error and miscalculation. My point is that the ethicopolitics of machine learning algorithms cannot be mapped onto the parameters of good and evil or the securing and imperiling of life. With the extraction of feature spaces, machine learning algorithms are actively generating new forms of life, new forms of boundary making, and novel orientations of self to self, self to other. To begin with robot surgery is to begin in a place where one could never definitively draw a line delineating the algorithmic moral good from some sense of immorality or evil.

Intuitive Surgery: Making the Singular Cut

The intuitive relation to mathematics noted by Turing finds a contemporary form in robotic surgical systems such as Intuitive Surgical's da Vinci robot. The application programming interface (API) and the cloud storage architecture of the da Vinci robot contain the data residue of multiple past human and machine movements. Figure 2.1 displays the "surgical gestures" of the movements of surgeons' hands on the remote console, as mediated through the robotic instruments of the da Vinci.[9] The surgical procedure modeled here is a routine four-stitch surgical suture to close an incision. Though the surgical gestures involved in the cutting and stitching of flesh are a matter of haptic routine for human surgeons, for the designers of the algorithms, the objective is to model the optimal suturing motion so that future human *and* robot surgeons have their intuitive movement shaped by the ingenuity of the model. As the Johns Hopkins computer scientists building the model explain, the process begins with the "automatic recognition of elementary motion" from the extraction of features in the data environment. The model extracts seventy-eight features, or "motion variables," from the vast quantity of video and sensor data archived by the da Vinci robot (figures 2.2 and 2.3)—twenty-five feature vectors from the surgeon's console, and fourteen from the surgical instruments attached to the patient-side robotic arms.[10] In the surgical gesture, the movement of the human hand is thus thoroughly entangled with the remote console and the surgical scalpel held by the robot's hand. The scientists describe the juxtaposed map of surgical gestures: "The left [top] plot is that of an expert surgeon, while the right [bottom] is of a less experienced surgeon."[11] Here, algorithms are enrolled to recognize surgical gestures and to extract the features of movement, to actively distribute cognition across human surgeons and robots, and to optimize the spatial trajectory of the act of suturing flesh. The future surgeon will learn to suture flesh optimally, via the robot's simulation functions, and the robot surgeon will learn to suture autonomously from the data of past gestures of expert human surgeons. Though the computer scientists do envisage autonomous actions by the robot—with "the possibility to automate portions of tasks, to assist the surgeon by reducing the cognitive workload"—this apparent autonomy is entirely contingent on the layered learning from models of past entanglements of human and robot gestures.[12]

The spring to action of surgical machine learning is not an action that can be definitively located in the body of human or machine but is lodged within a more adaptive form of collaborative cognitive learning.[13] Intimately bound together by machine learning algorithms acting on a cloud database of medical

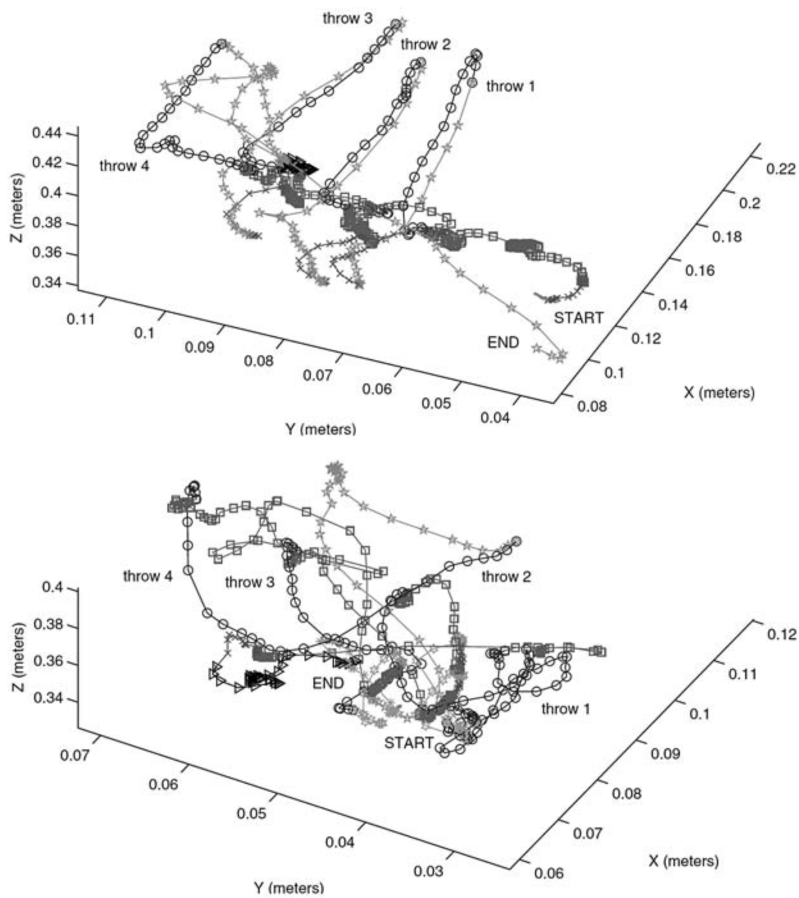
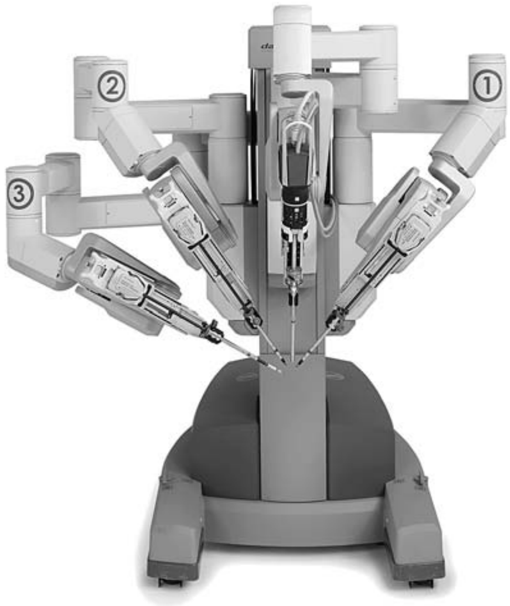


Figure 2.1 A plot of the trajectories of surgical gestures during a four-stitch suturing procedure. The top plot shows the trajectories of an expert surgeon, while the bottom plot shows the trajectories of an inexperienced surgeon, as extracted and analyzed by the surgical robot. Lin et al., "Automatic Detection."

data, the *we* of surgeon and robot restlessly seeks an optimal spring of action—the optimal incision, the optimal target of tumor or diseased organ, the optimal trajectory of movement. Intuitive Surgical's robots hold the promise of the augmented vision and precise trajectories of movement of a composite being of surgeon and machine. As one of the UC Berkeley robotics scientists describes what they term "iterative learning," machine learning allows "robotic surgical assistants to execute trajectories with superhuman performance in terms of speed and smoothness."[14] The drive for superhuman learning iterates back



Figures 2.2 and 2.3
The da Vinci surgical robot. Intuitive Surgical, 2018.

and forth across multiple gestures. As the hand and eye movements of a singular surgeon are tracked in the da Vinci's API, they commune with a mathematical model generated from a vast multiplicity of data traces of past surgical gestures. These gestures, in turn, modify the learning model to optimize and augment the future trajectories of future surgeons and robots as yet unknown. As the philosophers and scientists in the 1952 debate anticipated, the machine learning algorithms have something close to appetites, extracting and modeling features from the plenitude of cloud data, rendering springs of action, and acting on future states of being.

During the course of following the surgeons and robots of one world-leading oncology department in a UK teaching hospital, I began to note the many occasions when a surgeon used "we" in place of "I" to describe their daily collaborations with surgical robots. This *we* who learns is expansive. It includes the many humans in the research group and the surgical team, but also the many human and technical components of the simulation of a surgery—the multiple layers of medical imaging, video data from past surgeries, and algorithmic models that together composed a kind of virtual presurgery.[15] For example, when the UK surgical team were preparing to conduct a new surgical procedure on a specific type of tumor, they collaborated with other US surgeons who had previous experience of the procedure. This was not merely a dialogue between human experts, however; the research team also imported the data from the US surgeries, inputting them in an algorithmic model and experimenting with the parameters for a new context. As Rachel Prentice documents in her meticulous ethnography of surgical education, "surgical action must be made explicit for computers" so that "bodies and their relations in surgery are reconstructed" in a form that "can be computed."[16] This reciprocal and iterative relationship between human surgeon and computer is what she calls a "mutual articulation" in which "bodies affect and are affected by" one another.[17] When the movement of a surgeon's hands is rendered computable in the robotic model, Prentice suggests that this "instrumentalization" overlooks the "tacit" and "tactile experience" of surgery, such as the "elasticity of a uterus or the delicacy of an ovary."[18]

In Prentice's reading of surgical technologies, the tacit, tactile, and intuitive faculties of surgery define the human as the locus of care and embodied judgment and decision. Yet, in giving their accounts of working with robots, I found that human surgeons testify to their own body's capacities coming into being in new ways. What it means to be intuitive, to touch or to feel an organ, for example, alters with the advent of machine learning modalities of surgery. Seated at their virtual environment console, the surgeons access video feed

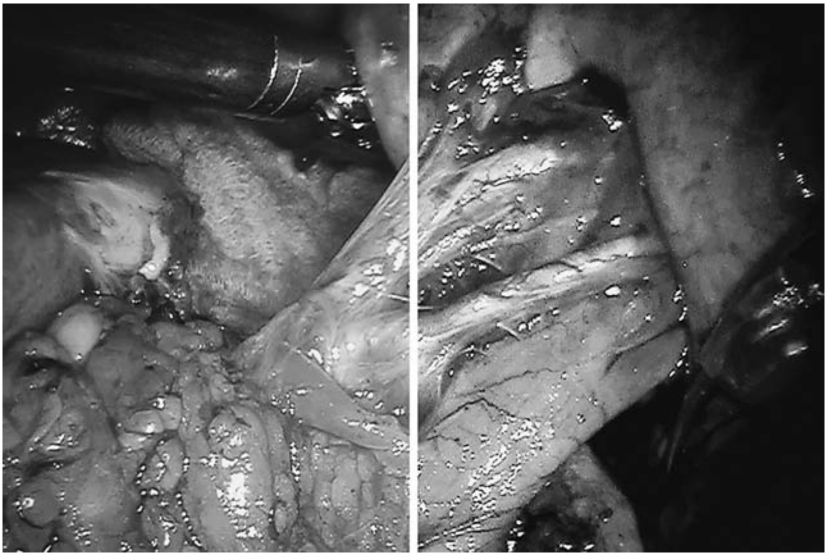


Figure 2.4 Fluorescence imaging of renal parenchyma, as viewed from the surgeon's console. Intuitive Surgical, 2018.

images from the endoscopic arm of the robot (figure 2.4). These images are overlaid with MRI and the fluorescence images of tumors so that, as one of the da Vinci computer scientists explains, "the system provides over a thousand frames per second and filters each image to eliminate background noise."[19] The work of the algorithms here is to extract the features of interest for perception by the surgeon, surfacing the optimal image from a background noise of teeming data.

One surgeon described to me the daily work of obstetric surgery with her da Vinci robot, noting that through the expanded mediated space, she is able to "see the unseeable" and "reach the unreachable" within the patient's body.[20] In contrast to Prentice's sense that human touch—and with it, judgment and decision—is evacuated and instrumentalized by surgical robotics, here the entangled touch of surgeon and robot invokes different relations of judgment and decision. The surgeon's relations to herself and to others—to her patients past and present, her operating theater colleagues, the robots, images, tumors, surgical instruments—are altered in and through the machine learning algorithms. "Touch engages us in a felt sense of causality," writes Karen Barad, so that "touch moves and affects what it effects."[21] Where Prentice foresees a loss of responsibility as the human surgeon's touch is evacuated by the robot, I propose instead an extension of responsibility to that which extends and exceeds human sensibility. The difference is an important one. The terrain of uncertainty regarding whether a specific tumor is operable without major damage to surrounding organs, for example, shifts with the 360-degree mobility of the robot's wrist—what can be precisely touched, sensed, and extracted from the body is altered. Indeed, leading computer scientist and artist Ken Goldberg, alongside his carefully presented accounts of the development of stochastic models and neural networks for the performance of surgical excision by robots, writes essays on the insensible and uncanny worlds that open up with his algorithms. Jochum and Goldberg's account of the "experiential uncanny" describes how robot actions "stretch the boundaries between the animate and inanimate" in new directions, serving to "challenge our beliefs about what, precisely, separates humans from machines."[22] The computer scientist's reflections on the embodied and intuitive capacities of his algorithms—to change the nature of what can be seen, reached, touched, or learned—run against the grain of a surgeon's hands being rendered computable by autonomous machines. Instead, machine learning algorithms work with the incomputable to open up new worlds of intuitive and insensible action.

As machine learning algorithms engage in "stretching the boundaries," the object that is surfaced for perception and action communes intimately with data on the events of past surgeries. This communion on *what is optimal*—the cut, the incision, the surgical strike—belongs properly to a composite being within the cloud analytic I describe in the previous chapter. The da Vinci data are no longer territorially limited to the memory of a specific robot, a server, an individual surgeon, or a group of scientists. Rather, the machine learning algorithms are deployed in a cloud architecture that yields the data residue of many millions of past surgeries. Lodged inside the actions of the singular cut—itself bordered by algorithms optimizing the thresholds of the instrument's trajectory—are the multiple data fragments of other entangled composites of surgeon, software developer, programmer, neural network, patient's body, images, and so on. The singular cut, within which teems a multiplicity, is present also in the autonomous vehicle, the drone, the smart borders system—always also with multiple data fragments lodged within.[23] In every singular action of an apparently autonomous system, then, resides a multiplicity of human and algorithmic judgments, assumptions, thresholds, and probabilities.

The Impossible Figure of the "Human in the Loop"

The neural network's capacity to learn by extracting features from its data environment has made it flourish in the algorithmic architectures of drones, autonomous vehicles, surgical and production robotics, and at the biometric border.[24] This capacity to learn something in excess of taught rules has also characterized the public concern and ethical debates around autonomous systems. Whether in the neural net algorithms animating surgical robots, autonomous weapons systems, predictive policing, or cloud-based intelligence gathering, what is most commonly thought to be at stake politically and ethically is the degree of autonomy afforded to machines versus humans as a locus of decision. I suggest, however, that the principal ethicopolitical problem does not arise from machines breaching the imagined limits of human control but emerges instead from a machine learning that generates new limits and thresholds of what it means to be human. As legal cases proliferate amid the errors, when the spring of action happens at the point of surgical incision, smart border, or drone strike, they consistently seek out an identifiable reasoning human subject to call to account: a particular named surgeon, a specific border guard, an intelligence analyst—the "human in the loop."

In Hayles's field-defining book, *How We Became Posthuman*, she proposes that the "distributed cognition of the posthuman" has the effect of complicating "individual agency."[25] Hayles does not argue that a historically stable category of human has given way, under the forces of technoscience, to an unstable and disembodied posthuman form. On the contrary, the conception of the human and human agency was, and is always, a fragile and contingent thing. As Hayles writes,

> The posthuman does not really mean the end of humanity. It signals instead the end of a certain conception of the human, a conception that may have applied, at best, to that fraction of humanity who have had the wealth, power, and leisure to conceptualize themselves as autonomous beings exercising their will through individual agency and choice. What is lethal is not the posthuman as such but the grafting of the posthuman onto a liberal humanist view of the self.[26]

Hayles's concerns for the grafting of the posthuman onto the figure of an autonomous liberal subject echo across the making of intuitive machine learning worlds. Though technologies such as Intuitive Surgical's robot actively distribute and extend the parameters of sight, touch, and cognition into posthuman composite forms, their ethical orientation is defined solely in relation to the control of an autonomous human subject. While human surgeons speak of an indeterminate *we* who learns, decides, and acts, nonetheless the capacity for judgment retains its Kantian location in the unified thought of a reasoning human subject.[27] Thus, when a violence is perpetuated or a harm is registered—damage, prejudicial judgment, or death—the only ethical recourse is to an imagined unified entity who secures all representations. So, for example, in a series of legal cases against Intuitive Surgical, the reported harms include the rupture of tissue, burns, and other damage to organs, severed blood vessels and nerves, loss of organ function, and fatalities.[28] In these juridical cases, where the robot's machine learning algorithms fail to recognize or to grasp precisely the outline of the target, what is sought is a unified locus of responsibility—a company, a negligent surgeon, or a hospital—an entity imagined juridically to be autonomous and unified, whose choices and agency can be held to account. Similarly, when autonomous weapons systems make errors in their target selection, or cause "collateral damage" amid the so-called precision strike, the ethical appeal is made to an accountable "human in the loop" of the lethality decision.[29] The notion of an ethical decision thus appears in the form of a reasoning human subject or a legal entity with a capacity to be a first person *I* who is responsible.

Yet, where would one locate the account of a first-person subject amid the limitless feedback loops and back propagation of the machine learning algorithms of Intuitive Surgical's robots? When the neural networks animating autonomous weapons systems thrive on the multiplicity of training data from human associations and past human actions, who precisely is the figure of *the* human in the loop? The human with a definite article, *the* human, stands in for a more plural and indefinite life, where humans who are already multiple generate emergent effects in communion with algorithms.[30] Recalling Geoffrey Jefferson's 1952 question, "Who was learning, you or the machine?," and Turing's reply, "We both were," the human in the loop is an impossible subject who cannot come before an indeterminate and multiple *we*.

Perhaps what is necessary is not a relocated human ethics—of feedback loops and kill switch control—for a world of the composite actions of human and algorithm. What is necessary, I propose, is an ethics that does not seek the grounds of a unified *I* but that can dwell uncertainly with the difficulty of a distributed and composite form of being. As machine learning changes the relations we have to ourselves and to others, the persistent problems of a Kantian unity of thought is newly dramatized by algorithmic formulations of learning and acting. To begin to address this different kind of ethicopolitics, one must dwell with the difficulty, as Donna Haraway suggests, making cloudy trouble for ourselves methodologically and philosophically.[31] Such a tracing of algorithmic threads as they meander through unilluminated space involves asking questions of how algorithms iteratively learn and compose with humans, data, and other algorithms. To be in the dark, to dwell there in an undecidable space, is to acknowledge that our contemporary condition is one in which the black box of the algorithm can never be definitively opened or rendered intelligible to reveal its inner workings. To trace the algorithm in the dark is not to halt at the limits of opacity or secrecy, but to make the limit as threshold the subject of study. Such a task begins by asking how machine learning algorithms learn things about the world, how they learn to extract features from their environment to recognize future entities and events, what they discard and retain in memory, what their orientation to the world is, and how they act. If intuition never was an entirely human faculty, and never meaningfully belonged to a unified *I* who thinks, then how does the extended intuition of machine learning feel its way toward solutions and actions? To this task I now turn.

Regimes of Recognition: How a Neural Network Makes the World

Allow me to begin by describing a scene—a laboratory designing machine learning algorithms for border and immigration control systems—where a series of neural networks learn to recognize people and things via the features in their data environments. One of the designers explains that his algorithms are trained on border and immigration data with many hundreds of thousands of parameters. He describes how he “plays with” his developing neural nets—taking the experimental model to the uniformed border operations team in the adjoining building to test it against the specific targets they are seeking in the algorithm’s output.[32] This traveling of the model between laboratory and operations center consists of a series of questions about whether the algorithms are useful, or if they are “good enough.” This question, Is it good enough?, illuminates some of the politically contested features of the algorithm’s emergence. Though for the border operations team, “good enough” may be a measure of the algorithm’s capacity to supply a risk-based target for a decision at the border, for the computer scientists, “good enough” means something quite different and specific. In computer science, a “good enough” solution is one that achieves some level of optimization in the relationship between a given target and the actual output of a model.[33] Understood in this way, it is not the accuracy of the algorithm that matters so much as sufficient proximity to a target. Put another way, the algorithm is good enough when it generates an output that makes an optimal decision possible. When the algorithm designers describe tuning or playing with the algorithm, they are experimenting with the proximity between the target value and the actual outputs from their model, adjusting the probability weightings in the algorithm's layers and observing how the actual risk flags generated by their model diverge or converge on the target.

The design of an algorithmic model, then, involves a contingent space of play and experimentation in the proximities and distances between the actual output and a target output. My concept of the space of play designates specifically the distance between a target output and an actual output of the model. This space of play, however, also opens onto an infinite array of combinatorial possibilities in terms of the malleable and adaptable inputs, parameters, and weights of the model. As one designer of machine learning algorithms for anomaly detection frames the question, "What is normal?" and "How far is far, if something is to be considered anomalous?"[34] Precisely this adaptive threshold between norm and anomaly was being negotiated between the laboratory and border operations. A small adjustment in the threshold will generate an entirely different set of outputs and, therefore, a change in the spring of action. I have observed this iterative process of playing with the threshold in multiple situations where algorithms are being trained for deployment, from police forces adjusting the sensitivity of a facial recognition algorithm to casinos moving the threshold for potential fraud:

> You must experiment to determine at what sensitivity you want your model to flag data as anomalous. If it is set too sensitively, random noise will get flagged and it will be essentially impossible to find anything useful beyond all the noise. Even if you've adjusted the sensitivity to a coarser resolution such that your model is automatically flagging actual outliers, you still have a choice to make about *the level of detection that is useful to you*. There are always trade-offs between finding everything that is out of the ordinary and getting alarms at a rate for which you can handle making a response.[35]

What is happening here is that the neural networks are learning to recognize what is normal and anomalous at each parse of the data. But, the shifting of the thresholds for that recognition embodies all the valuations, associations, prejudices, and accommodations involved in determining what is "useful" or "good enough." Sometimes, as with semisupervised machine learning, this regime of recognition emerges iteratively between humans, algorithms, and a labeled training dataset. In other instances, unsupervised machine learning will cluster the data with no preexisting labeled classifications of what is or is not useful or of interest. Even in this apparently unsupervised process, humans recalibrate and adjust the algorithm's performance against the target. In short, in all cases, machine learning algorithms embody a regime of recognition that identifies what or who matters to the event. Machine learning algorithms do not merely recognize people and things in the sense of identifying—faces, threats, vehicles, animals, languages—they actively generate recognizability as such, so that they decide what or who is recognizable as a target of interest in an occluded landscape. To adjust the threshold of what is "good enough" is to decide the register of what kinds of political claims can be made in the world, who or what can appear on the horizon, who or what can count ethicopolitically.

The kind of ethicopolitics I am opening up here is somewhat different from the attention that others have given to the inscription of racialized or other prejudicial profiles in the design of the algorithm.[36] Though identifying the human writing of prejudicial algorithms as a site of power is extraordinarily important, the regimes of recognition I have described actively exceed profiles written into the rules by a human. The machine learning algorithms I observed, from borders to surgery, and from facial recognition to fraud detection, are producing modes of recognition, valuation, and probabilistic decision weighting that are profoundly political and yet do not reside wholly in a recognizable human who writes the rules. They are also, of course, calculative spaces where prejudice and racial injustices can lodge and intensify, though not in a form that could be readily resolved with a politics of ethical design or the rewriting of the rules.[37]

The Hitherto Unseen: Detecting Figures, Detecting Objects

If machine learning algorithms are changing how something or someone comes to attention for action, then how does this regime of recognition come into being? Often the most apparently intuitive of human actions—to recognize a face in a crowd, to distinguish the features of a cat from a dog, to know how best to reach out and grasp an object—present some of the most difficult computational problems. When algorithms are understood as a series of programmable steps formulated as "if . . . and . . . then," it is precisely in the writing of the rules for the sequence that one decides the result: who or what will be of interest, or who or what can be recognized.[38] A common exemplar of the problem of recognizing the unseen is the capacity of algorithms to recognize handwritten digits.[39] The variability of the form of a figure—its "profile"—exposes the limit of rules-based algorithms that define the features in advance. How would one formulate rules for recognizing the handwritten number 3? In the traditional decision trees designed by J. R. Quinlan in the 1980s, the "product of learning is a piece of procedural knowledge that can assign a hitherto-unseen object to one of a specified number of classes."[40] To recognize an unseen object, the decision tree algorithm classifies it according to "a collection of attributes" describing its important features.[41] How would one begin to define the attributes of the figure 3 amid the variability of its form? As Quinlan acknowledged, the limits of the procedural knowledge of rules-based classifiers are encountered in unknown features that were absent in the training dataset. "The decision trees may classify an animal as both a monkey and a giraffe," wrote Quinlan, or the algorithm may "fail to classify it as anything."[42] Procedural classifiers such as decision trees, then, learn to recognize hitherto unseen objects according to the presence or absence of a set of properties encountered in the training data. Returning to the example of handwritten digits, would the variability of the form of the figure result in the decision that a 3 is not a 3? Or indeed that a 3 is not classifiable, or has the attributes of a 5? Though the recognition of handwritten figures is a common exemplar, the limit of procedural classifiers also applies to many other recognition problems, from facial recognition technologies to security threats, weather patterns, advertising opportunities, or the likely pattern of votes in an election. Put simply, while a rules-based classifier recognizes according to the profiled properties of an entity, the contemporary neural network algorithm learns to recognize via the infinite variability of features it encounters.

What does it mean to learn about the world in and through the variability of features in the environment? With the growing abundance of digital images and cloud data for training machine learning algorithms, the process of learning shifts from *recognition via classification rules* to *recognition via input data clusters*. Let us explore this further through the example of recognizing numeric figures. Figure 2.5 shows a simplified illustration of the spatial arrangement of a deep (multilayer) neural network—of the type I have sketched on many whiteboards at conferences and workshops. If the target of the algorithm is to optimize the likelihood of correctly identifying a handwritten digit, for example, then the training data will consist of a dataset of handwritten digits, each figure segmented to the level of pixels. What this means is that the algorithm does not learn to recognize the profile of the figure 3 per se, but rather learns to recognize the clustered patterns in the array of pixels in the image.[43] The input data in the neural net—and consider that in other instances, this could be anything: images, video, biometric templates, social media text—is assigned a series of probability weightings for its significance, with the output

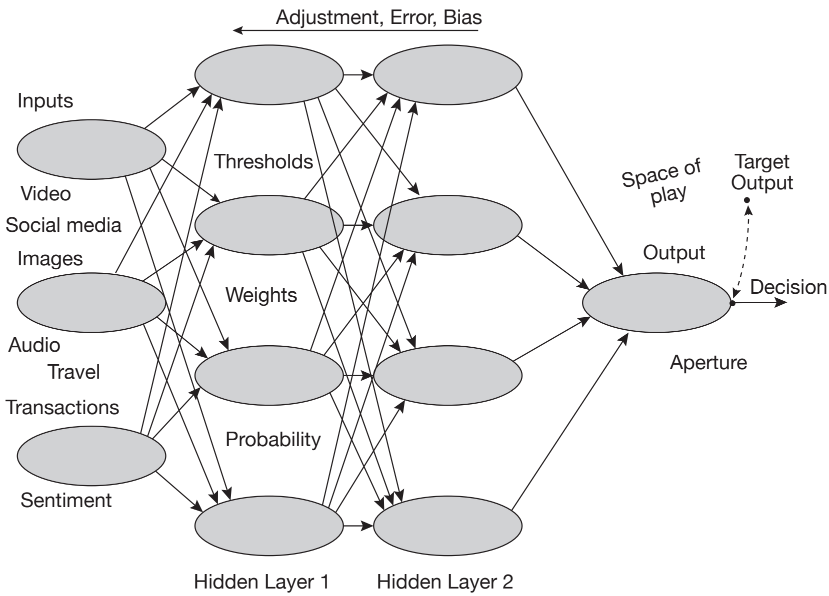


Figure 2.5 A representation of the arrangements of a deep neural network.

of each neuron to the next "hidden layer" dependent on whether the weighted sum of inputs is less than or greater than some threshold value. Each layer of the neural network detects some discrete aspect of the figure. As the computer scientists describe image recognition, "the learned features in the first layer" detect "the presence or absence of edges," with the second layer "spotting particular arrangements of edges," and the subsequent layers "assembling motifs into larger combinations that correspond to parts of familiar objects."[44] The recognition of edges, motifs, and familiar arrangements is not designed into rules by a human engineer but is definitively generated from the exposure to data. To be clear, this spatial arrangement of probabilistic propositions is one of the places where I locate the ethicopolitics that is always already present within the algorithm. The selection of training data; the detection of edges; the decisions on hidden layers; the assigning of probability weightings; and the setting of threshold values: these are the multiple moments when humans and algorithms generate a regime of recognition.

Adjusting the Features: The Variability of What Something Could Be

The computational problem of how to recognize people and things has become of such commercial and political significance that computer scientists enter their experimental algorithms in competitive image recognition contests. One particular algorithm, the AlexNet deep convolutional neural network, won an image recognition contest in 2012 and has become the basis for multiple subsequent commercial and governmental recognition algorithms, with the scientific paper cited more than twenty-four thousand times.[45] The AlexNet gives an account of itself—in the terms of a partial account I am advocating—that manifests just how its output is contingent on its exposure to data features, and the series of weightings, probabilities, and thresholds that make those features perceptible. As I recount something of how the AlexNet algorithm does this, I would like you to consider that if an algorithm is deciding "Is it a leopard?" or "How likely is it that this is a shipping container?" then it is also deployed to decide "Is this a face?" and "Is this face the same face we saw in the street protest last week?"[46] Understood in this way, the regime of recognition is political in terms of both arbitrating recognizability and outputting a desired target that is actionable.

As the computer scientists who designed AlexNet describe the relationship between recognition and cloud data, "objects in realistic settings exhibit considerable variability, so to learn to recognize them it is necessary to use much larger training sets."[47] The AlexNet CNN was trained on 15 million images, each image labeled by a human via Amazon Mechanical Turk's crowdsourcing labor tool.[48] Figure 2.6 shows eight test images for the algorithm, with the five labels considered most probable by the model assigned beneath each image. The algorithm is able to recognize a previously unseen image of a leopard or a motor scooter because the feature vectors of the image have close proximity to the gradients encountered in the training data. The capacity of the algorithm to recognize an incomplete creature at the edge of the frame (the mite) is considered to be a major advance in neural nets for image recognition. Where the algorithm failed to recognize an entity—the grille, the cherry—the scientists refer to the "genuine ambiguity" of which object is the focus of the image, and where the patterns of edges are occluded (e.g., the Dalmatian's spots and the cherries). To be clear, the logic of the AlexNet algorithm is that if one exposes it to sufficient data on the variability of what a leopard could be, then it will learn to anticipate all future instances of leopards. Indeed, deep neural nets are exposed to infinite variabilities—voting behaviors, faces in crowds, credit histories, kidney tumors, social media hashtags—to recognize the feature vec-

Figure 2.6 Test images of the AlexNet image recognition algorithm. Krizhevsky, Sutskever, and Hinton, "ImageNet Classification."

tors of all future instances. Whether someone or something can be recognized depends on what the algorithm has been exposed to in the world. Since the algorithm makes itself—adjusts thresholds and weights, for example—through its exposure to a world in data, it is becoming the contemporary condition of recognizability as such.

Like the cloud chambers of chapter 1, the propositional arrangements of the neural net are instruments of mattering, methods for making some things matter more than others. To seek to open or to make transparent the black box of this arrangement would be neither possible nor desirable, for the arrangement *is* an important site of politics, the spatiality of the calculus being politically significant in and of itself. This is significant for the ethicopolitical interventions one might wish to make because, for example, it could never be sufficient to demand that facial recognition algorithms that fail to recognize black faces be trained on a greater variability of images. For the algorithm also learns how to afford weight or value to one pixelated part of an image over others (the Dalmatian and not the cherry, the edges of this face and not that one). Indeed, as one computer scientist explained to me, a neural net like AlexNet, with six or eight hidden layers, is too complex even for the designer of the algorithm to explain the conditional probabilities that are learned. "I might adjust the weighting in that layer," he explains, "and I know that this will change the output, but I cannot say exactly how."[49] As with the design of AlexNet, the computer scientists work with the essentially experimental and unknowable nature of the algorithm. They perceive the fractional changes in the output of the model as they adjust the weightings, working with the emergent and unknowable properties of machine learning.

Bias Can Be a Powerful Ally

When deep neural network algorithms learn, then, they adjust themselves in relation to the features of their environment. To be clear, to learn, they have to weight some data elements of a feature space more than others—they have to have assumptions about how the world is ordered. Notwithstanding the widespread societal calls for algorithms to be rendered free of bias or to have their assumptions extracted, they categorically require bias and assumptions to function in the world. Indeed, even the textbooks used by the next generation of computer scientists address directly that "there can be no inference or prediction without assumptions," particularly the assumptions of "the probability assigned to the parameters."[50] Thus, when a team of European computer scientists discuss how they might move the threshold for their neural net algorithm to recognize the likelihood of a person of interest (a future person, yet to arrive) being a "returning foreign fighter" and not a "returning aid worker" from Syria, they mean that they will adjust the sensitivity of the algorithm to particular elements of weighted input data, such as increasing the probability weighting of particular past flight routes.[51] While some of this adjustment of the threshold is done by humans, today much of it is invested in the power of the algorithm to adjust itself in and through the emergent properties of the data, understood as a feature space. The "we both" of Turing's reflections seems to reassert itself here in the accounts of adjustment given by computer scientists Yann LeCun, Yoshua Bengio, and Geoffrey Hinton of Facebook AI, Google, NYU, and the University of Toronto: "We compute an objective function that measures the error (or distance) between the output scores and the desired pattern of scores. The machine then modifies its internal adjustable parameters to reduce this error. These adjustable parameters, often called weights, are real numbers that modify the input-output function of the machine. In a typical deep-learning system, there may be hundreds of millions of these adjustable weights."[52]

Like the abductive methods of intelligence gathering I discuss in chapter 1, this computational method observes the effect of the calculation—or the output signal—and theorizes back to the adjustment of parameters, like the mechanical knobs on a calculating machine. The distance between an agreed target output, or desired pattern, and the output scores is the *error* or the *bias*. Significantly, for the algorithm, *error is distance*; it is the playful and experimental space where something useful or "good enough" materializes. There is nothing normatively wrong about error in a machine learning algorithm, for it is a reduction of difficulty and difference into a mere matter of distance. Likewise, let me be clear, bias and weighting are not negative things for an algorithm. They are, on the contrary, essential elements of learning, so that, in computer science, "bias can be a powerful ally."[53] My point is that one could never satisfactorily address the ethicopolitics of algorithms by calling for a removal of human or machine bias and a reduction of error because the machine learning algorithm would cease to function at this limit point. Bias and error are intrinsic to the calculative arrangements—and therefore also to the ethicopolitics—of algorithms. At root, the algorithm can never be neutral or without bias or prejudice because it must have assumptions to extract from its environment, to adapt, and to learn. It is, ineradicably and perennially, a political being. To begin from here is to begin from the idea that all machine learning algorithms always already embody assumptions, errors, bias, and weights that are fully ethicopolitical. In the adjustment of parameters one can locate a shifting terrain of the relations of oneself to oneself and to others. The output of the algorithm is but a mere numeric probability, fragile and contingent, so that a tiny adjustment of the weights in the algorithm's layers will radically change the output signal, and with it the basis for decision and action.

Point Clouds and the Robot's Grasp

To extract something from the features in a data environment, to anticipate and to act, is a critical computational problem for deep machine learning algorithms in production line robotics, surgical robotics, drones, and IED (improvised explosive device) detection. Across these diverse domains, the capacity to recognize the three-dimensional form of an object and to decide on the optimal action is a challenge that animates computer science. Indeed, the failure to recognize multidimensional and mobile forms—such as those of human organs, vehicles, or facial features—has been a common feature of many highprofile mistakes and accidents by machine learners. In the fatal Tesla autonomous vehicle crash of 2016, for example, one way to articulate the error would be to say that the CNN algorithms failed to recognize the profile of a white van against a pale sky as the vehicle turned across the Tesla's path. The probabilistic answer to the question "Is this a vehicle?" was, fatally, "no." Discussions among computer scientists regarding the causes of such accidents are revealing in terms of a persistent determination to locate the source of the fatal flaw and to annex the algorithm from its milieu. For example, one group of IBM scientists urged caution "not to blame the algorithm for a failure of the sensors, the ambient lighting, or the human operator."[54] My point, though, is that what the sensor can sense, or the operator can decide, is only meaningful in the context of how the neural nets arbitrate what the objection could be, what it could mean. Similarly, in a Volkswagen factory in 2015, a robot failed to recognize the outline of a human coworker on the production line, mistaking him for a car door and crushing him to death. At the level of machine learning algorithms and their regimes of recognition, the da Vinci surgical robots' failures to recognize the boundary delimiting kidney tumor from human organ is not dissimilar to the biometric facial recognition systems that closed automated border controls at an airport when the setting sun changed the ambient lighting conditions. In all these instances, the algorithms have learned from the features of the environment they have been exposed to. Sometimes events and sensors in the environment will present them with a set of input features they have not encountered previously, and their assumptions and weightings may lead to a *spring of action* that misrecognizes the target.[55] Is this an error? Or is error merely a matter of distance?

The contemporary advent of cloud robotics has sought to address this problem of the limit point of exposure to features in a multidimensional environment. Cloud-based robotics, as we saw in the discussion of surgical robots, circulate data and aggregate computational power across a distributed system of machine and human learning. Just as the surgical robots are no longer limited to the data and computation stored within a bounded system, so the cloud-based intelligence system I discuss in chapter 1 recognizes its targets from exposure to data and analytics methods across borders and jurisdictions. Where machine learning intersects with cloud computing, the neural network algorithms are exposed to features from a vast archive of cloud data, including the Point Cloud Library of open source 2D and 3D images.[56] A point cloud is a set of topological data points mapping the 3D space of objects. Computer science research in cloud robotics is addressing the question of whether exposure to a vastly increased volume of point cloud data on objects can optimize the neural network's capacity to learn how to recognize and to act. Consider, for example, the computer science team at UC Berkeley's Automation Sciences Lab, whose research into cloud robotics is funded by the NSF, Google, and the US Department of Defense. Presenting their Dex-Net 1.0, or Dexterity Network algorithm, the Berkeley scientists experiment with CNNs to optimize the capacity of a robot to recognize and grasp a range of objects. The algorithm represents an advance on image recognition technologies such as AlexNet because it recognizes 3D objects from multiple viewpoints, and it outputs an optimal action based on this recognition.[57]

The Dex-Net algorithm is trained on an archive of Google point cloud data on "3D object models typically found in warehouses and homes, such as containers, tools, tableware, and toys."[58] The neural nets are learning to recognize the object's geometry and topology and then to optimize the robot's capacity to reach out and grasp the object. In a sense, the algorithm is asking a two-step question—What is this object? and How can it be most effectively grasped? This shift toward CNNs that can recognize *and* optimize an action is absolutely critical in the advance of robotics in manufacturing, medicine, and the military. The scientific papers on Dex-Net show something of the logics at work in coupling regimes of recognition to what I call a spring of action. When the Dex-Net algorithm is exposed to a training dataset of one thousand 3D point cloud objects (in figure 2.7, a household spray bottle), it is not able to find a "nearest neighbor" object that will allow it to recognize the query object. When the algorithm is exposed to the point cloud features of ten thousand objects drawn from the Point Cloud Library, however, it finds two proximal objects, or nearest neighbors, allowing it to recognize the object and optimize its grasp.

Though at first glance the Dex-Net's machine learning may appear as though the cloud is supplying "big data" volume to the algorithms, in fact the process of reduction and condensation I describe in chapter 1 is also taking place here. As the computer scientists propose, the significance of the "cloud-based network of object models" is actually to "reduce the number of samples required for robust grasps" and to "quickly converge to the optimal grasp."[59] Put simply, the Dex-Net algorithm is better able to condense and filter out the occlusions to recognize the most similar object. Each of the ten thousand cloud-based objects is prelabeled with 250 parallel-jaw robot grasps, each weighted with a probability of a successful grasp. "The goal of cloud robotics," as the computer scientists explain, is to "pre-compute a set of robot grasps for each object" so that "when the object is encountered, at least one grasp is achievable in the presence of clutter and occlusions."[60] Understood in these

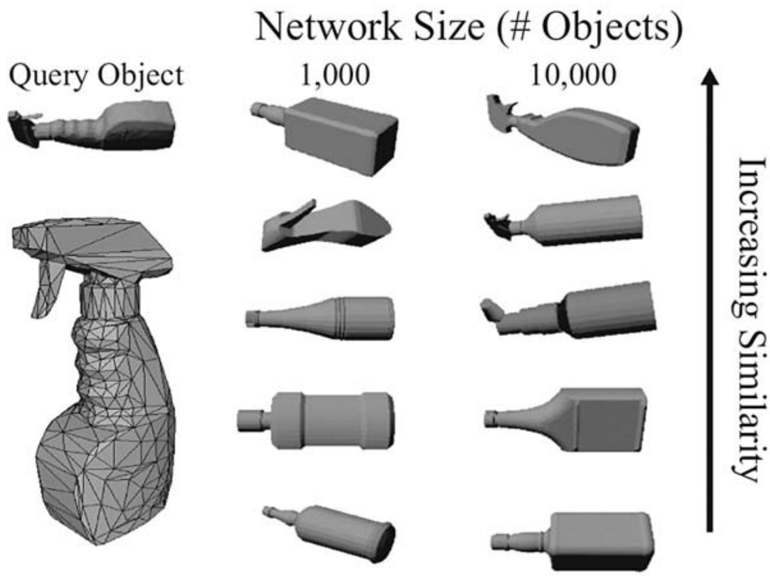


Figure 2.7 Point clouds for object recognition and grasp optimization. Mahler et al., "Dex-Net 1.0."

terms, the weightings and probabilities of the point cloud make it possible to precompute something, so that the encounter with the unknown object can always yield an action that is optimal. With the point cloud, an algorithmic system does not only ask "Is this a face?," "Is this a bottle?," "Is this a military vehicle?," but rather it has already precomputed an optimal action in relation to the topologies it has encountered.

Precomputation: "The Fundamental Thing Is We Know What Good Looks Like"

On the horizon of research in neural net algorithms for robotics, then, one finds this notion of *precomputation* to make an action achievable amidst *clutter* and *occlusions*. Precomputation captures the neural net computational problem that extends from the recognition and grasp of the shape of a human organ amid "noisy" surgical data, to the recognition of a civilian body in the screened occluded data of the drone. One computer scientist, who now designs algorithms for human gait recognition, described to me how he had been "working with surgeons modeling the perfect operation," this human-algorithm collaboration itself deploying point clouds to precompute the trajectory of "a perfect procedure."[61] He has developed his algorithms for multiple applications, with each iteration another step in optimization: "The fundamental thing is we know what good looks like," he explains, "and then if you're doing anomaly spotting, you can see when something's wrong." For example, when the online gambling platform BetFair sought algorithms to recognize and act on patterns of addition, the designer suggests, "that's something we can spot because we know what addicted play looks like." This malleable normative assumption of "what good looks like" or "what addictive play looks like" is generated in and through the algorithm learning in a feature space.[62]

Precomputation implies that some sense of what is a "perfect procedure" or "at least one achievable grasp" action is always already present within the algorithm as such. The ethicopolitics of machine learning algorithms like AlexNet and Dex-Net is in the bias, weights, thresholds, and assumptions that make recognition precomputable. To precompute is to already be able to recognize the attributes of something in advance, to make all actions imaginable in advance, to anticipate every encounter with a new subject or object, a new tumor or terrorist, by virtue of its proximity to or distance from a nearest neighbor. The condition of possibility of the algorithm's action is its exposure to an archive of cloud data, condensed via the infinitely malleable value system of weights, probabilities, thresholds, and bias.

This is a pressing problem of the politics of algorithms in our contemporary moment. All our handwritten digits; all our online data traces; the biometric templates of our facial geometry; the point clouds of household objects and military hardware; all the movements of the hands, eyes, and bodies of surgeons, pilots, soldiers, consumers, production line workers: these are the teeming conditions of possibility of the machine learning algorithm. We are it, and it is us. We could never stand outside it, even if we might wish to. Each of the data fragments that enters the point cloud has a part to play in the learning. Whose is the grasp that caused the injury? Which of the 2.5 million objects in the archive became the nearest neighbor? Which of the possible 250 grasps for each object? Which of the many tens of thousands, or millions, of cloud-derived probabilities was responsible for the grasp that intuitively decided to pull the trigger, so to speak?

The harms inflicted through machine learning are not located primarily in the ceding of human control to machines, as is so often assumed in the ethical and moral debates on algorithmic decisions. Indeed, as we have seen via the surgeon who learns to reach and touch differently with her da Vinci robot, what it means to be human is significantly transformed in and through the machine learning algorithm. To appeal to the human as locus of ethics, then, is to appeal to a being already entangled with new forms of knowing and learning. The principal harm, in contrast, is manifested instead in a specific threat to a future politics. The tyranny of proliferating machine learning algorithms resides not in relinquishing human control but, more specifically, in reducing the multiplicity of potential futures to a single output. The claim to precompute the future, or to know "what good looks like" at the border, in the operating theater, in the economy, forecloses other potential futures. To be clear, the neural net does not reduce multiplicity as such. After all, as I have outlined, the spatial arrangement of the neural net algorithm contains within it multiple probabilities, infinitely adjustable weights whose emergent effects can never be entirely known, even to the designer. The finite elements of each hidden layer of the neural net, one might propose, contain within them infinite possible correlations to other elements. The spatial arrangement of the neural net does not foreclose alternative readings, different arrangements of what or who matters and what or who does not. Crucially, however, at the point of action, this intrinsic multiplicity is reduced to a single output. The insistence on a single output is the algorithm's orientation to action. Though I am reminded by the computer scientists that the output need only be between 0 and 1, and that there are infinite numbers between 0 and 1, there is nonetheless a single numeric output. Let us not forget that the algorithm's output signal lies behind the risk score at the border, the credit decision, the target assessment of the drone, and the decision on sentencing, detention, or the incipient dangers of a gathered protest on a city street. It is as though all the many potentials held in parallel, simultaneously distributed across the layers of the neural net, could never have been. With the output of the machine learning algorithm, one might say, things could never have been otherwise. The output is a probability whose value is transformed by the smallest of adjustments in the parameters of the model. And yet, nonetheless, all political uncertainty is rendered tractable on the horizon of the action triggered by this single output.

At this point, one might reasonably ask how giving such an account of the contingent politics of machine learning algorithms is of any possible critical use. How might a cloud ethics work with the incompleteness, the undecidability, and the contingency of the algorithm's space of play? If one wants to inquire whether a given algorithm is responsible for a flash crash in the financial markets, or if one seeks a human rights law adequate to the task of holding autonomous weapons or autonomous surgery to account, then some ethical grounds might be considered essential—or at least some method for accountability. As Michel Foucault proposes in his discussion of ethics, however, what may be necessary is not to appeal to grounds or to the juridical domain of statutes, but rather to "ask politics a whole set of questions that are not part of its statutory domain."[63] A cloud ethics must be capable of asking questions and making political claims that are not already recognized on the existing terrain of rights to privacy and freedoms of association and assembly. Cloud ethics belong properly not to the individual as bearer of rights, but to the many touch points and data fragments that are aggregated from the relations between subjects and objects. Thus, a cloud ethics must be capable of asking questions such as How did that Dex-Net algorithm weight the probability of that future grasp?; Why did the training data teach the algorithm to recognize this and not that object amid the occlusions?; How was the distance between target and output signal (bias) used as a space of experimentation?; and, In outputting that score, what were the traces of the rejected alternative weights and parameters? Such questions are necessary and urgent, even and perhaps essentially when they are unanswerable. The unanswerable questions reawaken the multiplicity that was, in fact, always present within the machine learning algorithm. All the many contingencies and alternative pathways are reopened, and the single output bears the fully ethicopolitical responsibility for the actions it initiates. The processes and arrangements of weights, values, bias, and thresholds in neural nets are, I think we can safely say, not part of our statutory political domain. And yet, I suggest that they must be presented as questions and political claims in the world.