

Technology of the Surround

Beth Coleman

University of Toronto
beth.coleman@utoronto.ca

Abstract

In addressing the issue of harmful bias in AI systems, this paper asks for a consideration of a generatively wild AI that exceeds the framework of predictive machine learning. The argument places supervised learning with its labeled training data as primarily a form of reproduction of a status quo. Based on this framework, the paper moves through an analysis of two AI modalities—supervised learning (e.g., machine vision) and unsupervised learning (e.g., game play)—to demonstrate the potential of AI as mechanism that creates patterns of association outside of a purely reproductive condition. This analysis is followed by an introduction to the concept of the technology of the surround, where the paper then turns toward theoretical positions that unbind categorical logics, moving toward other possible positionalities—the surround (Harney and Moten), alien intelligence (Parisi), and intra-actions of subject/object resolution (Barad). The paper frames two key concepts in relation to an AI in the wild: the colonial sublime and black techné. The paper concludes with a summation of what AI in the wild can contribute to the subversion of technologies of oppression toward a liberatory potential of AI.

Keywords

artificial intelligence, black techné, ethics, ontology, predictive, surround, supervised learning, unsupervised learning

Meanwhile blackness means to render unanswerable the question of how to govern the thing that loses and finds itself to be what it is not. Stefano Harney and Fred Moten, *The Undercommons: Fugitive Planning and Black Study*

Introduction

My argument is to make AI more wild, not less. By wild, I indicate generative possibility for the technology in opposition to the reproduction of the same. The prompt for this line of inquiry is the call for transparency and accountability as an “ethics” in AI design.¹ I wonder if advocacy toward a *corrective* can produce the ends sought: less harmful bias and more equitable opportunity. What if—outside of the frame of the ethical corrective—one reorients AI application and ontology? I ask that question in looking at two models of AI production—supervised and unsupervised learning. Either modality can be applied toward harm (or benefit), depending on local conditions. And yet, with unerring regularity, AI reproduces systemic harmful bias in its design and application (e.g., Aran et al. 2020; Eubanks 2018; Noble 2018; O’Neil 2016; Raji et al. 2020). With this background in mind, I argue that unsupervised learning offers a potential AI pathway that challenges the reproduction of the status quo that is endemic to supervised learning. In lieu of a corrective, how might we consider an itinerant AI to “think” through things differently than “we” might? By “think” I indicate logics and processes to discovery that might work outside of (predetermined) dominant patterns of data processing as a type of generative collusion—AI as a partner as opposed to a prosthetic. By “we,” I point to the position of the self-determining (human) subject and the legacy of inclusions and exclusions that have informed that position over time.

I ask, in effect, what can AI learn from critical theory? (And reciprocally, what can critical theory learn from AI?) In particular, I engage three concepts toward the thinking of a liberatory function of AI technology: Stefano Harney and Fred Moten’s (2013) Black Studies figuration of the surround, Luciana Parisi’s (2019) techno-philosophic alien intelligence, and Karen Barad’s (2003) atomic intra-actions. The catalyst for bringing together Black Studies, philosophy of technology, and feminist technoscience, respectively, is to reframe artificial intelligence from a technology of oppression that surrounds in its global impact toward a potentially liberatory technology that is not bound to a replication of the past. To this end, I formulate AI—in its ubiquity and degrees of autonomy—as a technology of the surround. The defining features of a technology of the surround are ungovernability and difficulty of defining borders.

In recognizing these attributes, one of the complexities of an AI ethics rests with the dual challenge of the “black box” design of machine learning and ubiquity of its application: there are no clear boundaries. In a black box system, inputs and

outputs are legible, but the internal function of the system remains opaque or “black.” With the ubiquitous application of AI technology, the subject is not in communication with technology (a historical model of human-computer interaction) but the object of machine-to-machine decision making. One might say no to specific instantiations, such as AI-powered “killer robots”—the US Department of Defense’s pilot project for a drone-warfare cloud-processing system under contract with Google (Campaign to Stop Killer Robots, n.d.; Nolan 2020). And yet, the overflow, the interrelations of contracts, permission, obfuscation, and a civic not-knowing often render the ethical local at best. I put the conceptual frame of technology of the surround in play in relation to a discussion of two paradigms of machine learning (ML): supervised and unsupervised learning. My purpose is to investigate AI practices that might move beyond the reproduction of a biopolitics of classification. Biopolitics (with biopower) is a concept that has been developed primarily from Michel Foucault’s (1978) conceptualization of “technologies of power” or control apparatus that enacts at a societal level the sorting and managing of populations. In this sense, predictive AI carries on this legacy of ordering as an extension of societal discipline and control. In regard to AI, I summarize my argument in the following statements:

Thesis A: Supervised learning in AI reproduces the society it mirrors, which is often in the form of a system of eugenics or homophily: these are practices of sorting and prioritizing inherited from pre-computational sciences and reflected in social and political standards (Bowker and Star 2000). It is a practice of classification and the indexing of features that has its origin in the “science” of eugenics (Fisher 1936). Functionally, labeled data is used to train a ML system. Ontologically, supervised learning represents the automated reproduction of a categorical imperative, wherein the conditional for all agents is that of belonging to one category and not another. In this sense, subject and object are distinct in ontology, semiology, and practice. This is a fundamentally binary logic that takes sorting as an a priori—as the precondition for a thing or a state.

Thesis B: Unsupervised learning in AI simulates a monadic environment where an ML system encodes and structures a set of data relations on its own. Formally, data is left unclassified and the task of the ML is to find relations. Functionally, the system designers frame the unsupervised learning inputs and (often) assess and direct system outputs; but there are significant degrees of autonomy and self-determination in an unsupervised learning system. Ontologically, unsupervised learning represents generative possibility, signaled in this argument with the term “wild”: AI is a system of production that potentially works in a logics outside inherited expectation and conditioning (without going so far as to claim a *tabula rasa*). A (potential) primary difference of unsupervised learning is the destabilizing of binary process and outcome.

Machine Learning (ML)

I address second-generation AI, which is primarily characterized by ML with the application of neural networks. A basic description of a machine learning neural network is a computational system with inputs, parallel processing layers that influence each other but are hidden (in the sense of opaque) from system creators, and an output layer. The simple processing elements of the layers can produce complex behavior based on the relation between the processing elements and the system parameters. Based on statistical analytics, the dominant application of ML is predictive modeling. One of the key aspects to the success of modeling is access to big data for training, testing, and application. With the next generation of “sciences of the artificial” (Suchman 2008, 141; Simon 1969), in addition to the AI procedures of ML, one must also attend to the impact of AI as part of an ever-expanding technological array. There is a pronounced empirical aspect of the second generation of AI that enacts a surround: data sniffing, data extracting, data automation are commonplace affordances of the ubiquitous computing arrays that annotate the world, particularly world cities (Coleman 2018; Dourish and Bell 2011). In effect, the human subject is surrounded by a swarm of ubiquitous computing. The pervasive presence of sensor technology (internet of things, array of things, etc.) relates to AI processing in that such arrays feed the ravenous consumption of more data to model the world.

Model 1. Supervised Learning: Finding the White Dog

In “Deep Learning,” their 2015 article in *Nature*, Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, widely understood as the progenitors of the neural net era of AI, describe the process by which machines learn:

The most common form of machine learning, deep or not, is supervised learning. Imagine that we want to build a system that can classify images as containing, say, a house, a car, a person or a pet. We first collect a large data set of images of houses, cars, people and pets, each labelled with its category. During training, the machine is shown an image and produces an output in the form of a vector of scores, one for each category. We want the desired category to have the highest score of all categories, but this is unlikely to happen before training. 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 can be seen as “knobs” that define the input–output function of the machine. In a typical deep-learning system, there may be hundreds of millions of these adjustable weights, and hundreds of millions of labelled examples with which to train the machine. (LeCun, Bengio, and Hinton 2015, 436)

What they outline is a procedure that is deeply complex in computational parallel processing (“hundreds of millions of these adjustable weights”) and reliant on large sets of codified data (“hundreds of millions of labelled examples”) to produce the desired pattern of scores. In the example they give of an image recognition system, they train the machine to disambiguate Samoyeds (large fluffy white dogs) from other animals such as white wolves. The work of training in supervised learning is the classification of data (in this case images) based on features, or a set of quantifiable properties (Alpaydin 2010), such as “white” and “dog,” which must be seen in the feature set as distinct from “white” and “wolf” (LeCun, Bengio, and Hinton 2015). The algorithmic implementation of sorting is called a classifier, which maps input data to a category (Alpaydin 2010). Once the distance between the classification output scores and desired pattern of scores are reconciled, then the system has been sufficiently trained to engage with data “in the wild”—unlabeled images that the AI must identify based on its training. I highlight in this example the normative procedure of supervised learning to train an AI system toward its application “in the wild.” The world is reduced to a particular algorithmic lens that determines how to see the world.

It would be an error not to recognize the intricacy of functions in relation to the granularity of images—at the level of pixel—that the system produces. As LeCun et al. write, “its inputs...are simultaneously sensitive to minute details distinguishing Samoyeds from white wolves—and insensitive to large irrelevant variations such as the background, pose, lighting and surrounding objects” (2015, 438). There is not a theory of mind at work in this condition that attempts to simulate (human) thinking; rather, there is a model of reproduction (coded as probability). The precondition is quantities of data that direct learning toward a predetermined desired pattern: finding the white dog among images of other white canines. There is no world view of dogs and their habitats versus wolves. Nor is there an artificial intelligence animating insights manifested as a notable disruption to patterns of identification. If the AI is working effectively, it will reproduce the “correct” category distinction: dogs are dogs and wolves are not. There is only the automation of sorting (executing decision threshold) across a series of binaries or “weights” toward a correct output. A value above the threshold indicates “dog” and below that “not dog.” It is a powerful system for moving quickly, or optimizing, things that need sorting, such as who gets a loan, or an ad, or an interview, and so on. There is nothing as such that generates new patterns, as the system is designed to replicate predetermined valuations. Might it learn that wolves, as a function of being “not dog,” are wild? Certainly not, as what a wolf might be can only be framed in this paradigm as a partition of given inputs in relation to defined algorithmic analysis. And yet, this narrow framework in which meaning is constructed (or perhaps better said, extruded) is the foundation of predictive models: in a massive, complex, and closed system it learns to replicate as the future the conditions of the past. This process of training

does not always lead to harmful bias. But often it does, as the quotidian event of AI bias is most often a passive state of reproducing the status quo.

Regression and representation are two key aspects of predictive modeling. Both traits make functional pattern recognition. Pattern recognition addresses a statistical model of prediction based on sorting of category membership. Unambiguous category membership has its virtues—for example, when aimed at accurate and speedy identification of pneumonia in a lung X-ray (Adams et al. 2020). But in other contexts, particularly ones steeped in historical exclusions and harm, supervised learning produces a deficit, borrowing from the past to convene the future. Without the necessity of malicious intent, harmful bias will always haunt such a system in the empirical patterns of “big data” culture on which AI relies. If the standard machine vision training relies on massive, free internet search images, then systems trained in a certain era will have an over-indexing of former President George W. Bush: based on available databases and system designers’ lack of incentive, a demonstrated machine vision status quo is North American white male (Huang et al. 2008). In that sense, one witnesses the literal invisibility of black bodies in Global North machine vision systems to which Joy Buolamwini and Timnit Gebru (2018) point or the precarity that the Facebook algorithmic system of seek and expose demonstrates (Mattu et al. 2021). Such erasures and overexposures are symptoms not exceptions of a system design that will not be “fixed” with more diverse training sets or greater transparency of algorithmic design. Until the input/output is recalibrated toward a different end, fixing the training data or algorithm is often at best a post-facto plugging up of holes, “bugs,” and “errors in judgment.”²

And yet, it is not clear that the “black box” is the problem. Rather, one might locate an ontological entropy of AI system design, which is constrained in its reproduction of a biopolitics of hierarchy and valuation. The question of ethics moves from *how* to *toward what end* is AI being aimed. As Solon Barocas, Moritz Hardt, and Arvind Narayanan (2020) note, there is no single or clear path to “fair.” The outcome must be intentional in the design of the system. Without pretending AI in the wild is a panacea, I explore generative AI as a contrapuntal to the predictive. They are not always divergent pathways to an output. Nonetheless, they frame different epistemologies.

AI Theory of Mind

Historically, AI had been rare, exclusive, and narrowly applied. A primary goal was the effective simulation (and surpassing) of human expertise. Recall the chess matches between IBM supercomputer Deep Blue and world champion Garry Kasparov, the first of which Kasparov won in 1996. In the second match played in 1997, Deep Blue beat the Grandmaster (Campbell, Hoane, and Hsu 2002). Implicit in Deep Blue’s design is a theory of mind, a concept adopted by first-generation AI researchers from behavioral and brain sciences that underwrote the imaginaries

of artificial intelligence. Theory of mind frames the ability of the human mind to represent the mental states of others (Call and Tomasello 2008; Premack and Woodruff 1978). It is a theory that addresses the legibility of others' desires and intentions that prioritizes human cognitive behavior in comparison to animals, and in the case of AI, machines (Cuzzolin et al. 2020; Haenlein and Kaplan 2019; McCorduck 1979; Minsky 1986). As such, theory of mind offers another mode of measurement, hierarchy, and sorting mechanism. As Lucy Suchman and other feminist AI scholars have pointed out, theory of mind frames a distinctly conservative view of cognition and what kinds of beings and behaviors are included within its domain.³

In discussing the sociotechnological terms of artificial intelligence, one moves from first-generation AI theory of mind that worked toward the simulation of (human) thinking to the turn toward ML concepts, procedures, and mass implementation that prioritize effective predictive modeling with minimal interest in cognition. In other words, ML deprioritizes cognitive frameworks such as "understanding" and "knowledge" for efficiency, speed, and productive outcome (Anderson 2008). The great claim of second-generation AI is predictive acumen, which trumps mastery of a skill set. The implications of this turn from inherited Enlightenment imaginaries of the *cogito* to the signaling of a machine learning of the neural net points to a paradigm shift: the movement from an ontology of narrow machine intelligence that simulates human expertise to that of a broadly applied ML toolset that is trained on massive data to predict the most likely outcome.

As I have indicated, the predictive model is all too frequently a pernicious model in its reinscription of historical bias. The second-generation revival of artificial intelligence is largely based on an investment in machine learning whose architecture—the function of its functionality—is hidden. That is not a metaphor; it is an actual description of a neural net, which is the transformative system design of the AI surround. Neural networks are described as computational "black boxes," following the logic that while they can execute complex functions, the structure of the neural network will not illuminate the logic of the function. Procedurally, ML functions outside of human supervision. In this sense, one might understand the ML neural net as an itinerant technology; it moves between layers of information, weighing and counter weighing values/features within a prescribed frame. With that said, clearly articulated human frameworks remain critical to AI application—the inputs and (interpretation) of outputs are framed by the system designers.

Model 2. Unsupervised Learning: Mastering the Game of Go without Human Knowledge

If the recursive predictive model of supervised learning tethers pattern, then unsupervised machine learning generates sets of possibilities. The primary difference is that unsupervised learning identifies and “clusters” features through a logic of its own (e.g., “if the conditions of ‘car’ or ‘chair’ can be derived from the observed inputs, then a solution to generating a type of car or chair might follow multiple variations”). Unsupervised learning is wild in the sense of working outside of human parameters of association and prediction, with the clarification that it is the system designers who frame the elements to which the unsupervised learning system is exposed (Coleman 2019). The example of unsupervised learning I address is an AI system to solve the game Go. In the case of the AlphaGo Zero, the self-taught AI Go system, the mode of unsupervised learning is coded as “reinforcement.” As with the general category of unsupervised, reinforcement represents a dynamic, unlabeled computational environment. But the key considerations with reinforcement learning are the goal specificity and the ruleset needed to understand the conditions of that goal—in this case the game of Go and the goal to win by teaching itself and generating skills as it continues to beat its own best game (feedback).

The radical potential of unsupervised learning is a known, even if underexplored, phenomenon in ML. In their *Nature* article on deep learning, LeCun, Bengio, and Hinton (2015, 442) point to the “catalytic effect” of unsupervised learning. Notably, they move from the procedural rhetoric of the predictive to the invocation of analogy—a theory of mind as such—in how machines might learn untethered from pre-trained data. They write, “Human and animal learning is largely unsupervised: we discover the structure of the world by observing it, not by being told the name of every object” (LeCun, Bengio, and Hinton 2015, 442). In their speculative view, classification of data is antithetical to how *nature* models learning—which is described as a process of discovery with formal attributes: “Human vision is an active process that sequentially samples the optic array in an intelligent, task-specific way using a small, high-resolution fovea with a large, low-resolution surround” (LeCun, Bengio, and Hinton 2015, 442). Formally, unsupervised learning uses classifiers to perform *cluster analysis* (grouping objects that are similar in some way and dissimilar to objects in other clusters). But it is the ML system that decides what warrants similarity or dissimilarity. Classifiers modulate in relation to dynamic rules as the conditions of learning are different: the data for the most part are unlabeled, which means the algorithm must find its own structure from the input (Mishra 2017). Unsupervised learning must locate meaning (identify patterns) in the materials to which it is exposed, which does not necessarily coincide with the patterns of association humans would bring to a dataset.

As the authors of AlphaGo Zero write, “Supervised learning systems...are trained to replicate the decisions of human experts...In contrast, reinforcement learning systems are trained from their own experience, in principle allowing them to exceed human capabilities, and to operate in domains where human expertise is lacking” (Silver et al. 2017, 1). In mastering Go without human knowledge, the parameters of learning are still human-framed (i.e., what is Go and what are the rules?). But the process of learning the game does not simulate human expertise. For example, the machinic logic of “best game” technique is *winning* game technique, which is not in this case bound to simulation and prediction of expert human game play. The Monte Carlo tree search the system uses works in reference to self-play, not a priori world of Go play. AlphaGo Zero learns within the parameters (rule system/judgement of winner) of Go as an environment; but it does not simulate human Go play as such. In the three days of training the ML system, AlphaGo Zero “progressed from entirely random moves towards a sophisticated understanding of Go concepts...all discovered from first principles” (Silver et al. 2017, 10). The AlphaGo Zero designers describe a generative, as opposed to simply reproductive, event in which the machine engaged “non-standard strategies” outside of the scope of traditional game play. The authors stake their investment in a ML system that teaches itself to “exceed human capabilities” (Silver et al. 2017, 1). But beyond beating human experts (as stated, a long-standing telos of AI research), AlphaGo Zero demonstrates a quality that speaks to its wildness *outside* of human thinking. It executes “random” moves in the beginning of the learning cycle, demonstrating an active process toward determination that does not present a pre-given conclusion. In other words, the primary epistemological unit is not subject/object but phenomena.

If this can be said of a machinic system, unsupervised AI wanders, collecting and connecting, as it locates the solution horizon. In the sense that it “learns” what it is exposed to, unsupervised learning is itinerant and amoral. A particularly vivid example is unsupervised learning in Natural Language Processing (NLP) training. Unsupervised NLP experiments—such as Microsoft’s Tay and OpenAI’s GPT-3—set the system free to graze across linguistic data, “reading” the internet to gain natural language acumen. It is a process that has produced controversy and curiosity with the startling, ridiculous, and ugly utterances the NLPs have generated (Perez 2016; Metz 2020). In a demonstrated reproduction of the status quo, the internet teaches NLP AI racism, sexism, and other nastiness in record time. And yet, what if one experiments with the idea that such reproduction is not endemic to the system? That it is a design feature as opposed to its architecture? If GPT-3 were reading Franz Fanon and the corpus of anti-colonial anti-oppression literature (not as vast as the internet, but plenty big), it might speak a different language.

Outside of the judgement of good or bad outcomes, unsupervised learning offers an unbounded logic away from narrow conditions of the binary. It is not gauging

“white dog” or “not white dog;” it automates opportunistic clustering. Unsupervised learning offers behaviors outside of a preset condition. The system is not finite (it is also not infinite), in the sense that it can continue spinning off variations as “decisions” (Coleman 2019). In this sense, AI exceeds itself. By design, it generates, versioning possible outcomes until its humans decide which path to follow. The *generation of outcomes* as opposed to the reproduction of preset conditions may be the most experimental and exciting aspect of current AI.

Technology of the Surround

The sociotechnical state of AI sits at an ontological crossroads. The dominant paradigm of predictive AI simulates a command-control system that can be aimed like a weapon—the “killer robots” of a military postindustrial complex as well as the quotidian application of ubiquitous computing. In such a formulation, these are technologies of oppression that continue to power the extractive practices and constitutional imaginaries of a colonial sublime. With the term *colonial sublime*, I signal an event horizon wherein the *mechanisms* by which hierarchies of valuation of life are continuously erased for a violent logic of naturalization. In this sense, the colonial sublime produces its own biopolitic of “black box” logics, obscuring its own mode of reification in the production of technologies of oppression. In light of this protracted liminality, another direction is a turn to the wild—the possibility of an AI increasingly *outside* of a command-control scope. In this sense, AI exceeds itself as a technology of the surround.

A technology of the surround is both ubiquitous and unregulated. It is the manifestation of machine-to-machine communications that leave the human out of the loop in the data chatter. In the array of things—the sensors and other informatic relays—one is literally surrounded. Additionally (historically), a technology of the surround is an itinerant thing that moves at a tempo (adrift) outside of locked-in boundaries. If technologies of power rely on putting things in their proper place, then a technology of the surround presents a contrapuntal, as independent, adjacent, yet still in relation. To best follow the liberatory function of a technology of the surround, one must follow the root system of its genealogy.

Black studies theorists Harney and Moten, in their influential work *The Undercommons* (2013), describe one of their key figures, the surround, as a topos—a space outside of the governance of an Enlightenment legacy. In their text, it is the location in which blackness is unmoored from historical and ontological constraints, as the “thing that loses and finds itself” (Harney and Moten 2013, 49). By configuring the event of blackness—the surround—as “losing and finding,” Harney and Moten hail a long tradition of disruptive positionalities that abandon binaries such as master/slave, subject/object, and society/nature. The subversion of entrenched norms is the very event of “losing and finding” that

happens outside of the light of the fort, the reigning figure in their text of settler colonial empire.

Under various guises, the surround figures broadly into the telos of blackness in the Americas (and recursively in the contemporary world), as it is the space of the underground, where one slips away from the half-lives of the colonial sublime. In thinking the surround as a fertile space in which to decouple AI from the dominance of the predictive model and more broadly from an ontology of technologies of oppression, one encounters the liberatory possibility of black techné, a coalition of an aesthetics, a politics, and a positionality characterized by the itinerant and profoundly iterative. With the most historical relation to black agency, black techné is evident in Harney and Moten's concept of the surround. Yet it also arrives in key concepts of philosopher of technology Parisi and feminist technoscience theorist Barad where the mandate is to accelerate and augment the process of unbinding from a ruthless logic of repetition as reproduction. For Parisi, the site of potentiality is the "alien intelligence" of AI that offers a redirection beyond a reinscription of a cybernetic servo-mechanistic regime. With the Barad, it is the material-discursive "event" that constitutes being in the world—not subject/object but intra-action. This critical trifecta advances a formulation of black techné.

Black Techné and the Colonial Sublime

There would be no surround if not for the colonial sublime of the fort. But the surround is not a reinscription of the dialectical (master/slave). Rather, it is the outcome of escaping it. This complicated liberatory frame of the *subject unmoored* is central to a legacy of black techné. An iconic figure of black techné is the maroon (in French, *le marronnage*), who is the escaped (black) person occluded in the swamps and forests of the Americas. As the preeminent theorist of a poetics of relation, Édouard Glissant (1997) configures the maroon as the subject adrift. In Glissant's analysis, the maroon is a subject position always attached to an ebb and flow, even as it is detached from normative conditions of agency and (by extension) power. Assuming the mantle of Glissant's poetics of relation, Harney and Moten's concept of the surround takes up the maroon in the swamp, in the city, in the academy, in all places where slippage occurs—which is every place—to speak of a tempo of subversion. In this case, tempo is a critical quality of both temporality and the rhythm of a thing. Indeed, black techné as a temporality "loses and finds itself" across worlds of black aesthetics, black politics, and black life.

This is not a subject position but a critical framework of agential instrumentalization, as I have framed in "race as technology" (Coleman 2009). A modality of black techné, race as technology colludes with a sideways logic, the logic of the trapdoor, the escape hatch, the subversion of mastery in the usurpation of signs of power. It is a logics and a poetics of the surround, as such,

that troubles the stasis of the categorical: What if race were understood as a technology as opposed to a pseudo biological historical event (Coleman 2009; Reardon 2017). In this addition of race as technology, the twist of the screw is technology taking the place of the maroon out beyond the floodlights of the fort. This is not a revisiting of Foucault's panopticon, where formally, architecturally, bodies are conscripted to discipline themselves. In fact, it is quite the opposite, where the technology is out in the "wild" and proliferating. With the arrival of ubiquitous AI, the human subject as adjacent to technologies of the surround is brought into relief.

It is in this complex liberatory frame of the subject unmoored that I locate what might be rendered possible in the assumption of AI, which is a logic of the experimental as opposed to the recursive logic of the predictive. AI in the wild—a radical AI—departs from the recursively normative into the surround of the generatively exploratory. To think AI in relation to the maroon—the subject adrift from the dominion of command-control—is a coincidence of history and innovation. Empire locates the telos of technology as innovation—manifest destiny is always progressing and there is no legible collateral damage. Equally, the transatlantic trade in black bodies also evidenced a mode of innovative objectification (the equation of blackness with chattel slavery) that continues to animate the colonial sublime (Gilroy 1993). One can say, "Hold on, black techné, the radical tradition of black aesthetics as black freedom, cannot be equated with mindless machines." And that is certainly true. The murderous equation of the (en)slaved with machine is precisely what the maroons fled from into the swamp and darkness. And yet, the radical turn of AI is toward a technology of the surround—an agent of black techné that disrupts binary. To cite Denise Ferreira da Silva (2017), debunking the transcendental model of self-determination distinguishes a radical engagement from a critical one.⁴ AI addressed as a technology of the surround is a version of wild in concert with Parisi's argument of AI as an alien apparatus increasingly outside of a command-control scope.

Sorting Mechanisms: Alien AI

As a mode of predictive analytics, AI recursion in its data flows literally reinscribes history as the future—the wager of prediction is based on data of what has been before. As Laura Kurgan and collaborators have noted, homophily or heterophily are not preconditions of an analysis but effects of it (Kurgan et al. 2020). Following that logic, AI's reinscription of a eugenicist agenda is central to Wendy H.K. Chun's (2008) critique of software systems, as well as emergently in discourses of critical AI engineering and legal studies (Barocas, Hardt, and Narayanan 2020; Kuhlberg et al. 2020; Richardson, forthcoming). In the current state of design and application, AI carries on the extended, ruthless logic of modernity where technologies of power are sociotechnological sorting mechanisms. It is a persistent manifestation of the colonial sublime that

reinstates machines as the measure of man (and also the category of “not-man” by implication) along a recursive trajectory (Adas 2015). The persistent distinction of subject/object or master/slave traces back to the technological extension of “man” that is continuously enacted as a sorting mechanism. Prosthesis remains the dominant figure of *techné* in Western philosophy (Stiegler 1998); it carries across the historical mechanical arts to modern technology the conceptual framework of appendage in service to the subject (not object), with reinforced boundary markers. And the technological prosthesis as sorting mechanism has led to a profundity of violence evidenced in the automation of all others outside of the illuminated station of subject. As an ontology, the prosthetic continues to extend its reach across technological evolutions of command-control and cyber-servo-mechanistic apparatus.

Moving away from a paradigm of command-control, Parisi offers a view of AI that profoundly challenges the ontology of technology as prosthetic. In reconsidering AI as an alien intelligence, Parisi points to a change of state that moves the technology beyond tool and outside of the domain of what has historically—and increasingly hysterically—been referred to as the “self-determining subject.” In shifting from the paradigm of cyber-servo-mechanism to the alien subject of AI, Parisi signals the change of state from prosthetic to that of alien technology—outside of, adjacent to the transcendental self-determining subject. She queries “whether the servo-mechanic model of technology can be overturned to expose the alien subject of artificial intelligence as a mode of thinking originating at, but also beyond, the transcendental schema of the self-determining subject” (Parisi 2019, 27). In conceptualizing AI as “alien” outside of human control, even as it is of human design, Parisi offers a speculative window on what moving beyond a colonial sublime might portend. Parisi’s logic coincides with black *techné*: AI exceeds itself, loses and finds itself. To this end Parisi states, “However, how to describe an apparatus of capture that runs away from itself, how to understand the dominance of algorithmic forms of subsumption that challenge both the law of the subject and its crisis today?” (2019, 36). In keeping with the Harney and Moten figuration of blackness as the thing that “loses and finds itself,” Parisi summons with the “apparatus of capture”—the very technological modality that is meant to reinscribe the biopolitics of a surveillance state—the ethos of the itinerant. Despite its human maker/master, AI “runs away from itself.” This horizontal logic of exceeding itself in the sense of moving outside of its given ontological domain and toward uncharted territory (the wild, the swamp, the surround, the alien) offers an opening to other possibilities beyond the reinscription of technologies of the artificial that enact a violence of ordering. In citing the ongoing “crisis” of the subject, Parisi locates an opportunity for different relations articulated as living adjacently to technologies of the surround. In considering how such adjacency might be configured, I look to Barad’s account of agency not as a predetermined attribute but as an event with its own temporality and locality.

Categorical Imperative and the Intermittent Event of Becoming

Barad hails a material account of bodies (including bodies of knowledge) as not subject/object but locations of time and place. The frame—the rules of engagement—in this case are quantum physics as derived by Neil Bohr. In calling on the philosophy-physics of Bohr, Barad unbinds events from a categorical imperative in the sense that there is no *a priori* determination of position, e.g., subject/object. Position is determined *of a moment*. Barad describes the liberatory function of a technology of the surround in terms of a materialist agential realism of becoming: “For Bohr, things do not have inherently determinate boundaries or properties, and words do not have inherently determinate meanings. Bohr also calls into question the related Cartesian belief in the inherent distinction between subject and object, and knower and known” (2003, 813). The destabilizing of finite categories continues through the physics of wave/particle and the semiotics of subject/object. In her argument, the indeterminacy at the level of the atomic corresponds with an indeterminacy of language as signification. This is not a version of infinite regress, “turtles all the way down.” Rather, Barad points to tempo, the event of arrival and dissipation.

With this critical invitation to displace a false sense of certainty, Barad offers a logic outside of the categorical that speaks to a wildness of being that cannot be bound to a singular state in advance of the specificity of situation. The “event” as such is locative, particular, and not generalizable. As Barad writes, “Bohr resolves this wave-particle duality paradox as follows: the objective referent is not some abstract, independently existing entity but rather the phenomenon of light intra-acting with the apparatus...The notions of ‘wave’ and ‘particle’ do not refer to inherent characteristics of an object that precedes its intra-action. *There are no such independently existing objects with inherent characteristics*” (2003, 815 FN 21). The assessment at the atomic level is inherent instability that presents as a finite set of possible outcomes: wave or particle depending on the situation. The “intra-action” determines measurable datum, not a categorical imperative. In other words, the primary epistemological unit is not subject/object or “independent objects with inherent boundaries and properties” but phenomena.

At the atomic level, one understands this accounting of phenomena as demonstrated by science, even if one has no first-hand knowledge of atomic becoming. But at the societal scale the phenomenon is not to be believed; the investment of biopolitics is to lock in the subject/object, delineating distinct boundaries with visible markers of an optical regime.⁵ Predictive AI locks in the categorical as a condition of its function, effectively enacting “thingification”—“the turning of relations into ‘things,’ ‘entities,’ ‘relata’” (Barad 2003, 812). Thingification represents an ontology of datafication that enacts abstraction,

eliding materiality and contextual relations, as Donna Haraway (1988), N. Katherine Hayles (1999), and Michelle Murphy (2017) have argued. The outcome of setting things in order sustains a trace relation to histories of violent subjection that black techné troubles, enacting as such a power and politics of radical indeterminacy. In other words, indeterminacy is not exclusively an atomic feature, although the unrelenting regime of the indexical would demonstrate it as so. As Barad points out, neither “things” nor “words” respect a proper boundary. Semiotics had made that evident at the turn of the twentieth century. It has been a slower progression to acknowledge the intra-relation of subject/object among the observable things in the world. In other words, the biopolitics of a categorical imperative continue to play out. What a Kant, really. To map across these territories—the fluidity of atomic phenomena (wave/particle) to the entrenchment of biopolitical regime—is to reflect on and unbind the authentication of binary logic as unerring ground truth.

It is not unknown in the human conception of the world to recognize that the cat may be dead or not dead at the same time (until there is an event that resolves the state); but it is outside of human perception of the world to see possible outcomes as opposed to a given state. And yet the generation of many possible outcomes, as opposed to the reproduction of preset conditions, is exactly what an exploratory AI offers. Its itinerant wildness presents an opportunity to generate other worlds in relation to other types of beings.

Conclusion: AI in the Wild

AI exceeds itself. So very dumb, literally no common sense. And yet, it can be free—if not to imagine then to generate—speeding through possibilities, junctures that are idiotic until they are not. The radical turn at hand is the opportunity to look at artificial intelligence—the machinic making sense of—as a process of ongoing relations, as phenomena as opposed to “knowledge” represented in a database. I have argued that unsupervised learning, in particular, offers a procedural frame that does not inherently reproduce predetermined boundaries. Practically speaking, particularly in regard to the dominant paradigm of supervised learning, the need to audit persists—the “datasheets for datasets” must still be produced—as there is no context for trust and experimentation and there might never be (Geburu et al. 2018). And yet, one can see possible other worlds of AI in the wild. Throughout their work, Gilles Deleuze and Felix Guattari have written of Antonin Artaud’s (1976) infamous (non)figure of the body without organs, giving it a multiplicity of assignations as it is so vividly an unbounded thing that exceeds itself. As they write in *Anti-Oedipus*, “the body without organs is the deterritorialized socius, the wilderness where the decoded flows run free” (Deleuze and Guattari 1983, 176). In thinking technology of the surround as change of state, I interpolate such a narrative of black techné. The entanglement

of AI with the itinerant drift of the maroon and other such creatures of the wild would be a welcome one.

Acknowledgments

I would like to thank Alex Juhasz, Emily Denton, Michelle Murphy, and the *Catalyst* anonymous reviewers for their feedback over the development of the article. Additionally, thank you WUTFA for getting the jokes.

Notes

¹ The young but thriving existence of the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT), speaks to the growing need in AI, legal studies, critical policy, and critical data to address these emerging issues. One can find representative FAccT paper titles such as Barocas, “Problem Formulation and Fairness,” Gebru, “Closing the AI Accountability Gap,” and Hardt, “The Social Cost of Strategic Classification.”

² In this procedural vein, the AI Now *Algorithmic Accountability Policy Toolkit* (2018) is an excellent example of applied methods that relate AI design frames to policy accountability.

³ Suchman (2008) has outlined a feminist counter history of first-generation artificial intelligence from its inception in the 1950s to the early 2000s. One primary aspect of the feminist critique of historical AI, from scholars such as Adam (1998) and Kember (2003), is the theory of mind scientists brought to the discipline. Suchman outlines the critique in the following manner, “AI builds its projects on deeply conservative foundations, drawn from long-standing Western philosophical assumptions regarding the nature of human intelligence” (2008, 142). She points to a primary ethos of feminist technoscience engagement with AI as the exposure of a “politics of ordering” that manifests in binaries such as subject/object, same/other (Suchman 2008, 140).

⁴ Ferreira da Silva critiques the transcendental model of self-determination in the distinction between a radical engagement and a critical one. She writes, “as a category of racial difference, blackness *occludes* the total violence necessary for this expropriation, a violence that was authorized by modern juridical forms—namely, colonial domination (conquest, displacement, and settlement) and property (enslavement). Nevertheless, blackness—precisely because of how, as an object of knowledge, it occludes these juridical modalities—has the capacity to unsettle the ethical program governed by determinacy, through exposing the violence that the latter refigures” (Ferreira da Silva 2017).

⁵ Mirzoeff (2011) in “The Right to Look” and Virilio (1994) in *The Vision Machine*, among other works, have addressed this topic extensively.

References

Adam, Alison. 1998. *Artificial Knowing: Gender and the Thinking Machine*. New York: Routledge.

Adams, Scott J., Robert Henderson, Xin Yi, and Paul Babyn. 2020. "Artificial Intelligence Solutions for Analysis of X-Ray Images." *Canadian Association of Radiologists Journal*. <https://doi.org/10.1177/0846537120941671>.

Adas, Michael. 2015. *Machines as the Measure of Men: Science, Technology, and Ideologies of Western Dominance*. Ithaca, NY: Cornell University Press.

AI Now. 2018. *Algorithmic Accountability Policy Toolkit*. <https://ainowinstitute.org/aap-toolkit.pdf>.

Alpaydin, Ethem. 2010. *Introduction to Machine Learning*. 3rd ed. Cambridge, MA: MIT Press.

Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired*, July 2008.

Artaud, Antonin. 1988. "To Have Done with the Judgment of God." In *Selected Writings*, edited by Susan Sontag, 555–575. Berkeley: University of California Press.

Barad, Karen. 2003. "Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter." *Signs: Journal of Women in Culture and Society* 28 (3): 801–31. <https://www.journals.uchicago.edu/doi/10.1086/345321>.

Barocas, Solon, Moritz Hardt, and Arvind Narayanan. 2020. *Fairness and Machine Learning: Limitations and Opportunities*. E-book. <https://fairmlbook.org/>.

Bowker, Geoffrey, and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of the 1st Conference on Fairness, Accountability and Transparency. *PMLR* 81, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.

Call, Josep, and Michael Tomasello. 2008. "Does the Chimpanzee Have a Theory of Mind? 30 Years Later." *Trends in Cognitive Sciences* 12 (5): 187–92. <https://doi.org/10.1016/j.tics.2008.02.010>.

Campaign to Stop Killer Robots. n.d. Accessed July 7, 2021. <https://www.stopkillerrobots.org/act/>.

Campbell, Murray, A. Joseph Hoane, and Feng-Hsiung Hsu. 2002. "Deep Blue." *Artificial Intelligence* 134 (1–2): 57–83. [https://doi.org/10.1016/S0004-3702\(01\)00129-1](https://doi.org/10.1016/S0004-3702(01)00129-1).

Chun, Wendy H.K. 2008. "The Enduring Ephemeral, or the Future Is a Memory." *Critical Inquiry* 35 (1): 148–71. <https://www.journals.uchicago.edu/doi/10.1086/595632>.

Coleman, Beth. 2009. "Race as Technology." *Camera Obscura: Feminism, Culture, and Media Studies* 24 (1 (70)): 177–207. <https://doi.org/10.1215/02705346-2008-018>.

———. 2018. "Smart Things, Smart Subjects: How the 'Internet of Things' Enacts Pervasive Media." In *The Routledge Companion to Media Studies and Digital Humanities*, edited by Jentery Sayers, 222–29. New York: Routledge.

———. 2019. "Bauhaus Generative: Avant-Garde to Algorithmic Aesthetics in Three Chairs." In *Bauhaus Futures*, edited by Laura Forlano, Molly Wright Steenson, and Mike Ananny, 287–298. Cambridge, MA: MIT Press.

Cuzzolin, Fabio, A. Morelli, Bogdan Ionut Cîrstea, and Barbara J. Sahakian. 2020. "Knowing Me, Knowing You: Theory of Mind in AI." *Psychological Medicine* 50 (7): 1057–61. <https://doi.org/10.1017/S0033291720000835>.

Deleuze, Gilles, and Felix Guattari. 1983. *Anti-Oedipus: Capitalism and Schizophrenia*. Translated by Robert Hurley, Mark Seem, and Helen R. Lane. Minneapolis: University of Minnesota Press.

Dourish, Paul and Genevieve Bell. 2011. *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing*. Cambridge, MA: MIT Press.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Ferreira da Silva, Denise. 2017. "1 (life) ÷ 0 (blackness) = ∞ – ∞ or ∞ / ∞: On Matter beyond the Equation of Value." *e-flux* 79 (February). <https://www.e-flux.com/journal/79/94686/1-life-o-blackness-or-on-matter-beyond-the-equation-of-value/>.

Ferrer, Xavier, Tom van Nuenen, Jose M. Such, Mark Coté, and Natalia Criado. 2020. Bias and Discrimination in AI: A Cross-Disciplinary Perspective. *IEEE Technology and Society Magazine* 40 (2). <https://doi.org/10.1109/MTS.2021.3056293>.

Fisher, Ronald A. 1936. "The Use of Multiple Measurements in Taxonomic Problems." *Annals of Eugenics* 7 (2): 179–88. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>.

Foucault, Michel. 1978. *The History of Sexuality*. Vol. 1, *The Will to Knowledge*. Translated by Robert Hurley. New York: Pantheon Books.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2018. "Datasheets for Datasets." arXiv: 1803.09010. <https://arxiv.org/abs/1803.09010>.

Gilroy, Paul. 1993. *The Black Atlantic: Modernity and Double Consciousness*. New York: Verso.

Glissant, Édouard. 1997. *Poetics of Relation*. Translated by Betsy Wing. Ann Arbor: University of Michigan Press.

- Haenlein, Michael, and Andreas Kaplan. 2019. "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence." *California Management Review* 61 (4): 5–14. <https://doi.org/10.1177/0008125619864925>.
- Haraway, Donna. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14 (3): 575–99. <https://doi.org/10.2307/3178066>.
- Harney, Stefano, and Fred Moten. 2013. *The Undercommons: Fugitive Planning and Black Study*. Wivenhoe, UK: Minor Compositions.
- Hayles, N. Katherine. 1999. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press.
- Huang, Gary, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. 2008. "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments." In *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*. October 2008. <https://hal.inria.fr/REALFACES2008/inria-00321923v1>.
- Kember, Sarah. 2003. *Cyberfeminism and Artificial Life*. London: Routledge.
- Kurgan, Laura, Dare Brawley, Brian House, Jia Zhang, and Wendy H.K. Chun. 2020. "Homophily: The Urban History of an Algorithm." *E-flux Are Friends Electric?* November 2020. <https://www.e-flux.com/architecture/are-friends-electric/289193/homophily-the-urban-history-of-an-algorithm/>.
- Kuhlberg, Jill, Irene Headen, Ellis Ballard, and Donald Martin. 2020. "Advancing Community Engaged Approaches to Systems." White paper. Drexel University, Washington University in St. Louis, June 2020.
- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44. <https://doi.org/10.1038/nature14539>.
- Mattu, Surya, Leon Yin, Angie Waller, and Jon Keegan. 2021. "How We Built a Facebook Inspector." *The Markup*, January 5, 2021. <https://themarkup.org/citizen-browser/2021/01/05/how-we-built-a-facebook-inspector>.
- McCorduck, Pamela. 2004. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. 2nd ed. New York: Routledge.
- Metz, Cade. 2020. "Meet GPT-3. It Has Learned to Code (and Blog and Argue)." *New York Times*, November 24, 2020. <https://www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html>.
- Minsky, Marvin. 1986. *The Society of Mind*. New York: Simon & Schuster.
- Mishra, Sanatan. 2017. "Unsupervised Learning and Data Clustering." *Towards Data Science*, May 19, 2017. <https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a>.
- Mirzoeff, Nicholas. 2011. "The Right to Look." *Critical Inquiry* 37 (3): 473–96. <https://doi.org/10.1086/659354>.

Murphy, Michelle. 2017. *The Economization of Life*. Durham, NC: Duke University Press.

Noble, Safiya. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Nolan, Laura. 2020. "Software Engineer on Ethics in Tech." <https://otia.io/2020/02/27/laura-nolan-software-engineer-on-ethics-in-tech/>.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Parisi, Luciana. 2019. "The Alien Subject of AI." *Subjectivity* 12 (1): 27–48. <https://link.springer.com/article/10.1057%2Fs41286-018-00064-3>.

Perez, Sarah. 2016. "Microsoft Silences Its New A.I. Bot Tay, after Twitter Users Teach It Racism." Techcrunch.com, March 24, 2016. <https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/>.

Premack, David and Guy Woodruff. 1978. "Does the Chimpanzee Have a Theory of Mind?" *Behavioral and Brain Sciences* 1 (4): 515–26. <https://doi.org/10.1017/S0140525X00076512>.

Raji, Inioluwa Deborah, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. 2020. "Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing." In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES '20)*, 145–51. New York: Association for Computing Machinery. <https://arxiv.org/abs/2001.00964>.

Reardon, Jenny. 2017. *The Postgenomic Condition: Ethics, Justice, and Knowledge after the Genome*. Chicago: University of Chicago Press.

Richardson, Rashida. Forthcoming 2022. "Racial Segregation and the Data-Driven Society: How Our Failure to Reckon with Root Causes Perpetuates Separate and Unequal Realities." *Berkeley Technology Law Journal* 36 (3). <https://ssrn.com/abstract=3850317>

Silver, David, Julian Schrittwieser, Karen Simonyan, Ionnis Antonoglou, Aja Huang, Arthur Guez, and Yutien Chen. 2017. "Mastering the Game of Go without Human Knowledge." *Nature* 550 (7676): 354–59. <https://doi.org/10.1038/nature24270>

Simon, Herbert. 1969. *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Stiegler, Bernard. 1998. *Technics and Time, 1: The Fault of Epimetheus*. Translated by Richard Beardsworth and George Collins. Stanford, CA: Stanford University Press.

Suchman, Lucy. 2008. "Feminist STS and the Sciences of the Artificial." In *The Handbook of Science and Technology Studies*, 3rd ed., edited by Edward Hackett, Olga Amsterdamska, Michael Lynch, Judy Wajcman, 139–64. Cambridge, MA: MIT Press.

Virilio, Paul. 1994. *The Vision Machine*. Translated by Julie Rose. Bloomington: Indiana University Press.

Author Bio

Beth Coleman is an Associate Professor of Data & Cities at the University of Toronto, where she directs the City as Platform lab. Working in the disciplines of Science and Technology Studies and Black Studies, her research focuses on machine learning, urban data, and civic engagement.