I. INTRODUCTION
In one of the side wings off the main entry hall of the British Museum there is a little-visited exhibition that takes as its object of analysis the Enlightenment obsession with classification. Here, in Room 1, or the “Enlightenment Gallery”—it is fitting that it is the museum’s first room, constituting the ideological ground on which the rest of its imperialist treasures rest—we find von Linné’s taxonomies on full display, a revelry of classificatory apparatuses, in an Age of Reason that eagerly sought to organize all of reality. But this ordering mania also entailed a veritable Age of the Category, its dialectical, shadowy underside: the sorting of the individual into groups is a *katēgorō*, an “I accuse,” an “I speak against,” from *kata* (“against”) and *agoreūō* (“to speak in an assembly”): the category is this thing that stands against the agora, a reactionary *j’accuse*. To categorize is always to hierarchize: no symbolic power exists that is not also a symbolic violence (Bourdieu used the terms interchangeably). Kant’s categories were not just metaphysical but racializing formations: “And it might be that there was something in this which perhaps deserved to be considered; but in short, this fellow was quite black from head to foot, a clear proof that what he said was stupid” (cit. Franklin 2002: 281).

Racism is not the ascription of false properties or incorrect assumptions to certain groups or individuals, as the notion of “prejudice”—an epistemic “pre-judgment” to be corrected with adequate facts—would have us believe. Nor is it the inability to accept certain properties in another as the term “intolerance” suggests: Žižek (2008) rightly points out that it would be absurd were Martin Luther King, Jr. and the Civil Rights movement to have advocated “tolerance” as the necessary condition of a post-Jim Crow America. To merely tolerate does not go far enough; tolerance places a partial burden of responsibility on racialized subjects themselves, as if the fault of racism was not with the fault-ridden racializers but with subjects of racialization processes.

Let us instead venture the hypothesis that racism is an interruption, a cessation of movement, an obstruction of the fulfillment of desire or...
desires: racism impedes before it “prejudges” or “intolerates.” A few years ago, a study suggested that racism involves an absence of goodwill: in an experiment, black and white participants were instructed to board a city bus, explain to the driver that they lacked money for the fare, and ask whether they could still ride the bus. “Bus drivers were twice as willing to let white testers ride free as black testers (72 percent versus 36 percent of the time)” (Ayres 2015). We should accept the radically universalist implications of this study: racism is a refusal of the benefit of the doubt. Importantly, goodwill is a crucial social lubricant; very little is possible in this world—certainly not without much angst and anguish—in the absence of a benefit of the doubt, from college applications to job interviews to home mortgage applications and rental agreements, to the formation of friendships, romances, and professional associations, interactions with strangers, coworkers, neighbors, the police: the whole vast apparatus of state, markets, and civil society require this invisible grace, which is not so much trust, or Putnam’s “social capital,” as it is a concession, a giving to those strangers we surround ourselves most of the time a free pass, a live and let live. The intricate social machinery of advanced societies requires this symbolic substance to operate smoothly. The tragic condition of the racialized body is this fundamental withholding of concessions, of a civil laissez-faire. Racism is an obstacle: the road through life is not smoothed but obstructed. Racism is the differential distribution of processes of smoothing, causing some to experience their journey across the “sea of whiteness” (Ahmed) as tranquil, others as tempestuous: “By not having to encounter being white as an obstacle, given that whiteness is ‘in line’ with what is already given, bodies that pass as white move easily. [...] White bodies are comfortable as they inhabit spaces that extend their shape.” (Ahmed 2006)

Only secondarily does racism involve the fixing of identity; but racism does involve this, too. This act of fixing is a transgression against Whitman’s (2004: 66) famous formula, “I am large, I contain multitudes.” Racism shrinks racialized bodies, denying them their Whitmanesque largeness, compressing their multiplicity into a oneness. It is a false harmony—tautologically, a “harmony that violates...the rules of harmonious composition” (Weber 1841: xxxviii)—because it is an imposition of harmony onto an object that resists unity, whose “rules of composition” are non-decomposable: recall that Leibniz’s idea of universal harmony, a “diversitas identitate compensata” (Leibniz 2004: 30), or “diversity compensated by identity,” a “unity in multiplicity” (Antognazza 2009), was ultimately guaranteed by the goodness and omnipotence of God; in the process of racialization, it is not the a deity but the wholly secular racializers themselves who stand as the final guarantors of harmonized identity: a false corpora, like Mussolini’s Chamber of Fasci and Corporations, which demanded the appearance of harmony – the resolution of conflict between capital and labor in the case
of Fascist corporatism, or the reduction of the racialized being to the brute fact of raciality, rather than their loves, hatreds, desires, sublimity, relations, and aspirations, in the case of processes of racialization. Racism entails the flattening of multiplicities; or, to speak in auditory terms, the reduction of polyphony to univocality. Racism is the imposition of a unicity that would otherwise tend towards a state of fracture: identity is this falsehood of unified meaning.

The true proponents of “identity politics,” then, are not so much members of ethnoracialized minorities as those who are committed to the process of racialization: it is the racializers (really racists; let’s drop the centuries-old euphemisms), who insist on settling once and for all the question of identity, which would otherwise return to its natural state: hazy, inexact, ill-defined. Our ideal here should be Deleuze and his confederate Félix Guattari: “Félix and I, and many others like us, don’t feel we’re persons exactly. Our individuality is rather that of events, which isn’t making any grand claim, given that heccities can be modest and microscopic” (Deleuze 1995: 141). The haecceitas, or “this-ness,” of a thing always involves a radical uniqueness, which is perhaps the central claim of the Christian notion of personhood. Heceity comes from Aristotle’s τὸ τὶ ἐστὶ (to ti esti), “the what (it) is,” which, strangely, coincides with that oft-repeated question of 1970s black American slang: What it is? (“What’s going on?”) The τὸ τὶ ἐστὶ of Black Vernacular English could be read as this reassertion of haecceitas, a refusal of ascribed identity: the “what it is?” understood as both self-interrogation and an exteriorizing interrogation: what are we? Are we really what you say we are?

Recall Josiah Wedgwood’s anti-slavery medallion in the mid-1700s, which asked, “Am I not a man and a brother?” To this we might reply in the affirmative – with the caveat that the “I” must extend beyond these cardinal categories. The universality implied in Wedgwood’s humanism ends in a flattening reduction (the complexity of the ego reduced to its foundational humanity), which, counterproductively, only eases the racializing maneuver: peeling back the onion-like layers of an individual’s inward complexity to a human core opens for the re-inscription of difference through racialization. Instead of universalizing as a defense against such racializing particularity, one might instead particularize even more, to affirm difference all the more, and more radically: abstraction is a sort of falsehood, a betrayal of the self’s multiplicity. To become truly humanistic one must become entangled in the empirical, accepting Deleuze’s assertion that a life is “impersonal and yet singular” (2001: 28), demanding a “wild” sort of transcendental empiricism (ibid. 25), asserting difference all the way down, so to speak. The unity in difference postulated by all hitherto existing humanisms only denies the true wealth of multitudinous difference in the self, leveling it thorough a kind of “ultramonomization” of man’s infinity: the self is “cooked” in the service of preparing a humanistic stew...
Moreover, the apparent race-egalitarianism of humanism really fails to disrupt the racializing project. What we might call more “sophisticated” racists—a dangerous label, to be sure—are more than willing to concede heterogeneity, dispersal of traits, or the lack of essential unity in a race. Paradigmatically, in a 1920s sociological tract, *The Racial Basis of Civilization*, which was an “attack on the Nordic doctrine,” that is, the idea that the “Nordic element” was inherently superior to other racial groups, the sociologist Frank H. Hankins (1926) presented a strangely humanistic defense of miscegenation, or the interbreeding of races. Hankins was not just anybody: in the 1930s he became president of the American Sociological Association, and he was one of the first to sign the 1933 Humanist Manifesto, a cross-breeding, so to speak, of American interwar positivism with aspirational multiracialism that left “race” intact as an ordering principle. Thus, Hankins wrote, “While we are denying the extravagant claims of the Nordicists, we also deny the equally perverse and doctrinaire contentions of the race egalitarians. There is no respect, apparently, in which races are equal; but their differences must be thought of in terms of relative frequencies, and not as absolute differences in kind. They are like the differences between classes in the same population” (Hankins 1926: ix). Perfectly willing to concede that the category of “race” concealed an internal differentiation of qualities, such that some blacks were superior to some whites, his was a sort of Bayesian, or probabilistic, account of racial hierarchization. While Arthur de Gobineau claimed that miscegenation would cause the collapse of civilizations—a “peoples degenerate only in consequence of the various admixtures of blood which they undergo,” de Gobineau (1915: 211) wrote in his blood-curdling tract, *The Inequality of Human Races*—Hankins’s amalgamationist screed readily promoted the harmony of apparently incongruous racial combinations. “Arguments against the crossing of white and negro must...be purely sociological. [...] The mulatto, being in general superior in intelligence to his darker brother and less diverse in physical appearance from the general American population, finds a larger opportunity in the white world. In fact, it seems obvious that large numbers of mulattoes have latterly crossed the line into the white stock. [...] Many such individuals are undoubtedly superior in their biological inheritance to millions of white citizens by whatever traits this be judged” (Hankins 1926: 348). The goal of a more refined, and properly sociological account of racial hierarchy, then, was to promote a eugenic selection of advantageous traits, with the aim of advancing all of humanity, so that “in the ever-changing texture of racial qualities and the infinite combinations still to be made there may in the future arise race blends quite as excellent as those which produced the Age of Pericles, the wonderful thirteenth century, the Renaissance, or the present era in European civilization” (Hankins 1926: 351). In this admittedly incoherent, occasionally rambling creed, we find the outlines of a humanistic racialization, which precisely in its efforts to
unite and advance humanity, finds itself all the more deeply entrenched in the logic of race.

My digital copy of de Gobineau’s *The Inequality of Human Races*, which is freely available on Archive.org for scholars and neo-Nazis alike, was scanned on the basis of a physical copy housed in the Library of the University of North Carolina. (As many no doubt will recall, this university was itself the site of ferocious struggles over its checkered Confederate history and imbrication with the slave economy.) Tellingly, prior to its digitization, the library index card glued to the inside of the front cover shows that the book was checked out by a library patron on 14 September 2001, three days after the World Trade Center attacks on 11 September. Is not this a formidable clue – that the post-9/11 era is also an era of Gobineauist revivalism, from its politically valorized form in Samuel Huntington’s “clash of civilizations” thesis, to the racist, imperial “adventures” in Afghanistan, Iraq, and Libya (a war on the Arabo-Islamic world that finds its echoes in the Norwegian mass shooter Anders Behring Breivik’s “manifesto” hailing the “Gates of Vienna”), and culminating in the alt-right explosion following Trump’s ascendancy?

The only proper response to the humanist project of abstract universality, grounding racial groups in their shared human properties, is to particularize even more. In practice, this entails propounding the necessity of a *race of one*: we are each one of us a “singular race” (Deleuze and Guattari 1987: 379), that is, a radically particular race unto ourselves, whose trajectories through the lifeworld are recognized as entirely unique. No life has ever been lived twice over. In this sense, the racializing project is right: we really do have biologically distinct properties, but they are so suffused with the social as to make them indistinguishable from the social; our lives are social through and through. Even our bloodstream is shaped by the socially constituted availability of nutrients, the presence of environmental pollutants: in short, the external environment is a societal environment. But more importantly, the singularity of our experiences means that our racial identity is wholly singular, a oneness: there are no non-arbitrary ways in which these properties can be clustered together and aggregated upwards. Skin color? Ask the Brazilians, who emphasize non-phenotypical properties, such as income (“money whitens”) and “culture,” as well as physical features only vaguely to skin color, such as hair texture and the shape of the nose, in their constitution of a uniquely Brazilian folk concept of *raça*. Think also of the ancient Egyptians, whose Nubians in the Sudanese south were depicted with dark skin, but nowhere suggested that this blackness was a socially constitutive feature that impinged on allocations of social resources. Skin color is arbitrary, a particular feature of American “race relations” that has been falsely universalized to cover the global domain of racial thought and racial critique.
In Europe, a series of moral, or behavioral, panics broke out in 2018 over (parts of) the exterior surface of the immigrant body: the face, the hand. In Switzerland, a Muslim couple were denied citizenship because they refused to shake the hands of their state interlocutors. Similarly, in Norway, a Muslim man was “discontinued” from his temporary job in Oslo as a schoolteacher because he refused to shake the hands of his female colleagues; he referenced the putative restrictions in the Islamic regulation of the genders said to prohibit physically touching women. The handshake, his critics said, was one of the primary coordinates of Western liberal society; a refusal to shake the hand is a refusal of the Eliasian civilizing process, originally functioning (or so the story goes) to signal pacification, the absence of arms: the proffered hand is a signal that one is not on a war footing. Thus the refusal of the handshake was read as a confirmation of the suspicion of a Muslim invasion, the paranoid phantasy—the fascists need a good war to mobilize the “populace” (itself a pejorative term, from the Italian popolaccio, or the “common people”)—of a second Battle of Vienna, where the Ottoman horde would once again be turned back from European shores. These behavioral cues were magnified, blown out of all proportions, read as the excessive, “monstrous” precursors to a clash of civilizations: recall that the word is derived from monstrum (“portent”), and monere, “to warn,” making the monster a signal of the shape of things to come, the advent of an apocalyptic event.

There is an inherent duality in racism. It is both the frustration of desire (the inability to breathe clean air, a failure of admissions) and a corresponding production of particular desires. Both the constraining and the productive power of racism must be mapped. How was it experienced, and why did it arise? In short, the phenomenology and genesis of racism. But let us never forget that in late modernity, racism has primarily become a behavioral, not physicalist, concept. Racism has become almost completely biopolitical, concerning itself with the efficient regulation of life itself: from a concern with “broken” family structures to maintaining safe streets, integrating racialized others into the demands of the “knowledge economy,” adapting bodies to the demands of the state and its markets. Racism has become biopolitical in so far as racists are able to mobilize popular support based on (purported) behavioral inadequacies and maladaptations in their targets of racialization, recalcitrant elements that resist the exigencies of modern capitalist economies. Race is no longer, if it ever was, primarily about appearance: it turns on (purported) behavior. Late-modern racism revolves around the realm of the cultural, and how it interacts with the political-economic, more than the (supposedly) biological, so much so that we can say that a racism of the twenty-first century stands at the intersection of three key behavioral domains: sexuality, law, and work.
If we have still not arrived at a proper concept of racism, and if racism as an analytic concept is increasingly falling out of favor, this is in no small part due to its narrow circumscription to micro-interactionism: to speak of racism is all too often to be caught up in a paradigm of interactions, meetings; in brief, to be trapped in the individual-level encounter. Racism as a tool for seeing has been largely limited to this encounter between two beings; we are lost in the encounter. To move beyond this limiting conception, we must expand the notion to encompass a kind of production: a production of bodies, a production of brains, minds, dispositions, *habitus*, worldviews. To this can be added external actants, like environments, a neighborhood, material objects. We need a theory of racism that can recognize that a *highway* can be racist—“...black and Hispanic people experienced 37% higher exposures to the pollutant [NO2] than white people in 2010”), that an *iPhone* can be racist (“A Chinese woman, identified only by her surname Yan, was offered two refunds from Apple for her new iPhone X, as the AI-powered facial recognition technology was unable to tell her and her other Chinese colleague apart”), that an *algorithm* can be racist (auto-corrections reflect in part corpora of natural language use, scraped from the Web, which mirrors ongoing linguistic racialization). It is easy enough to multiply such examples across the entire breadth of “inanimate” objects, which are precisely *not* inanimate in so far as they possess the potential to act, to become vectors of racialized power, “to affect and be affected” (Deleuze and Guattari 1987: 261). And even in typing these words, my word processor automatically corrects, so to speak, the word “racialized” to “radicalized”; perhaps we should take such instances of algorithmic thought as a serendipitous opportunity to deconstruct racialization – for to be racialized *is* in some sense, at least potentially, to be radicalized, to witness the otherness in oneself as a first step towards recognizing the unity of struggles, the otherness in others’ struggles, as with the Black Lives Matter campaigners’ recognition of commonalities with Gazan Palestinians and their struggle against the blockade, against daily terrors.

II. AN ETHNORACIAL CRITIQUE OF ARTIFICIAL REASON
Artificial intelligence (AI) will soon become so complex, so inscrutably advanced, that it will exhibit behavior that, to the human eye, will seem obscure, even mysterious. To meet this challenge, we must establish a behavioral science of AI, a disciplined attempt to explain, predict, and control the behavior of AI. Just as human thinking and action can be puzzling, or even mystifying, thereby mandating the emergence of a science of the human being—from sociology to physiology—so too will

---

an AI that approaches or even exceeds human capabilities require a coherent, well-organized, and incisive science whose aim is the unravelling of the obscurity of AI.

As AI grows human-like in its capabilities and inner workings, its outward face will only become more impenetrable. This only matters, of course, because AI will suffuse our lives to a growing degree: The heightened power of AI is what makes it an important object of study. To understand it calls for a structured framework for the scientific study of AI. This science, technitology—the science of explaining, predicting, and controlling the behavior of artificially intelligent systems—will draw on the methods, concepts, and approaches, found in social science, computer science, law, and philosophy, to name just a few obvious wellsprings of knowledge.

Technitology is a neologism, modeled on the Greek τεχνητός (“artificial”). As Wilhelm Pape’s (1914) Handwörterbuch der griechischen Sprache notes, this Greek term means that which is “künstlich od. listig gemacht” (“artificially or cunningly made”). The double entendre of the term suits our purposes well, for AI is not only artificially intelligent, but, especially following an intelligence explosion (see e.g. Barrat 2013), but also in its intermediate stages leading up to it, including present-day technologies, will be—and in some ways already is—cunningly intelligent. From a human perspective, a sufficiently advanced AI system will become a form of technology that is potentially evasive or deceitful: in this sense, AI may very well entail the culmination of Hegel’s “cunning of reason,” the realization of an Idea that always lay dormant in the Promethean potential. Already, algorithms, which for all practical purposes could be described as artificially intelligent, distribute social resources and forms of capital in ways that are distinctly political, involving the exercise and deployment of symbolic power. Technitology is little more than the science of exposing the cunningness of AI, of laying it bare for human observers to comprehend, critique, and even oppose. Technitology is the science of the cunning of artificial reason. It is an instrument—a scalpel—to cut into the flesh and bone of advanced algorithms and sophisticated cybernetic systems, to produce digital autopsies. In turn, these autopsied digital corpses will form the basis of an informed public discussion of the proper regulation of the digital, which has hitherto only been haphazardly thought by the political.

The central premise here is that the most probable future course of AI’s effects on human societies lies in the reproduction of already-existing hierarchies, asymmetries, power structures, and relations of domination. The extreme, dystopian visions of some science-fiction-inspired authors are unlikely to come to pass: “If you roll your eyes when people talk of gun-toting Terminator-style robots taking over, then you’re spot on: this is

---

areally unrealistic and silly scenario” (Tegmark 2017: 134). Algorithmic automation and autonomy is far more likely to reproduce class difference than drastically modulate the human condition. While no-one can predict the future with any degree of absolute certainty, the disconnect between the physical world and the digital domain is likely to prevent the formation of such problems as the “grey goo problem,” i.e. the fevered imagination of some speculative dystopians, which holds that a superintelligent computer will simply take over the planet, restructuring matter at the molecular level through self-replicating nanobots or similar – and thereby eviscerating the human species. Far more probable, certain in the short-to-medium term is the slow, steady reproduction of social differences through such thing as the allocation of information, resources, and power to already-dominant actors – at the expense of minorities, workers, and the (broadly speaking) disenfranchised.

What are we talking about when we talk about artificial intelligence? Definitions vary, but all share the idea that computers are able to handle tasks intelligently, that is, with some modicum of autonomy, automation, and aptitude. As one recent historical overview of the subject noted, “AI ranges from machines truly capable of thinking to search algorithms used to play board games.” Clearly, much technology that is already available to the average consumer possesses AI-like properties. Google Translate is able quite effectively to translate between dozens of languages. Uber’s car-hailing software can locate available drivers and route them to my location. Google Maps can chart an efficient commute between my home and workplace. In some sense, these programs, and thousands of other applications besides, already fit the description of AI, because the “intelligence” part of AI is so broad.

Much of the literature on AI, however, is concerned with a higher-order form of intelligence, or what is called artificial general intelligence (AGI), which is considered the capacity to “truly” think, to “truly” understand. Quite what this “truly” is supposed to consist of is less clear. As one early report on the first annual symposium of the American Society of Cybernetics in 1967 reported, “The ordinary function of computers consists of performing prodigious numbers of high speed additions and subtractions and storting and retrieving immense amounts of data.” But, as the writer pointed out, “These functions do not constitute thinking—the use of judgment and imagination” (Schmeck 1967). Famously, John Searle (1999) elaborated the so-called Chinese room problem: imagine that you are trapped in a room and are handed notes written in Chinese. With a wide array of dictionaries and grammar tables at your disposal, you are able to parse the notes and formulate replies. In this sense limited sense, you are able to communicate in Chinese. But do you truly understand Chinese? AGI is on some accounts held to involve a truer understanding

---

than that exhibited by the person trapped in the Chinese room problem, demonstrating instead a human-level (or higher) insight into the problems at hand and a fluency in applying the requisite skills to solve them. While the laws of physics do not as far as we can tell exclude the possibility of some day creating AGI systems that match or supersede human brain-power, it is much less clear “if/when we humans will succeed in building such superhuman AGI” (Tegmark 2017: 130).

One might ask why a science of technitology is necessary. After all, unlike the object of the human sciences—human beings themselves—we will have designed and implemented the procedures, programs, and systems that make up advanced AI. Therefore, one might think, all one would have to do is look beneath the hood of the machine, so to speak, to examine the source code, design protocols, inputs, and outputs of AI, to understand it: there should be no mystery, properly speaking, when the whole system is readily available for inspection. But there are a number of problems with this argument. First, it is not at all clear that AI will always be made transparent in this way. In fact, this is why philosophers like Nick Bostrom and Eliezer Yudkowsky (2018) have made transparency one of the key objectives of good AI design. There is even a name for this kind of transparency, Explainable Artificial Intelligence (XAI), and proponents of this idea have argued that we need “opaque machine learning (ML) models” so that we can understand why AI produces the outputs it does – and, if need be, change the underlying algorithms through legal regulation and state-led intervention.

The main impetus for AI development at present comes from state actors, primarily in the form of various military and other semi-clandestine research programs, and from private corporations, who see a tremendous profit-making potential in the development and maturation of this sort of technology. Neither of these parties would have any reason to fully, or even partially, display the inner workings of their AI systems for the benefit of interested outsiders: in fact, the intellectual property found in the “innards” of their AI would be valuable, either in monetary or in geopolitical or military terms, only to the degree that it would hidden from view. This, of course, also holds for any other sufficiently advanced piece of software: Microsoft shares would plummet were the source code of Microsoft Windows to be made widely available for all to tinker with and compile on their own. But even as AI programs are likely to be kept hidden from public view, their effects on daily life will increase, to the point that our daily lives will be shaped and structured by the (potentially) opaque actions of AI. We will live in the realm of algorithmic governance, algorithmic discipline, and algorithmic control. In short, as AI becomes more advanced and opaque, a social demand arises for non-governmental and non-corporate agents to be able able to explain how we moved from

---

some set of inputs to some outputs. Technitology is the science of prying open the “black boxes” of corporate and state-run AI systems with the aim of illuminating them. Explanations are always retroactive, and technitological explanations will be no exception, but they will offer a natural starting point for society-wide discussions about the powers exercised by AI systems, the differential distribution of social energies and resources to particular groups, places, institutions, and things.

Second, and more importantly, AI will remain obscure even to its makers. This is far more important, for it suggests that AI will possibly outsmart its creators, and therefore elude even the most minimal of internal checks on its future behavior. But equally important, the emergent complexity in AI systems means that the creators simply will not understand what it is they have made. Even with full transparency and total access to all the inner workings of AI, and even with a complete catalogue of its source code and relevant inputs, the makers of advanced AI will more likely than not fail to fully grasp how their AI creations work. They may be unable to predict in advance or retroactively explain how advanced AI produces certain outputs because of complexity. In fact, even at this relatively early stage in the history of AI development, we are beginning to see the contours of this problem taking shape even now. As Carl Miller notes, the complexity of algorithms in advanced computer systems have already reached such a high degree of sophistication, and have become so “ridiculously complicated,” as Miller phrases it, that their programmers no longer entirely understand how they work. Here’s what Miller learned from speaking with a programmer at one unnamed tech giant:

The researcher knew, of course, what data he’d fed into the process. He knew why he’d designed it, the problem it was trying to solve and the outputs that it produced. However, after he’d been trying to explain it for over an hour, he sat back in his chair, exhausted. “Yes, as you can see, the gap between input and output is difficult to understand,” he said. He’d flooded the algorithm with a huge amount of information, “a trend”, he said, because in the tech giant he could, and everyone did. But the amount of data meant it was hard to tell what the salient inputs within it were. “From a human perspective you’re not sure which of the inputs is significant; it’s hard to know what is actually driving the outputs. It’s hard to trace back, as a human, to know why a decision was made.” (Miller 2018)

These kinds of statements smack of science fiction, but they are quickly becoming reality. As Demis Hassabis, the head of DeepMind, a British AI firm owned by Google, has said, “There’s a whole bunch of interesting and difficult philosophical questions…that we’re going to have to answer about how to control these systems, what values we want in them, how do we want to deploy them, what do we want to use them for.” It is true

that we must think long and hard about what sorts of values, preferences, and ideological priorities we should tolerate in AI systems.

These systems will have ideologies: They will privilege certain outlooks, behaviors, and social groups. How? By ranking certain kinds of information, by prioritizing certain kinds of individuals. In short, they will exercise what the French sociologist Pierre Bourdieu called symbolic power. Already AI designers are beginning to demonstrate an awareness of these issues. In October 2017, DeepMind announced that it was establishing an “ethics and society” unit. "At DeepMind, we start from the premise that all AI applications should remain under meaningful human control, and be used for socially beneficial purposes," two DeepMind employees wrote in announcing the establishment of an ethics division. “Understanding what this means in practice requires rigorous scientific inquiry into the most sensitive challenges we face...If AI technologies are to serve society, they must be shaped by society’s priorities and concerns.” Similarly, in June 2018, Google announced a set of “AI principles,” consisting of seven “objectives” and four broad areas Google pledged it would studiously avoid. This is a step in the right direction. Only some years ago, the technology observer Evgeny Morozov could write in all seriousness, “What’s truly rankling here is Google’s insistence on the supposed neutrality and objectivity of its algorithms. Instead of acknowledging that its algorithms might have shortcomings and biases that ought to be corrected, Google behaves as if introducing humans to occasionally review the work of its algorithms would be tantamount to abandoning all faith in artificial intelligence as such” (Morozov 2013: 142). Clearly, Google’s recognition that AI does involve something like the exercise of symbolic power—a term they, of course, never use—and that AI is not necessarily (or even probably) a site of the neutral deployment of objective computational ordering, is at least the beginning of a society-wide conversation about the potential iniquities promoted and transformed by the large-scale adoption of AI.

What these developments portend is the coming of ideology to AI. When an AI system that black Americans a home mortgage or produces an “automated” lower credit score than their white counterparts, this is an example of ideological AI in action. Now, we should take care to place the term “automated” in quotation marks, for nothing in the realm of AI is so far fully automated, if only because these systems are still always in the first instance produced by human hands and minds, and in the last instance used by human operators. Even at this early stage, it is readily apparent that the realm of law is one social domain where AI’s ideological effects have been and are likely to continue to be felt most strongly. In the United States, PredPol is a “predictive policing” software package developed to help police officers select which areas to patrol and police.
Given constraints of time and energy, the predictive policing paradigm holds that it is reasonable to target “hotspots” and “high-crime” areas. Informational scarcity also means that police officers, and their software overlords, must rely on shorthand techniques and heuristic signals to target geographic areas for intensive policing. Predictive policing can, however, involve powerful ethnoracial, spatial, and social class biases. In 2016, the Human Rights Data Analysis Group discovered that PredPol, which promises to “predict critical events and gain actionable insight,” tended to send police officers into poor, black neighborhoods (Lum and Isaac 2016). Researchers at the University of Utah also discovered that “because the software learns from reports recorded by the police rather than actual crime rates, PredPol creates a ‘feedback loop’ that can exacerbate racial biases” (Cossins 2018). In short, the purported rationality of such heuristics and the practical consequences that flow from them—increased police vigilance, heightened arrest rates, and a growing incidence of incarceration in certain underprivileged neighborhoods and dominated social groups—is severely curtailed, even undermined, by the noxious social effects of artificially intelligent policing systems.

Let’s think a little more closely about PredPol. It is a system that is likely to be rife with biases because of the way it is designed. According to the creators of this system, PredPol “uses a machine-learning algorithm to calculate its predictions. Historical event datasets are used to train the algorithm for each new city (ideally 2 to 5 years of data)...PredPol uses ONLY 3 data points – crime type, crime location, and crime date/time – to create its predictions. No personally identifiable information is ever used. No demographic, ethnic or socio-economic information is ever used. This eliminates the possibility for privacy or civil rights violations seen with other intelligence-led policing models.” While this sounds reasonably good on paper, we should think about what is contained in the historical data used to train the machine learning system. Clearly, a recorded crime, as registered by the police, is not at all the same as the actual incidence of crime: this is why crime victimization surveys are carried out; instead, we can think of the two elements as conjoined in a Venn diagram, with a certain amount of overlap, but just how much overlap will depend on the making-visible of certain kinds of crime (e.g. crime committed in poor black neighborhoods) versus the making-invisible of other kinds of crime (e.g. crime carried out in wealthy white suburbs).

If historically registered incidences of crime are spatially, racially and socioeconomically skewed, then the training set will feed the algorithm with a biased sample of urban space for future policing - and reproduce these inequities in policing by targeting police activities in socially definite ways. Clearly, we would need to know a lot more about

---

*See e.g. [http://www.predpol.com/](http://www.predpol.com/)
* [http://www.predpol.com/how-predictive-policing-works/](http://www.predpol.com/how-predictive-policing-works/)
the inner workings of PredPol to carry out a comprehensive critique of its behavioral ramifications. But that is in part the point of technitology: we urgently need to establish a coherent framework for the rigorous study of the significance of AI for social life. One method would be to regulate transparency: any AI system with a certain amount of bearing on social life would be subject to inspection and analysis by committees of technitologists possessing the requisite competence in computer science, criminology, and philosophy.

Similar and related examples are comparatively easy to multiply. Take the example of facial recognition. In 2018, the MIT researcher Joy Buolamwini conducted research on the ethnoracial biases in leading facial recognition software packages. Buolamwini took a closer look at software provided by Microsoft, IBM and Face++, a Chinese tech startup. By showing the programs 1,000 different faces, Buolamwini discovered that all three programs were relatively good at distinguishing between the faces of white participants. “But when it came to dark-skinned females, the results were dismal: there were 34% more errors with dark-skinned females than light-skinned males...As skin shades on women got darker, the chances of the algorithms predicting their gender accurately “came close to a coin toss.” With the darkest skin women, the face-detection systems were getting their gender wrong close to half the time.” Part of the explanation probably lies in the training sets used by AI developers. AI relies on huge quantities of data to train machine-learning systems, such as neural networks, in recognizing certain patterns. In the case of facial recognition technologies, programmers feed the system with thousands of faces that have already been tagged by humans as possessing certain properties – male or female, white or black, for instance.

But what happens when these training sets themselves contain biases? If a training set contains too few minority individuals—because they are a minority, in the literal sense of being in the minority, or because they have been ignored, either willfully or inadvertently, by AI’s creators—the neural network is unlikely to be able to properly isolate and distinguish black individuals in real-world use. Similarly, if the tagging of the original training set is mistaken, machine learning will get off on the wrong track. As one New York Times report has noted, “[T]he fact that neural networks are probabilistic in nature means that they’re not suitable for all tasks. It’s no great tragedy if they mislabel 1 percent of cats as dogs, or send you to the wrong movie on occasion, but in something like a self-driving car we all want greater assurances” (Lewis-Kraus 2016). And, we might add, in something like policing.

Why does racial bias in facial recognition matter? For one thing, it could lead to dangerous outcomes for ethnoracial minorities. If facial recognition software is used to identify wanted fugitives by being

---

connected to closed-circuit television (CCTV), as is already done in China, it becomes increasingly important that facial recognition is able to accurately identify members of the public. In China, facial recognition is being used in this manner: the country boasted of having some 170 million CCTV cameras in 2017 and was on track to install an additional 230 million cameras by 2020 (Russell 2017), and AI is being used to track and apprehend suspected criminal offenders. In one spectacular case, Chinese authorities were able to apprehend a suspect in the midst of a 60,000-strong concert audience: the 31-year-old man, wanted for “economic crimes,” was reportedly “shocked” that he was discovered in the middle of this vast crowd (BBC 2018). A racially biased form of AI-driven facial recognition technology, hooked into CCTV and steering policing behavior, could easily lead to the erroneous arrest, incarceration, or even police killing of innocent individuals – in the style of that old and overtly racialized police statement, “the suspect is described as a black male,” only now taken to a higher, Panopticon-style level of technological surveillance.

III. CONCLUSION
To return to the problem of racism means returning to the problem of what Deleuze termed “a life of pure immanence.” In his final essay before his death, Deleuze (2001) offers an appraisal of the meaning of “a life” through a reading of Charles Dickens’s Our Mutual Friend. “Suddenly, those taking care of him manifest an eagerness, respect, even love, for his slightest sign of life,” Deleuze (2001: 28-29) writes. “Between his life and his death, there is a moment that is only that of a life playing with death. The life of the individual gives way to an impersonal and yet singular life that releases a pure event freed from the accidents of internal and external life, that is, from the subjectivity and objectivity of what happens.” This “haecceity no longer of individuation but of singularization: a life of pure immanence, neutral, beyond good and evil,” is a vision of life as at once the highest abstraction and the deepest particularity: a human life that is valuable not because it is human but because it is a life. And it is this radically universalist and particularizing vision of human dignity that is encapsulated in the passage in Pauline theology: “There is neither Jew nor Greek, there is neither slave nor free, there is no male and female […]” (Galatians 3:28). There is only a life. The violence of racialization lies in the denial of this indefinite form. Racializers see only particular forms of life: the life of this person or that. To this the soundest of responses is that a black life matters – because it is a life.

REFERENCES


