The Impossibility of Automating Ambiguity

Abeba Birhane*
University College Dublin School of Computer Science
Lero – The Irish Software Research Centre
abeba.birhane@ucdconnect.ie

Abstract On the one hand, complexity science and enactive and embodied cognitive science approaches emphasize that people, as complex adaptive systems, are ambiguous, indeterminable, and inherently unpredictable. On the other, Machine Learning (ML) systems that claim to predict human behaviour are becoming ubiquitous in all spheres of social life. I contend that ubiquitous Artificial Intelligence (AI) and ML systems are close descendants of the Cartesian and Newtonian worldview in so far as they are tools that fundamentally sort, categorize, and classify the world, and forecast the future. Through the practice of clustering, sorting, and predicting human behaviour and action, these systems impose order, equilibrium, and stability to the active, fluid, messy, and unpredictable nature of human behaviour and the social world at large. Grounded in complexity science and enactive and embodied cognitive science approaches, this article emphasizes why people, embedded in social systems, are indeterminable and unpredictable. When ML systems "pick up" patterns and clusters, this often amounts to identifying historically and socially held norms, conventions, and stereotypes. Machine prediction of social behaviour, I argue, is not only erroneous but also presents real harm to those at the margins of society.

Keywords
Complexity, machine learning, artificial intelligence, embodiment, equity, racial justice

1 Introduction

Post-Cartesian frameworks, including developments within the embodied and enactive cognitive sciences, complex systems science, and dialogical approaches to cognition, strongly emphasize the inherently indeterminable nature of the person and the inextricably entangled relationship between person, other, and technology. These traditions have challenged Cartesian ambitions that neatly delineate human behaviour and actions into dichotomies, instead emphasizing ambiguities, continuity, and fluidity. The person exists in a reciprocal relationship with others in a social, cultural, and increasingly digitized and automated milieu. People, far from being static Cartesian selves, are active, dynamic, and continually moving. The *interactive turn* (e.g., De Jaegher et al., 2010), for instance, has been playing a crucial role in shifting emphasis from the view of the individual as a relatively stable and fully autonomous entity that can be fully understood, to the view of the person as active and dynamic, pregnant with a myriad of open-ended possibilities. On a similar note, distributed cognition and extended mind (Clark & Chalmers, 1998) frameworks have challenged the idea that cognition ends at the skull and that the skin marks the contours of the self, fuzzing the traditionally held neat

* Corresponding author.

understanding of cognition and self. There is no clear line demarcating where the mind ends and the world begins. These nuanced approaches recognize that uncertainty, ambiguity, and fluidity, not static dichotomies, exemplify human beings and their interactions. We are fully embedded and enmeshed with our designed surroundings and we critically depend on this embeddedness to sustain ourselves. Furthermore, our historical paths and the moral and political values that we are embedded in constitute crucial components that contribute to who we are. The idea of defining the person once and for all, drawing static classifications, and making accurate predictions thus appears a seemingly futile endeavour. In complexity science terms, human beings and their behaviour are complex adaptive phenomena whose precise pathway is simply unpredictable (Juarrero, 2000).

Automation, on the one hand, is something that is achieved once a given process is complete, that is, it is understood, and discrete such that it can be implemented from a set beginning to a set finish reliably. People and social systems, on the other hand, are partially-open, always becoming, and inherently unfinalizable (Bakhtin, 1984). Automation as complete understanding, therefore, stands at odds with human behaviour, which is inherently incomplete, making machine classification and prediction futile. Given the open and incomplete nature of human beings and social systems, automating sensible (as opposed to automating nonsense and random) ambiguity and indeterminability is ill-conceived. A machine capable of grasping humanity by definition is capable of grasping open-endedness, incompleteness, fluidity, and ambiguity. Alas, this becomes something other than machine or automation as we know it.

Cartesianism, in nuanced forms, remains pervasive in various fields of enquiry from the physical to the human sciences, and computation and AI are no exception (Dreyfus, 2007; Weizenbaum, 1976; Winograd & Flores, 1986). Nonetheless, not all computation and AI is Cartesian. Grounded in a dynamically interactive, embodied, distributed, and fluid understanding of the world, various approaches address questions of AI in a manner that goes beyond Cartesianism. The broadly construed field of Artificial Life (ALife), for example, is concerned with generating artificial systems—via computer simulations, robotic agents, or biochemical processes—that behave like living organisms (Langton, 1997). Technology is conceived of as dynamic, interactive, and embedded in social systems within ALife inspired approaches. Through a proposal for *dynamic interactive artificial intelligence*, Dotov and Froese (2020), for example, call for systems that emphasize user-machine inter-dependence over autonomy. In recognition of the adaptive and self-organizing nature of technology, the term *living technologies* has gained momentum within the field of ALife (Aguilar et al., 2014; Bedau et al., 2010; Gershenson, 2013). And as technology becomes more "living", the question of safety also becomes more central (Gershenson, 2013). The morphing of technology into society, brings both benefits and challenges, according to (Aguilar et al., 2014). This in turn, the authors argue, calls for the establishment of ethical principles for artificial life. Similarly, Bedau et al. (2010) have argued that the creation of living technologies requires the considerations of ethical issues, the development of safeguards, as well as "proper mechanisms to prevent its misuse" (p. 95). Echoing similar sentiments, Helbing et al. (2012), have pointed out the risk of such technologies benefiting only a few stakeholders instead of all humanity.

Although ALife puts dynamicity, embeddedness, and reciprocal and interactive technology–society relationships at its core, the social, moral, ethical, and political concerns of technology are merely explored. Although ethics and safety concerns are gaining more attention, they largely remain peripheral, and studies rarely treat these subjects in and of themselves in-depth and rigorously. This article builds on the fluid technology–society relationship underlying ALife, but focuses on the impossibility of predicting human behaviour. Furthermore, it examines the ethical consequences of attempting such predictions, as well the concrete impact on specific populations.

Machine learning (ML) systems increasingly pervade the social, political, legal, and commercial spheres sorting, classifying, and predicting human behaviour. Networked and ubiquitous AI systems such as the Internet of Things (IoT) and smart technologies that pervade day-to-day life reduce every corner of lived experience as behavioural data to be used as input into such systems (Zuboff, 2019). Patterns discerned from this huge volume of data by ML systems are used to infer and predict human behaviour and actions. The practice of sorting, classifying, and predicting using ML tools is often applauded as a beacon of technological progress and a revolutionary marvel that provides answers

to long-standing problems. In a world marked by complexity, change, and uncertainty, shortcuts and simple answers are often championed (Birhane, 2021). Analytics companies boast their ability to provide insight into the human psyche and predict human behaviour (e.g., Qualtrics [https://www.qualtrics.com/uk/]). Some even go so far as to claim to have built AI systems that are able to map and predict "human states" based on speech analysis, images of faces, and other behavioural data (e.g., Affectiva [https://www.affectiva.com/]). Such practice of sorting, organizing, and forecasting the world necessarily has actionable impact with grave consequences. ML systems are not only an academic research endeavour, but a multi-billion dollar business where these tools are deployed into the real world in high-stakes decision-making including in hiring (Ajunwa et al., 2016; Sánchez-Monedero et al., 2020); medicine (Ferryman & Pitcan, 2018; Obermeyer et al., 2019); and criminal justice systems (Angwin et al., 2016; Lum & Isaac, 2016).

In this article, I place machine categorization and predictions within the broader and historical Western science and philosophy that aspires to pin down, taxonomize, and simplify the complex and interconnected world. Although ML and AI[1] deal explicitly in probabilities and risks rather than in Newtonian determinacies, I contend that machine categorization and prediction of social outcomes limits possibilities and creates a world partially determined by prediction itself. The social world is messy and fluctuating but also inundated with persistent social norms, power asymmetries, and historical injustice. Historical norms and traditions are often unkind and unjust to individuals and groups at the margins of society, and accordingly, attempts to find stable patterns to sort and categorize the social world pick up these deeply ingrained norms and injustices. Far from being static, social realities are continually co-constructed and the integration of ML systems into day-to-day life increasingly plays a crucial role in influencing the kind of social reality that exists. In Barad's words, "Reality is sedimented out of the process of making the world intelligible through certain practices and not others. Therefore, we are not only responsible for the knowledge that we seek but, in part, for what exists" (Barad, 1998, p. 105). As systems that interact with and are inextricably linked to the social sphere, ML systems partly create social orders. However, while recognizing ML systems as practices that alter the social world, it is also important to acknowledge that responsibility and opportunity to create social orders are unequally distributed. Social, economic, and other privileges mean that a small homogeneous group is endowed with the creation of ML systems, and in part for what exists, contributing to the maintenance of the status quo, while the least privileged are forced to live in such realities that the few create, oftentimes subject to machine harm and injustice.

The rest of the article is organized as follows. Section 2 outlines the underlying Cartesian tendencies of current ML systems that strive for stability, order, and predictability. In section 3, I argue that machine classification and prediction impose determinability and limit possibilities. This is followed by section 4, where I illustrate the fluid, ambiguous, and non-determinable nature of people and social systems. I review current research that illustrates how prediction is a self-fulfilling prophecy in section 5. Next, I look at how machine-imposed determinablility, opportunity, and harm are distributed disproportionately within society in section 6, and I examine how the very practice of sorting and predicting are inherently political in section 7. Section 8 takes a brief look at creativity, which stands outside determinability as a potential transformative force to a just world, and I close in section 9.

2 Cartesian and Newtonian Inheritances

> Traditional science in the Age of the Machine tended to emphasize stability, order, uniformity, and equilibrium...[whereas] most of reality, instead of being orderly, stable, and equilibrial, is seething and bubbling with change, disorder, and process. (Prigogine & Stengers, 1984, pp. xiv & xv)

1 AI can generally be conceived of within two broad categories: narrow AI and general AI, the latter commonly known as Good Old Fashioned Artificial Intelligence (GOFAI). The use of AI throughout this paper refers to narrow AI, specifically ML systems that deal in mathematical probabilities which are increasingly used in decision-making in the social sphere.

In *Order Out of Chaos*, Prigogine and Stengers (1984) remark that dissecting problems into their smallest components marks one of the most highly valued and developed skills in contemporary Western philosophy and science. A subject of enquiry is broken down into bite-size chunks and each chunk isolated from another and its environment by means of various tricks and thought experiments. Through the tactic of *ceteris paribus*, scientists and philosophers presume to work from the assumption that some factors can be held constant and that inextricably entangled interrelations can be isolated. Dissection, isolation, and separation characterize the epitome of the burning desire in Western philosophy and sciences for control, manipulation, and formalization of the world around us and the deep quest for certainty, stability, order, and predictability.

Certainty and order have always been highly sought after in Western philosophy and sciences. Through the process of elimination of all things that can be doubted, Descartes attempted to get rid of unreliable and fallible human intuitions, senses, and emotions. This was fundamental in the quest to establish a secure foundation for absolute knowledge based solely on solid grounds: reason and rational thought (Descartes, 1984). Central to Descartes' work was uncovering the permanent structures beneath the changeable and fluctuating phenomena of nature in which he could build the edifice of unshakable foundations of knowledge. The view of the person that emerged from such a worldview was a primarily rational, static, self-contained, and self-sufficient subject that contemplates the external world from afar in a "purely cognitive" manner as a disembodied and disinterested observer (Gardiner, 1998, p. 129). In the desire to establish timeless and absolute certainty, cognitive capabilities and mental processes were privileged as of primary importance to what it means to be a person. Complete understanding, control, order, manipulation, formalization, and prediction find a comfortable home in this worldview. Although few, if any, scholars identify with the Cartesian view as originally proposed by Descartes, this worldview still prevails today in subtle forms.

In a similar vein, and with a similar fundamental influence as Cartesianism, the Newtonian worldview aspired to impose order and to arrive at universal and objective knowledge in a supposedly observer-free and deterministic world. This worldview sees the world as containing discrete, independent, and isolated atoms. Within the physical world, Newtonian mechanistic descriptions allowed precise predictions of systems at any particular moment in the future, given knowledge of the current position, speed, and acceleration of a system. This view fared poorly, however, when it came to the messy, interactive, fluid, and ambiguous world of the living, who are inherently context bound, socially embedded, and in continual flux. In a worldview that aspires for certainty and predictability, the very idea of ambiguity, complexity, and multivalence—the essence of being, so far as there can be any—is not tolerated. Despite the inadequacy of the billiard ball model of Newtonian science in approaching complex adaptive systems such as human affairs, its residue prevails today, directly or indirectly (Juarrero, 2000).

Descartes and Newton did not single-handedly carve out lasting worldviews that have come to dominate much of Western thought. Nonetheless, they represent the quintessential figures that envisaged an objective, universal, and relatively static worldview governed by laws. This striving for a universal law, Daston argues (Gross, 2020), is a predicament that fails when confronted with unanticipated particulars since no universal ever fits the particulars. Commenting on current ML practices Daston explains:

> I think machine learning presents an extreme case of a very human predicament, which is that the only way we can generalize is on the basis of past experience. And yet we know from history—and I know from my lifetime—that our deepest intuitions about all sorts of things, and in particular justice and injustice, can change dramatically. (Gross, 2020, para. 45)

Furthermore, as Ahmed (2007) contends, all bodies inherit history and, more fundamentally, the inheritance of Cartesianism is grounded in white straight ontology. The reality of the Western straight white male masquerades as the invisible background that is taken as the normal, standard, or universal position. Anything outside of it is often cast as dubious or as an edge case or outlier.

ML systems embody the core values of the Cartesian and Newtonian worldviews where historical, fluctuating, and interconnected behaviour is presumed to be formalized, clustered, and predicted in a

value-free and neutral manner. The historic Bayesian framework of prediction is a primary example (Bayes, 1763). This framework has played a central role in establishing explanations of behaviour based from "rational principles alone" (Jones & Love, 2011, p. 169; see also Hahn, 2014). Bayes' approach, which is increasingly used in various areas including data science, machine learning, and cognitive science (Jones & Love, 2011; Seth, 2014), played a pivotal role in establishing the cultural privilege associated with statistical inference and set the "neutrality" of mathematical predictions. Bayes' essay, which was published after his death, included a note that Bayes' method of prediction "shows us, with distinctness and precision, in every case of any particular order or recurrency of events, what reason there is to think that such recurrency or order is derived from stable causes or regulations in nature, and not from any irregularities of chance" (Bayes, 1763, p. 374). However, despite the association of Bayes with rational predictions, Bayesian models are prone to spurious relationship and amplification of socially held stereotypes, a point I expand on in sections 6 and 7. Horgan (2016, para. 30) notes, "Embedded in Bayes' theorem is a moral message: **If you aren't scrupulous in seeking alternative explanations for your evidence, the evidence will just confirm what you already believe**" [emphasis in original].

3 Machine Imposed Determinability

Current AI and ML tools that are increasingly becoming an integral aspect of the social world are direct descendants of the Cartesian and Newtonian worldview insofar as they are tools that impose order and pin down the fluctuating nature of human behaviour through taxonomies, classifications, and predictions. These tools force determinability, limit possibilities, and in the process, create a world that resembles the past. Historical patterns and socially accepted norms are rife with histories of discrimination and injustice, and the implication of automating a future that resembles the past for those historically disadvantaged is dire. I discuss this in sections 6 and 7. Below, I examine how ML systems are tools that create a certain type of future through prediction.

Technological developments and their intimate connection to what it means to be a social, dynamic, embodied living being are not new but go as far back as the history of humankind itself, to prehistoric tools such as stone and spear. However, the current mass scale development and deployment of AI and ML systems pose new and unprecedented challenges and overall negative impacts towards marginalized communities, which are disproportionately negatively affected. Technological artifacts constitute a crucial part of the socio-technological milieu. They mediate and enrich our living world but also hold invisible and unprecedented power in shaping and altering reality. In other words, the design of technology is the design of possibilities and constraints (Suchman, 2007).

Technological tools constrain or enable actions while making day-to-day life seamless. A GPS application on a device, for example, can make travelling from point A to B considerably easier. In some circumstances, technological tools form crucial components that sustain lives—pacemakers, for example. Technological developments, especially ML systems, are not something that stand *above and over* humans but are integral parts of the active, fluid, and dynamic environment of complex, adaptive, self-organizing social systems. Through their power to classify and predict, ML systems direct behaviours and actions towards some things and *away* from others.

ML systems work by identifying patterns in vast amounts of data. Given immense, messy, and complex data, an ML system can sort, classify, and cluster similarities based on seemingly shared features. Feed a neural network labelled images of faces and it will learn to discern faces from not-faces. Not only do ML systems detect patterns and cluster similarities, they make predictions based on the observed patterns (O'Neil & Schutt, 2013; Véliz, 2020). Machine learning, at its core, is a tool that predicts. It reveals statistical correlations but with no understanding of causal mechanisms.

Furthermore, machine classification and prediction are practices that act directly upon the world and result in tangible impact (McQuillan, 2018). Various companies, institutes, and governments use ML systems across a variety of areas. These systems process and datafy people's behaviours, actions, and the social world, at large. Machine-detected patterns often provide answers to fuzzy, contingent,

and open-ended questions. These "answers" neither reveal any causal relations nor provide explanation on why or how (Pasquale, 2015). Crucially, the more socially complex a problem is, the less capable ML systems are of "accurately"[2] or reliably classifying or predicting. Narayanan (2019) broadly maps the application of AI systems into three crude categories: Perception (e.g., face recognition), Automating Judgment (e.g., detecting spam), and Predicting Social Outcomes (e.g., predictive policing). There has been rapid progress in the first category and "the fundamental reason for progress is that there is no uncertainty or ambiguity in these tasks" (Narayanan, p. 7). Automating Judgment, such as toxic language detection, presents a somewhat contested practice because the *correct decision* can often be subjected to disagreement. The task of predicting social outcomes, however, remains fundamentally dubious involving "a lot of snake oil" (Narayanan, p. 9) and is marked with numerous drawbacks and harms. Similarly, in a recent work Salganik et al. (2020) examined the predictability of social trajectories of children from vulnerable families. A team of 160 ML researchers built predictive models using a rich dataset. The authors found that not one model made an accurate prediction and the best predictions were only slightly better than those from a simple benchmark model. Thus, Salganik et al. caution those considering using predictive models to forecast social outcomes. Nonetheless, predictive systems continue to pervade decision-making of social outcomes with disastrous consequences.

4 Indeterminability of the Person

Traditional cognitive science, so far as its desires for a universal, objective, and predictable science of the mind goes, is the heir to the Cartesian and Newtonian worldviews. The continually fluctuating and interconnected state of human affairs finds a stable point through the conjecture that positions the individual person as the seat of knowledge. As such, the individual person is often isolated and taken as the unit of analysis. Great emphasis is placed on her individual mental capabilities as the mind is assumed to be the property of the single individual (Linell, 2009; Marková, 2016). The nature of experimental design in scientific psychology, for example, illustrates the subtle remnant of the desire for cleansing thinking of cultural influences and political dimensions. Research in memory testing, for instance, to a large extent, proceeds from the assumption that memory is a purely cognitive process that resides in the brain (Harris et al., 2011). The individual is removed from her lifeworld and tasked with recalling a series of images or words (often meaningless to the person) using flashcards or a screen in the artificial confines of a laboratory. Subsumed by objective and universalizable formulations, cognitivist approaches paint a picture of the person that equates *persons* with *brains*. Emphasis on dynamic relations, contextual and historical embeddings, and messy interactions, on the other hand, are perceived as a threat that blurs and contaminates neat classifications and universalizable conceptions.

Individualistic and reductionist approaches are irrevocably ingrained in Western thought. It is a continual struggle, even for the most aware researcher and practitioner, to steer clear of them. Taking the individual self as the unquestioned origin of knowledge of the world and of others is a legacy of this tradition (Linell, 2009). Traditional social cognition research, supposedly an endeavour that turns attention to the social, falls short of recognizing the dynamic and entangled nature of bodies and environments, and how each influences the other. The individual person, in social cognition, is portrayed as the meaning generator and is the primary interest of study (Marková, 2016). Pushing back against these individualistic traditions, the broadly conceived approaches of embodied and enactive cognitive science offer views of persons, brains, and nature of reality in general that are active, dynamic, and inextricably connected with environments (and others).

2 The term *accuracy* vaguely denotes how closely models or data represent "the ground truth" or things as they are in the world. However, critical and genealogical examination of the use of this term remains scarce in the ML literature. In the absence of such critical examination, "accurate classification or prediction," especially in the context of social affairs, risks corresponding representations with stereotypically held views.

Fluidity, multivalence, and precariousness are not perceived as obstacles that stand in the way of final and universal understanding, but are acknowledged and celebrated as necessary conditions for existence. The embodied and enactive turn (Chemero, 2011; Kyselo, 2014; McGann & De Jaegher, 2009; Varela et al., 2016), at its core, places living bodies, with their peculiarities, fluidity, and messiness, at centre stage. Living bodies are not stationary entities that can be captured in neat taxonomies, rather they are active, dynamic, historical, social, cultural, gendered, politicized, and contextualized organisms. People are not solo cognizers that manipulate symbols in their heads and perceive their environment in a passive way, but they actively engage with the world around them in a meaningful and *unpredictable* way. Living bodies, according to Di Paolo et al. (2018), are processes, practices, and networks of relations which have "more in common with hurricanes than with statues" (p. 7). They are *unfinished* and always *becoming*, marked by "innumerable relational possibilities, potentialities and virtualities" (p. 6) and not calculable entities whose behaviour can neatly be automated and predicted in a precise way. Bodies "grow, develop, and die in ongoing attunement to their circumstances. … Human bodies are path-dependent, plastic, nonergodic, in short, historical. There is no true averaging of them" (Di Paolo et al., 2018, p. 97).

Universalizable theories of bodies, taxonomies, and statistical predictions of future behaviours all rely on similarities and abstraction of features that are common among particulars. Unique, contingent, and idiosyncratic features and behaviours pose challenges when it comes to deriving elegant taxonomies. However, idiosyncrasies and peculiarities make someone the particular, novel, and creative person they are. Living bodies each face unique challenges defined by the particular trajectory of history of enactments, history of adaptations, and social circumstantial interactions as they continually navigate the social world. Social interactions themselves, De Jaegher and Di Paolo (2007) contend, are active and dynamic engagements that take on a life of their own in an *unpredictable* way. They shift in moods, aims, and levels of intimacy, without the participants intentionally seeking these changes. Most fundamentally, "Our most sophisticated knowing is full of *uncertainty*, *inconsistencies*, *ambiguity*, and *contradictions* [emphasis added]. These characterize how we most often deal with the world, ourselves, and each other" (De Jaegher, 2019, sec. 1, para. 4). Furthermore, Buccella (2020) singles out *indeterminacy* as a key factor that is important in understanding human perceptual experience. Perception is necessarily open-ended and the environment presents unlimited possibilities and offers many ways of life (Merleau-Ponty, 1945/2012; Nonaka, 2020).

On a similar note, examining the problem of meaning in artificial beings, Froese and Taguchi (2019) single out *indeterminability* as the key characteristic of humans that differentiates them from artificial beings. Emphasizing the futility of reductionist approaches to complex adaptive systems, Cilliers (posthumously noted by Preiser, 2016, p. 64) further points out that, "From the argument for the conservation of complexity—the claim that complexity cannot be compressed—it follows that a proper model of a complex system would have to be as complex as the system itself." Precise[3] predictions of behaviours and actions, therefore, are impossible and, when enforced, dire ethical consequences emerge. This might raise the question of whether people and social systems are unpredictable in principle or unpredictability is only a practical limitation and a matter of even more data and compute power. Following Cilliers' argument for the "conservation of complexity," I contend that people and social systems are unpredictable in principle. Having said that, this is not to oppose those that might aspire towards and attempt "accurate" predictions. However, the argument remains that, so long as classification and predictive systems operate within a white straight ontology (Ahmed, 2007), precise and accurate prediction risks measuring how closely behaviours or actions adhere to socially and historically held stereotypes.

Indeterminability and unpredictability do not, by any means, mean that people and social systems wander aimlessly without pattern, habit, or relatively stable behaviour. For any given society there

3 The very term *precision* (in prediction) assumes an observer-free "ground truth," a correct description of reality, and a correct trajectory against which things can be compared. This line of thinking follows Cartesian logic. Precision, accordingly, marks proximity to the presumed "ground truth" or the "correct description of reality" while deviation from it might signal lack of precision. Contrary to these presumptions, descriptions of reality or "ground truth" are never given in an observer-free manner (see section 4).

exist socially and culturally accepted norms and historically established conventions. People self-organize with these dynamic and contextual constraints that serve as the reduction of possibilities (Juarrero, 2000). However, these relative stabilities and habitual patterns do not mean a an individual person can be rendered fully knowable and predictable with precision. Any prediction of future behaviour based on past patterns is at best a statistical probability. We may, therefore, be able to predict a person's general dynamics, under certain conditions, within certain context and time but precise prediction of a person's specific behaviour and action, due to their nonlinear interactions and endless possibilities, are impossible. Moreover, as discussed in section 7, relatively stable patterns and established conventions and norms are charged with social, political, and power asymmetries that benefit or disadvantage groups and individuals depending on one's position in society. When ML systems "pick up" stable patterns, they also identify harmful current and historical norms, prejudices, and injustices. Taking such historical and current patterns as the ground truth, from which to model the future brings forth a machine-determined world that resembles the past and raises a host of ethical and justice issues.

5 Prediction, a Self-fulfilling Prophecy

In the age of ubiquitous and interconnected systems, it has become outdated to conceive of the digital and the physical as separable realities. Outcomes from ML systems are used to justify action in the social world. Who we are is not signified by our bodies alone but also by algorithmic identifications, often assigned to us without our knowledge or consent. Marketing and web analytics companies who gather (and/or purchase) huge amounts of data construct our algorithmic identities from quantifiable attributes that emerge from various input data, observed patterns, and algorithmic inferences and extrapolations. Traces of data and metadata, including behavioural data such as statistics from visited websites, online purchase history, device location, and data from cameras, sensors, and IoT devices that proliferate public and private spaces, all contribute to the construction of the "person" in the digital realm. Algorithmic identifications that are assigned to an individual or a group carry tangible implications as such identifications increasingly play a central role in determining the outcomes of various aspects of an individual's endeavours.

As the hiring process becomes more and more automated, for example, algorithmic hiring tools become most consequential in determining key aspects of a person's life such as how much one earns, where one works and lives. Just like automating any social processes, automating best candidates is prone to automating and reproducing historically and socially held inequities and stereotypes all while providing a veneer of objectivity (Raghavan et al., 2020). Depending on how appropriate or fit candidates are deemed, job opportunities are automatically surfaced to some and withheld from others. More accurately, automated hiring systems serve as tools to reject applicants who do not fit in certain boxes (Ajunwa & Greene, 2019). And given that best, appropriate, or fit applicants are often defined and measured by past success, candidates that do not fit within that box risk exclusion. This was case in point with the hiring algorithm that Amazon deployed and disbanded in 2018 upon discovering that the tool had been discriminatory against women (Dastin, 2018). Amazon's hiring tool was trained to identify the best candidates based on observed patterns in résumés submitted to the company over a 10-year period. Given the male dominance in the tech industry, most résumés came from men. Automating such patterns, by definition, then is a process to automating historical inequities. Predictions based on past hiring decisions reproduce patterns of inequity even when tools explicitly ignore race, gender, age, and other protected attributes (Bogen & Rieke, 2018).

ML tools are not simply methods that sort and classify people and the social world, but are also apparatuses that directly act upon the world transforming social realities and producing certain subjectivities (and not others) (McQuillan, 2018). For instance, faced with an automated assessment system, a job seeker is likely to alter her behaviour in a manner that guarantees positive outcome; awareness that one's social media post has the potential to impact one's perceived characteristics or fitness for a job has the potential to alter her actions and behaviour.

Algorithmic classifications, sortings, and predictions, when enacted in the world, create a social order. For any individual person, group, or situation, algorithmic processes give advantage or they inflict suffering. Jobs are made and lost (Ajunwa et al., 2016; Sánchez-Monedero et al., 2020). Who is visible and legible is legitimized through algorithmic predictions as some subjectivities (and not others) are recognized as a pedestrian (Wilson et al., 2019), or hire-able (Ajunwa et al., 2016; Speicher et al., 2018), or in need of medical care (Obermeyer et al., 2019), or likely to engage in criminal acts (Angwin et al., 2016; Lum & Isaac, 2016).

Furthermore, the very practice of forecasting the future partly acts directly upon the world—machine prediction plays a part in creating what exists whenever such predictions inform decision-making. In a recent paper, Perdomo et al. (2020) illustrate that prediction often influences the outcome that it is trying to predict. They refer to such practice as "performative prediction": "Traffic predictions influence traffic patterns, crime location prediction influences police allocations that may deter crime, recommendations shape preferences and thus consumption, stock price prediction determines trading activity and hence prices" (p. 7599). Similarly, Benjamin (2019) argues that "crime prediction algorithms should more accurately be called crime production algorithms" (p. 83) because predictive policing software predominantly targets historically underserved communities, wherein hyper-surveillance partly produces crime, partly creating a self-fulfilling prophecy.

In another recent study Milano et al. (2020) looked at ubiquitous recommender systems and found that recommender systems shape user preferences and guide choices at both the individual and social levels. The authors contend that attempts to predict preferences have socially transformative effects and impinge on personal autonomy. Furthermore, machine classification and prediction is a multi-billion dollar business, meaning that business objectives play an active role in the direction with which social outcomes and behaviours are moulded. Through "nudging," persuasion, and limiting the range of options available to individuals, recommender systems cajole people in particular directions, often in a way that maximizes profit. Recommender systems often have commercial objectives and are developed for business applications. Consequently, by predicting preferences, recommender systems not only shape individual experience and social interactions, they also hold transformative impact on society in a manner that aligns with commercial values (Milano et al., 2020).

6 Imposed Determinability in Unequal Measures

> Reality is sedimented out of the process of making the world intelligible through certain practices and not others. Therefore, we are not only responsible for the knowledge that we seek but, in part, for what exists. (Barad, 1998, p. 105)

ML systems, for many societies, constitute part of the social ecology forming components of a self-organizing system. These systems exist in and evolve with social norms, trends, and societal uptakes. They morph into the background of daily life to the extent that we forget they exist. Weiser (1999, p. 3) remarked, "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it." Good infrastructure is, by definition invisible and forms part of the background for other work (Star, 2002). ML systems have become inextricably intertwined with what it means to be a human being, yet remain invisible forces that shape lives and opportunities of countless individuals and communities.

On the one hand, machine–human relation and interaction is a process that constitutes a systems-level organization. On the other, it is vital to acknowledge power asymmetries within this systems-level organization. Influences are not bi-directional and benefits, disadvantages, and negative impacts are distributed unequally among individuals and communities. In the case of recommender systems, for example, commercial agents behind the development and deployment of these systems exert power in moulding future realities. The individual person or the end user, has little to no direct influence.

In fact, when ML is used to rank, sort, score, and predict social outcomes, what Narayanan (2019, p. 9) calls "AI snake oil," those being ranked and scored are rarely aware of it or merely know why

they are given certain scores. This makes it difficult to contest and negotiate algorithmically assigned identities and scores. Furthermore, the very practice of scoring, characterizing, and assigning algorithmic identities without people's awareness risks treating people like objects. As Maturana (2004, p. 108) remarks,

> If you deprive people of the opportunity [to contest and protest against their characterization], you treat them like freely disposable objects; they have the status of slaves, compelled to function without the opportunity of complaining when they do not like what is happening to them.

Current data extraction, classification, and prediction (without the awareness of the data subject) practices across analytics firms and big technology corporations, where people are often treated as data objects, bear close resemblance to Maturana's remark.

Predictive models, due to their use of historical data, are inherently conservative. They reproduce and reinforce norms, practices, and traditions of the past. Historical norms and traditions are often unkind and unjust to individuals and groups at the margins of society. Decisions made in the past align with the maintenance of the status quo. The practice of constructing predictive models based on the past and directly deploying them for decision-making amounts to constructing a programmed vision of the future based on an unjust and socially conservative past. Through the application of predictive systems in the social sphere, historically and socially unjust norms, stereotypes, and practices are reinforced. A robust body of research on algorithmic injustice (Benjamin, 2019; Birhane, 2021; Eubanks, 2018) shows that predictive systems perpetuate societal and historical injustice. In a landmark study, Buolamwini and Gebru (2018) evaluated gender classification systems used by commercial industries. They found huge disparities in image classification accuracy; lighter-skin males were classified with the highest accuracy and darker-skin females were the most misclassified group. Similarly, object detection systems designed to detect pedestrians display higher error rates when identifying dark skin pedestrians while light-skinned pedestrians are identified with higher precision (Wilson et al., 2019). The use of these systems ties the recognition of subjectivity to skin tone. Recidivism algorithms unfairly score black defendants as higher risk compared to white defendants of similar criminal conviction (Angwin et al., 2016). Hiring tools tend to disproportionately disadvantage women (Ajunwa et al., 2016). Additionally, the notion of gender that ML systems depend on is a fundamentally essentialist one that operationalizes gender in a trans-exclusive way resulting in disproportionate harm to trans people (Barlas et al., 2021; Hamidi et al., 2018; Keyes, 2018). Machine classification and prediction, thus, negatively impact individuals and groups at the margins the most.

7 Sorting and Predicting Are Moral and Political

A central tenet of the linear, stable, and predictable Cartesian–Newtonian worldview is the idea of objectivity—the assumption that observation, description, and classification of the world can be done from a "View from Nowhere" (Nagel, 1989). Heinz von Foerster famously decried, "Objectivity is a subject's delusion that observing can be done without him [sic]. Invoking objectivity is abrogating responsibility—hence its popularity" (Glasersfeld, 1992, p. 3). The practice of categorizing, ordering, and forecasting a future necessarily entails making moral and ethical choices as deemed "correct" from a given point of view. Through the very act of clustering similarities, boundaries around which behaviours and actions are good or acceptable and which are not are drawn. Furthermore, through their performative powers, predictive systems cast certain ways of being as "normal" while others are deemed "deviant" and in need of correction. Practices that were previously understood to be moral and political, and historically required a great deal of dialogue and negotiation, are obfuscated as apolitical endeavours with the advent of machine classification and prediction. Moreover, the veneer of objectivity that ML is entrusted with presents an added challenge to seeing machine classification and prediction as anything but a technical and mathematical task.

Categorization by human beings itself arises within context and goal-directed activities. Human categories, rather than carving nature at its joint, are developed on the fly to address goal-directed actions. Categories, therefore, are dynamic, unstable, contextualized, and inherently embedded in ongoing activities (Barsalou, 1991, 2009). Machine categorization, therefore, can only be created within the context of a broader goal rather than being an austere, abstract, and mechanical process. Creating categories and drawing boundaries is not primarily a technical choice or a purely scientific question but necessarily an ethical and moral one, especially when such practice has a direct and tangible impact on vulnerable lives. Acknowledging this is a crucial step in taking responsibility for *what exists*. Having said that, responsibility needs to be selectively attributed. Benjamin (2019), in *Race After Technology*, notes that

> it might be tempting to point to the smart technologies we carry around in our pockets and exclaim that "we are all caught inside the digital dragnet!" But the fact is, we do not all experience the danger of exposure in equal measure. (Benjamin, 2019, p. 111)

Although, as Barad (1998, 2007) assessed, reality is something we create together through active practices, a few have genuine say towards what kind of world needs to be co-created while many others are forced to live in it. Existing social, political, and financial power dynamics mean those at the bottom of societal hierarchies have little say, if any at all, in the co-creation of realities.

It is impossible to operate in a value-free space. The type of concerns, questions, and design all reflect the motivations, commitments, and interests of those at the helm of creating ML systems. When values are not explicitly laid out, the values that are taken as "universal" or "neutral" are those values that represent the status quo and the values that are implicit within a given field (Ahmed, 2007; Collins, 2002). Within both the academic fields and corporate industry currently developing ML systems en masse, the values that are taken as "universal" are predominantly the values and interests of the Western, white, male. Computer science (as well as its subfield, machine learning), since its conception as an academic field in the 1950s in the US has always been a field that strove for impactful application within the military, education, and the general social sphere. The field has since come to exert unprecedented social, political, and economic power. Within major technology corporations—from Microsoft to Amazon—Western white men with homogeneous backgrounds and conservative leanings (Cohen, 2018) remain the most predominant powerful figures that influence and redefine social realities (Broussard, 2018). What we find then is a huge power disparity between powerful corporations (and the individuals behind them) and end users whose agency, opportunity, and options have been limited in the process of algorithmic classification and prediction. Such process amounts to financial and personal gain for the former at the expense of the latter. In fact, as Zuboff (2019) argues, the technology industry is built on capitalization and monetization of lived experiences and on building tools of surveillance.

Given massive power disparity, those engaged in the practices of designing, developing, and deploying ML systems—effectively shoehorning individual people and their behaviours into predefined stereotypical categories—carry a great proportion of the responsibility in creating *what exists*. Such demography decides what questions are worthy of investigation, what problems need to be "solved", and what is sufficient performance for a model to be deployed into the world. Consequently, this group bears much greater responsibility and accountability.

Scientific enquiries carry inherent ethical and moral dimensions. The more a topic of enquiry veers towards human and social affairs, the more apparent its moral and ethical dimensions. The apparent dissociation of science from ethics has historically allowed science to evade accountability and responsibility, and similarly so will algorithmic systems if they are allowed to. Ubiquitous deployment of ML models to high-stake situations creates a political and economic world that benefits the most privileged and harms the vulnerable. It also creates a social world where the status quo is maintained and historical injustice perpetuated. For most scholars working at the intersections of algorithmic injustice and science and technology studies, it has become trivial and common knowledge that in speaking of ML models "works well" often equates to "picks up historical patterns." The social world as it is, is

filled with beauty, ugliness, and cruelty. And as Benjamin (2019) notes, to think that one can feed a model with all the world's beauty, ugliness, and cruelty and expect only beauty is a fantasy.

Within the context of complex systems thinking, Artificial Life, or similar fields of enquiry that are primarily concerned with the creation and/or simulation of intelligent systems, the notion of ethics often revolves around the moral status of the "intelligent system." Should the supposedly intelligent system have moral or legal rights on a par with a human being? Does the experience of pain or the capacity for a "theory of mind" differ depending on whether the entity is carbon- or silicon-based? Does working towards ethical systems come down to Isaac Asimov's conception of laws of ethical robotics? How do we prepare humanity for the Singularity? These concerns, for the most part, focus on hypothetical and/or future "First World Problems" (Birhane & van Dijk, 2020). These quests might be a valid intellectual exercise in and of themselves but in light of the mass integration of ML systems into society and the harms they impose on vulnerable individuals and communities, I argue that attention regarding machine ethics should primarily focus on current and tangible concerns.

8 On Creativity

Human creativity is marked by imagination and thinking of things that were not thought of before. Creative innovations that have come to define and revolutionize the world, from music to medicine, are often marked by surprise, spontaneity, and uncertainty. Creativity, Juarrero (2000) reminds us, stands in stark opposition to certainty and predictability. It requires unexpected and spontaneous behaviour and not repeating past patterns and trajectories. Creativity, by definition, defies expectation.[4]

As ML systems attempt to order the spontaneous and non-determinable social world, they create a future that resembles the past leaving us no room for a chance to be different. Such classifications and predictions reinforce stereotypes and impinge on the inherently open-endedness of being, limiting a person's potential by defining them by what people like them have done or liked or how people like them have behaved in the past. When future behaviours are predicted based on past stereotypes, individuals are deprived of the opportunity to challenge stereotypes and to realize their full potential.

In the age of ML where accurate categorization and precise prediction are highly valued, unexpected and spontaneous behaviour poses a challenge and is seen as a deviation to be corrected—not an inherent, indeterminable part of human beings that should be celebrated. In fact, the idea of the "average" can be traced to the explicit logic in the origins of statistics in sociology, crime, and public health by Quetelet in the 1800s. The "average" was considered an ideal. "Quetelet applied the same thinking to his interpretation of human averages: He declared that the individual person was synonymous with error, while the average person represented the true human being" (Rose, 2016, p. 26). To this day, uniquely and idiosyncratically expressed unrepeatable behaviours that defy systemic rules or clusters of patterns—fundamental to creativity—are seen as undesired anomalies and edge cases. ML processes codify the past. They do not invent the future. Doing that, O'Neil (2016, p. 204) emphasizes, "requires moral imagination, and that's something only humans can provide."

Creativity, which stands outside machine determinability, holds the potential to transform the way we approach and use algorithmic systems. From organized resistance, to strict regulations, to disrupting the current capitalist ecosystem, to imagining a fundamentally new type of technology that celebrates differences instead of forcing uniformity (discussed below in the conclusion) all require creativity. As opposed to accepting a machine-determined world as inevitable, creativity is fundamental to imagining an alternative world and the disruptive technologies to actualize such world.[5] Artificial life, underpinned by fundamental recognition of uncertainty and non-determinability as an inherent condition of life and its distinctly creative understanding of humans and society, holds promising paths to a just world.

4 Having said that, the view of creativity as the creation of something novel is not to remove the creative process from its historical, social, and contextual embeddings.

5 This is not to succumb to techno-solutionism where technology is sought as the only viable source of answer. Far from it, and in some cases, the option of *no technology at all* can be the most efficient way to a just world.

9 Conclusion and Discussion

> It is essential for the thing and for the world to be presented as "open," to send us beyond their determinate manifestations, and to promise us always "something more to see." (Merleau-Ponty, 1945/2012, p. 348)

We have so far looked at how individual people and social systems, as complex adaptive systems, are active, dynamic, and necessarily a historical phenomenon whose precise pathway is *unpredictable*. Contrary to this, we find much of current applied ML classifying, sorting, and predicting these fluctuating and contingent agents and systems, in effect, creating a certain trajectory that resembles the past. Such practice, in the process, brings forth a predetermined social reality that places people in stereotyping boxes and maintains the status quo, harming and disadvantaging individuals and communities that are historically and socially disadvantaged.

ML systems, tools that fundamentally classify, order, and predict, I argue, are practices that reincarnate Cartesian and Newtonian worldviews that seek stable, predictable, and complete understanding. But, people (and the social systems that they are embedded in) are partially open, indeterminable, and fluctuating, meaning a complete understanding would imply death of the person or that the social system has come to a stall. Automation, which requires complete understanding, thus stands at odds with human behaviour, which is inherently incomplete and unfinalizable, making machine classification and prediction futile.

Arguably, systems of classification are inherent to humans and part of all cultures, although modern Western culture has produced more than most, without realizing it (Bowker & Star, 2000). Therefore, I do not argue against machine classification altogether. However, given that people and social systems are dynamic and unpredictable and that social structures are hierarchical and saturated with power asymmetries, forcing order and taxonomies brings forth harm and injustice to those at the margins. Furthermore, as Narayanan (2019) remarks, predicting social outcomes is a fundamentally dubious endeavour with many disadvantages and problems but with few, if any, benefits. In this regard, those behind ML systems (from conception, to design, to development, to deployment)—individuals and corporations alike—bear the responsibility for the unjust and harmful social reality they are creating.

Knowledge of self and of the world is fluid, dynamic, and continually moving. Any understanding of the person–society–technology relationship is partially open and evolving. This accommodates the uncertainties of ongoing change and provides room for dialogue, negotiation, reiteration, and revision of any claims and positions. This means that how algorithmic systems are designed and implemented requires continual negotiation between the different stakeholders. The views and input of vulnerable communities that are disproportionately negatively impacted by algorithmic decision-making needs to be central at all stages of the design, development, and deployment process. One way of moving forward to a just society is to envisage a fundamentally different kind of technology that is grounded in ambiguity, fluidity, and diversity of experience. In their proposed vision in *Diversity Computing*, Fletcher-Watson et al. (2018) envisage fundamentally new kinds of computing devices that reflect, promote, and embrace differences rather than eliminating them. This requires a fundamental shift from striving for uniformity towards a technology that underpins the inherent diversity and indeterminability of the world.

The discourse surrounding ML development and deployment is marked by the rush to "solve problems" with little, if any, thought, consensus, or reflection of what the problems are. Machine learning, in this regard, produces "thoughtlessness, the inability to critique instructions, the lack of reflection on consequences, a commitment to the belief that a correct ordering is being carried out," McQuillan (2020, p. 3) argues. Understanding of contingent and underlying factors is crucial and needs to be prioritized over prediction. This requires asking questions such as "Why are we finding these clusters and similarities?" and investigating that further instead of using the patterns that we find to build predictive models (Birhane, 2021). On a more radical way forward McQuillan (2020) puts forward a "non-fascist AI." A non-fascist AI, according to McQuillan, requires resistance to toxic application of AI through self-organization and broader worker mobilization to an alternative social vision. This vision

aspires for AI that confronts us with the injustice of current systems and the tools that form part of the movement for social liberation. Similarly, Kalluri (2020) notes that current ML systems centralize power where it is already concentrated. Consequently, ML models that shift power from the most to the least powerful hold the key to social realities that serve the least privileged.

Finally, the main points that I have discussed in this section are crucial for a radical transformation of AI that embraces uncertainties and indeterminabilities and serves the most disadvantaged. Although acknowledging responsibility and accountability is an important first step, pleading with the powerful to take responsibility and be considerate to the vulnerable is simply not sufficient. Just as science has found ways to evade ethical responsibility by means of systematic separation of "objective science" and ethics, AI, if allowed, will do the same; indeed, in some respects, it has been doing so with impunity. In fact, global technology giants spend millions (Molla, 2019) actively lobbying to influence legislation in their favour, which comes at the expense of less agency and more harm to the masses. The capitalist ecosystem in which ML systems are built and deployed presents one of the greatest challenges. Even for the most well meaning technologists, the incentive structures pressure individuals to develop technology that maintains the status quo, wields existing power, and produces maximum profit. Technology that envisages a radical shift in power (from the most to the least powerful) stands in stark opposition to current technology that maximizes profit and efficiency. It is an illusion to expect technology giants to develop AI that centres on the interests of the marginalized. Strict regulations, social pressure through organized movements, strong reward systems for technology that empowers the least privileged, and a completely new application of technologies (which require vision, imagination, and creativity) all pave the way to a technologically just future.

Acknowledgments

The author would like to thank Alicia Juarrero, Anthony Ventresque, Dan McQuillan, Elayne Ruane, Hanne De Jaegher, Johnathan Flowers, Marek McGann, Os Keyes, Sergio Graziosi, Thomas Laurent, Tony Chemero, and Vinay Uday Prabhu for their useful feedback on an earlier version of this manuscript.

Funding Information

This work was supported, in part, by Science Foundation Ireland grant 13/RC/2094 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero—the Irish Software Research Centre (www.lero.ie).

References

Aguilar, W., Santamaría-Bonfil, G., Froese, T., & Gershenson, C. (2014). The past, present, and future of artificial life. *Frontiers in Robotics and AI*, *1*(8). https://doi.org/10.3389/frobt.2014.00008

Ahmed, S. (2007). A phenomenology of whiteness. *Feminist Theory*, *8*(2), 149–168. https://doi.org/10.1177/1464700107078139

Ajunwa, I., Friedler, S., Scheidegger, C. E., & Venkatasubramanian, S. (2016). *Hiring by algorithm: Predicting and preventing disparate impact*. SSRN. http://sorelle.friedler.net/papers/SSRN-id2746078.pdf

Ajunwa, I., & Greene, D. (2019). Platforms at work: Automated hiring platforms and other new intermediaries in the organization of work. In S. P. Vallas & A. Kovalainen (Eds.), *Work and labor in the digital age* (pp. i–x). Emerald Publishing. https://doi.org/10.1108/S0277-283320190000033005

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bakhtin, M. M. (1984). *Problems of Dostoevsky's poetics*. (Caryl Emerson, Ed. & Trans.). MPublishing, University of Michigan Library. https://doi.org/10.5749/j.ctt22727z1

Barad, K. (1998). Getting real: Technoscientific practices and the materialization of reality. *Differences: A Journal of Feminist Cultural Studies*, *10*(2), 87–91.

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press. https://doi.org/10.1515/9780822388128

Barlas, P., Kyriakou, K., Guest, O., Kleanthous, S., & Otterbacher, J. (2021). To "see" is to stereotype: Image tagging algorithms, gender recognition, and the accuracy-fairness trade-off. *Proceedings of the ACM on Human–Computer Interaction*, *4*(CSCW3), 1–31. https://doi.org/10.1145/3432931

Barsalou, L. W. (1991). Deriving categories to achieve goals. *Psychology of Learning and Motivation*, *27*, 1–64. https://doi.org/10.1016/S0079-7421(08)60120-6

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281–1289. https://doi.org/10.1098/rstb.2008.0319, PubMed: 19528009

Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418. https://doi.org/10.1098/rstl.1763.0053

Bedau, M. A., McCaskill, J. S., Packard, N. H., & Rasmussen, S. (2010). Living technology: Exploiting life's principles in technology. *Artificial Life*, *16*(1), 89–97. https://doi.org/10.1162/artl.2009.16.1.16103, PubMed: 19857142

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. John Wiley & Sons.

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, 2(*2*), 100205. https://doi.org/10.1016/j.patter.2021.100205, PubMed: 33659914

Birhane, A., & van Dijk, J. (2020). *Robot rights? Let's talk about human welfare instead*. arXiv preprint arXiv:2001.05046. https://doi.org/10.1145/3375627.3375855

Bogen, M., & Rieke, A. (2018, December). *Help wanted: An examination of hiring algorithms, equity, and bias*. Upturn. https://www.upturn.org/reports/2018/hiring-algorithms/

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. MIT Press. https://doi.org/10.7551/mitpress/6352.001.0001

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press. https://doi.org/10.7551/mitpress/11022.001.0001

Buccella, A. (2020). Enactivism and the "problem" of perceptual presence. *Synthese*, 1–15. https://doi.org/10.1007/s11229-020-02704-1

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *FAT'20: Proceedings of the 2020 Conference on Fairness, Accountability and Transparency* (pp. 77–91). ACM.

Chemero, A. (2011). *Radical embodied cognitive science*. MIT Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Cohen, N. (2018). *The know-it-alls: The rise of Silicon Valley as a political powerhouse and social wrecking ball*. Simon & Schuster.

Collins, P. H. (2002). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. Routledge. https://doi.org/10.4324/9780203900055

Dastin, J. (2018, October 10). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazonscraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

De Jaegher, H. (2019). Loving and knowing: Reflections for an engaged epistemology. *Phenomenology and the Cognitive Sciences*, 1–24. https://doi.org/10.1007/s11097-019-09634-5

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507. https://doi.org/10.1007/s11097-007-9076-9

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447. https://doi.org/10.1016/j.tics.2010.06.009, PubMed: 20674467

Descartes, R. (1984). *The philosophical writings of Descartes* (J. Cottingham, R. Stoothoff, & D. Murdoch, Trans., Vol. 2). Cambridge University Press. https://doi.org/10.1017/CBO9780511818998

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic bodies: The continuity between life and language*. MIT Press. https://doi.org/10.7551/mitpress/11244.001.0001

Dotov, D., & Froese, T. (2020). Dynamic interactive artificial intelligence: Sketches for a future AI based on human-machine interaction. In J. Bongard, J. Lovato, L. Hébert-Dufresne, R. Dasari, & L. Soros (Eds.), *ALIFE 2020: The 2020 Conference on Artificial Life* (pp. 139–145). MIT Press. https://doi.org/10.1162/isal_a_00350

Dreyfus, H. L. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philosophical Psychology*, *20*(2), 247–268. https://doi.org/10.1080/09515080701239510

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Ferryman, K., & Pitcan, M. (2018). *Fairness in precision medicine*. Data & Society. https://datasociety.net/library/fairness-in-precision-medicine/

Fletcher-Watson, S., De Jaegher, H., van Dijk, J., Frauenberger, C., Magnée, M., & Ye, J. (2018). Diversity computing. *Interactions*, *25*(5), 28–33. https://doi.org/10.1145/3243461

Froese, T., & Taguchi, S. (2019). The problem of meaning in AI and robotics: Still with us after all these years. *Philosophies*, *4*(2), 14. https://doi.org/10.3390/philosophies4020014

Gardiner, M. (1998). The incomparable monster of Solipsism: Bakhtin and Merleau-Ponty. In M. Mayerfeld Bell & M. Gardiner (Eds.), *Bakhtin and the human sciences* (pp. 128–144). Sage. https://doi.org/10.4135/9781446278949.n9

Gershenson, C. (2013). Living in living cities. *Artificial Life*, *19*(3–4), 401–420. https://doi.org/10.1162/ARTL_a_00112, PubMed: 23834590

Glasersfeld, E. von, (1992). Declaration of the American Society for Cybernetics. In C. V. Negiota (Ed.), *Cybernetics and Applied Systems* (pp. 1–5). Marcel Decker. https://www.vonglasersfeld.com/065

Gross, J. (2020). Historicizing the self-evident: An interview with Lorraine Daston. Los Angeles Review of Books. https://lareviewofbooks.org/article/historicizing-the-self-evident-an-interview-with-lorraine-daston/

Hahn, U. (2014). The Bayesian boom: Good thing or bad? *Frontiers in Psychology*, *5*, 765. https://doi.org/10.3389/fpsyg.2014.00765, PubMed: 25152738

Hamidi, F., Scheuerman, M. K., & Branham, S. M. (2018). Gender recognition or gender reductionism? The social implications of embedded gender recognition systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). ACM. https://doi.org/10.1145/3173574.3173582

Harris, C. B., Keil, P. G., Sutton, J., Barnier, A. J., & McIlwain, D. J. (2011). We remember, we forget: Collaborative remembering in older couples. *Discourse Processes*, *48*(4), 267–303. https://doi.org/10.1080/0163853X.2010.541854

Helbing, D., Bishop, S., Conte, R., Lukowicz, P., & McCarthy, J. B. (2012). FuturICT: Participatory computing to understand and manage our complex world in a more sustainable and resilient way. *The European Physical Journal Special Topics*, *214*, 11–39. https://doi.org/10.1140/epjst/e2012-01686-y

Horgan, J. (2016, January 4). Bayes's theorem: What's the big deal? *Scientific American*. https://blogs.scientificamerican.com/cross-check/bayes-s-theorem-what-s-the-big-deal/

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169. https://doi.org/10.1017/S0140525X10003134, PubMed: 21864419

Juarrero, A. (2000). Dynamics in action: Intentional behavior as a complex system. *Emergence*, 2(*2*), 24–57. https://doi.org/10.1207/S15327000EM0202_03

Kalluri, P. (2020). Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature*, *583*(7815), 169–169. https://doi.org/10.1038/d41586-020-02003-2, PubMed: 32636520

Keyes, O. (2018). The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–22. https://doi.org/10.1145/3274357

Kyselo, M. (2014). The body social: An enactive approach to the self. *Frontiers in Psychology*, *5*, 986. https://doi.org/10.3389/fpsyg.2014.00986, PubMed: 25309471

Langton, C. G. (Ed.). (1997). *Artificial life: An overview*. MIT Press.

Linell, P. (2009). *Rethinking language, mind, and world dialogically*. Information Age Publishing.

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, *13*(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

Marková, I. (2016). *The dialogical mind: Common sense and ethics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511753602

Maturana, H. R. (2004). *From being to doing: The origins of the biology of cognition*. Karnac Books.

McGann, M., & De Jaegher, H. (2009). Self–other contingencies: Enacting social perception. *Phenomenology and the Cognitive Sciences*, *8*(4), 417–437. https://doi.org/10.1007/s11097-009-9141-7

McQuillan, D. (2018). Data science as machinic Neoplatonism. *Philosophy & Technology*, *31*(2), 253–272. https://doi.org/10.1007/s13347-017-0273-3

McQuillan, D. (2020). *Non-Fascist AI*. SocArXiv. https://osf.io/preprints/socarxiv/b64sw/

Merleau-Ponty, M. (2012). *Phenomenology of Perception* (D. Landes, Trans.). Routledge. (Original work published in 1945). https://doi.org/10.4324/9780203720714

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI and Society*, *35*, 957–967. https://doi.org/10.1007/s00146-020-00950-y

Molla, R. (2019). Google, Amazon, and Facebook all spent record amounts last year lobbying the US government. *Vox*. https://www.vox.com/2019/1/23/18194328/google-amazon-facebook-lobby-record

Nagel, T. (1989). *The view from nowhere*. Oxford University Press.

Narayanan, A. (2019). *How to recognize AI snake oil*. 2019 Arthur Miller lecture on science and ethics. https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf

Nonaka, T. (2020). Locating the inexhaustible: Material, medium, and ambient information. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00447, PubMed: 32231630

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342, PubMed: 31649194

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

O'Neil, C., & Schutt, R. (2013). *Doing data science: Straight talk from the frontline*. O'Reilly Media, Inc.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press. https://doi.org/10.4159/harvard.9780674736061

Perdomo, J. C., Zrnic, T., Mendler-Dünner, C., & Hardt, M. (2020). *Performative prediction*. arXiv preprint arXiv:2002.06673.

Preiser, R. (Ed.). (2016). *Critical complexity: Collected essays* (Vol. 6). Walter de Gruyter.

Prigogine, I., & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature*. Verso Books.

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *FAT '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 469–481). ACM. https://doi.org/10.1145/3351095.3372828

Rose, T. (2016). *The end of average: How to succeed in a world that values sameness*. Penguin UK.

Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., Altschul, D. M., Brand, J. E., Carnegie, N. B., Compton, R. J., Datta, D., Davidson, T., Filippova, A., Gilroy, C., Goode, B. J., Jahani, E., Kashyap, R., Kirchner, A., McKay, S., … McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(15), 8398–8403. https://doi.org/10.1073/pnas.1915006117, PubMed: 32229555

Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to 'solve' the problem of discrimination in hiring?: Social, technical and legal perspectives from the UK on automated hiring systems. In *FAT '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 458–468). ACM. https://doi.org/10.1145/3351095.3372849

Seth, A. K. (2014). The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND* (pp. 1–24). MIND Group.

Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F. N., Arvanitakis, G., Benevenuto, F., Gummadi, K. P., Loiseau, P., & Mislove, A. (2018). Potential for discrimination in online targeted advertising. In *FAT '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 5–19). ACM.

Star, S. L. (2002). Infrastructure and ethnographic practice: Working on the fringes. *Scandinavian Journal of Information Systems*, *14*(2), 6.

Suchman, L. (2007). *Human–machine reconfigurations: Plans and situated actions*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808418

Varela, F. J., Thompson, E., & Rosch, E. (2016). *The embodied mind: Cognitive science and human experience*. MIT Press. https://doi.org/10.7551/mitpress/9780262529365.001.0001

Véliz, C. (2020). *Privacy is power: Why and how you should take back control of your data*. Random House.

Weiser, M. (1999). The computer for the 21st century. *ACM SIGMOBILE Mobile Computing and Communications Review*, 3(*3*), 3–11. https://doi.org/10.1145/329124.329126

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W H Freeman & Co.

Wilson, B., Hoffman, J., & Morgenstern, J. (2019). *Predictive inequity in object detection*. arXiv preprint arXiv:1902.11097.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. Addison-Wesley.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Hachette Book Group.