JOURNAL OF DIGITAL SOCIAL RESEARCH

WWW.JDSR.IO ISSN: 2003-1998

VOL. 2, NO. 3, 2020, 42–64

COPRODUCTION, ETHICS AND ARTIFICIAL INTELLIGENCE: A PERSPECTIVE FROM CULTURAL ANTHROPOLOGY

Leah Govia*

ABSTRACT

Over the past five years, artificial intelligence (AI) has been endorsed as the technical underpinning of innovation. Sensationalist representations of AI have also been accompanied by assumptions of technological determinism that distract from the ordinary, sometimes unassuming consequences of interaction with its systems and processes. Drawing on scholarship from cultural anthropology, along with science and technology studies (STS), this paper examines coproduction in a Canadian AI research and development context. Through interview responses and field observations it presents sites of sociotechnical entanglement and ethical discussion to highlight potential spaces of mediation for anthropological practice. Emerging themes from the experiences of AI specialists include the negotiability of technology, an ethics of the everyday and critical collaboration. Together this returns to an initial approach into a situated understanding of artificial intelligence, negotiating with broad, sensationalist perspectives and the more commonplace, backgrounded cases of narrow research.

Keywords: cultural anthropology; artificial intelligence; ethics; coproduction

* University of Waterloo, Canada.

1 INTRODUCTION

AI research and development has seen substantial investment in Canada. Previous government commitment has sought to position the country as "a world-leading destination for companies seeking to invest in AI and innovation"[1] and in 2017, the federal government implemented a "$125 million Pan-Canadian Artificial Intelligence Strategy, the world's first national AI strategy".[2] This, along with membership in the Global Partnership on Artificial Intelligence[3] and "fast-track visa programs"[4] for tech talent have seen multiple Canadian cities placed among the fastest growing tech markets in North America. More recently too, in response to COVID-19 we've seen the development of AI-supported contact-tracing apps contributed by companies, universities and national AI institutes.[5] It is in and among these many venues that the Canadian AI context is both flourishing locally and displaying significance internationally, at least in view of globalizing, capitalist development discourse. Acting within and between these strategies are researchers and developers whose work becomes hyper-publicized as intelligent technologies continue to enter into our daily lives. Questions about research and design further emerge in this public view, urging specialists to face the social or ethical implications of the work that they do.

While it may not be possible to ensure non-harmful application of artificial intelligence, it is important to guarantee less harmful processes in its research and development. For instance, a commonly expressed concern is the need for AI to be designed with transparency. In many domains where it integrates tangibly with varying publics and stakeholders, such as in the health and financial industries, transparency has become synonymous with trust and accountability (Kim et al. 2014; Manderson et al. 2015). When seeking such transparency, it is necessary to understand how specialists engage with dynamic, sociotechnical articulations — in and of their work — as a nexus where ideas of trust and accountability also configure and emerge. Here too, as AI is drawn into common social, political, public venues, anthropological mediation becomes useful when accounting for negotiation between the imagined and realized sociotechnical contexts specialists grapple with.

---

[1] Government of Canada https://www.canada.ca/en/department-finance/news/2017/03/growing_canada_sadvantageinartificialintelligence.html

[2] CIFAR https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy

[3] Global Partnership on AI https://www.therecord.com/news/waterloo-region/2020/06/16/canada-joins-international-partnership-to-promote-responsible-ai.html

[4] Fast-track visa programs https://dmz.ryerson.ca/the-review/artificial-intelligence/

[5] TraceSCAN https://uwaterloo.ca/stories/news/new-ai-technology-will-be-used-improve-contact-tracing-covid; Mila https://globalnews.ca/news/6951846/coronavirus-contact-tracing-app/

Within anthropology, literature on artificial intelligence is still emerging and less established than in other areas of STS, but topics such as post-humanism, virtual worlds, human-machine interaction, big data and algorithms offer related insights (Born 1997; Robertson 2010; Nardi 2010; Boellstorff et al. 2012; Richardson 2015; Irani 2015; Seaver 2018). More specific to AI, an earlier inquiry by Mariella Combi (1992) focuses on the AI imaginary to display how problems and solutions, both technical and social, are constructed through human-computer relation. Similarly, the late Diana Forsythe's work during the early 1990's involves an ethnographic account of knowledge-making in an AI scientific community. She investigates shared practice and meaning to present how knowledge is localized rather than representative of a universal commonsense (Forsythe 1993ab). This is further accompanied by an extensive body of literature from science and technology studies (STS), which aims to critically examine the construction of scientific knowledge and practice. Through this field of study one can investigate the interplay between "epistemic and political processes" to demonstrate how technologies and social orders are co-produced. Including theory on the agency of things, when extended to artificial intelligence STS considers symbolic and material agencies that transform spaces, facilitate experience and create different kinds of relations — sociocultural, ethical, technical or otherwise — through coproduction (Latour and Woolgar 1986; Haraway 1988, 1990; Latour 1991, 1999; Hacking 1999; Bille and Sorensen 2007; Solomon 2008; Sismondo 2008; Ingold 2008; Jasanoff 2004, 2016). To simply illustrate, the programming decisions that specialists make when coding and the computational agencies of algorithms that arise as such, while typically viewed as technically independent are situated with certain historical, cultural or political arrangements that already inform available choices and potential outcomes. Which theories and algorithms are framed as most suitable for varying software applications, how data is represented, or the ways in which coding practices come to "matter"; these emerge through the interrelation of technical practice and the particular contexts where specialists construct knowledge of said practice (Reardon 2001).

An anthropological perspective is suited for sociotechnical analysis or for identifying sites of coproduction. It provides reflexive understanding that phenomena are all at once situated, dynamic, emergent, and in this seemingly conflicted, yet grounded plasticity is the presence of negotiation. When accessed in the creation of regulatory frameworks and policies, for example, it foregrounds a heterogeneity of publics and stakeholders for intelligent technologies constructed to suit a wide range of experiences and contexts, not only those that reproduce hegemonic, normative subscriptions of being (Mosemghvdlishvili and Jansz 2013). An anthropological approach concerned with situated knowledges and embedded action may help to

manage the technological landscapes that influence how we interact with our worlds, sometimes in ways we've yet to imagine (Haraway 1988).

2 METHODS

Data collection was supported by methods including semi-structured interviews, unobtrusive observation, archival research and textual analysis. Semi-structured interviews were conducted both in-person and virtually, guided by questions that asked respondents to share their experiences working within the field of artificial intelligence, or working with any of its associated techniques for cognate disciplines such as quantum computing. Respondents were also asked to discuss the social or ethical implications of artificial intelligence, both particular to their work and to more publicized examples. This ranged from industry professionals using machine learning techniques for business strategy, to graduate students exploring ethical algorithms in their dissertation. Interview audio was then transcribed and thematically coded, manually and with analysis software Atlas.ti.

An academic conference, AI guest talks and group meetings were the primary sites of unobtrusive observation. For example, at the Fifth Annual Conference on Governance of Emerging Technologies: Law, Policy and Ethics, I attended talks presented by Canadian and international researchers on topics specific to regulation, ethics and artificial intelligence. Entering these spaces and "becoming the phenomenon" or attempting to simulate a position similar to that of the specialists attending increased access to epistemological processes that membership within an AI-related community might afford (Franklin and Roberts 2006). Standard to an anthropological approach, this interpretive process is informed by a notion of ethnography as embodied practice and highlights the dynamic activity of the field (Cerwonka and Malkki 2008). Key respondents were later identified and include computer science professors, PhD students/candidates, a post-doctoral researcher and an industry professional (P1, P2…P7).

While each individual worked within an AI community or related space, particular emphasis was placed on those based at a Canadian institution in Southwestern Ontario, Canada. The faculty of computer science at this university is renowned for its connections to the tech industry, with graduates often finding placement in positions at companies such as Apple, Facebook, and Google. Existence of an Artificial Intelligence Group and its most recent collaboration with the Partnership on AI further evidences a concentration of AI research at the university, adding to its appeal as a source of data. Interviews were approached with the concept of engaged listening and an ethnographic imaginary meant to provide insight similar to that of participant observation (Forsey 2010). For additional data,

basic textual analysis was applied to public policy recommendations, reports, and design guides from two North American R&D organizations with designated initiatives that address the social implications of artificial intelligence. These are the Canadian Institute for Advanced Research (CIFAR) [6] and the Institute of Electrical and Electronics Engineering Standards Association (IEEE SA)[7].

3 UNDERSTANDING COPRODUCTION AND AI

Artificial intelligence has been categorized in different ways to distinguish function and capability, with terms such as "weak"/ "narrow" and "strong" AI being used, although these categories overlap and are not consistently taken up by researchers (Warwick 2013). The specialists I spoke with predominantly worked on narrow AI, which has been described as deliberately programmed, task-specific, or with capabilities restricted to a single domain (Bostrom and Yudkowsky 2014). When discussing their thoughts on the discipline, a common theme among my interlocutors was that artificial intelligence is nowhere near the level of capability displayed in the media. As one PhD student succinctly explained: "public conversation glosses over critical distinctions in what's actually possible, and what we foresee as feasible, and what's currently being done" (P3). Though others confirmed that various forms of artificial intelligence can and will continue to surpass human performance, as intended, they also echoed the words of the PhD student with reference to current applications of AI being single-purpose. One doctoral candidate recounted their conversation with a "bleeding edge researcher", stating that from a few years ago:

> The bleeding edge development is the robot can figure out when a chair is in its way, and move the chair out of its way so it can continue rolling down the hallway…so if AI were to take over the world you would not be able to stop them by putting chairs in their way…not that particular model of chair. (P2)

These specialists are aware of the sensationalist expectations crafted with public understandings of AI, in coexisting, historicized and emerging sociotechnical imaginaries, but it may not align with the technical realities of their work. The non-technical is sometimes placed external to these realities too. Here I return to the foundational suggestion that distinctions between the social and technical are often fabricated rather than actual. Technology is not constructed in isolation, but instead co-produced with

[6] CIFAR https://www.cifar.ca/ai

[7] IEEE SA https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

"social practices, identities, norms, conventions, discourses, instruments and institutions" (Jasanoff, 2004, p.3; Latour and Woolgar 1986). Identifying sites of coproduction in artificial intelligence can expose how its features are in constant entanglement while simultaneously emphasizing said features and the ways they hold potential, contingent configurations specific to the AI context, while moving beyond narratives of technological determinism.

Studies in educational settings show that like other subfields of computer science, AI is considerably practice-oriented (Kay et al. 2000). While it is seemingly obvious to state, students learn various coding languages and become familiar with how developer input influences the functions of a system. With primary actions virtually facilitated through a computer, to specialists "the central meaning of work may be writing code and building systems" (Forsythe 1993a, p. 470). This was similarly noted by a professor of computer science who explained that in AI, "at the research level it's just studying algorithms", sharing an example where "you create some image database and then you write some algorithms to classify images or something like that, but you can do all that without asking a human being to do anything" (P5). There are moments in daily practice that are conventional and distance specialists from the sociality of their work, but when the characterization of work in AI is assigned to certain structures of discourse, other topics can be sidelined or positioned as external to the technical aspects in focus (Forsythe 1993ab). It may also mask other considerations and consequences of the technologies at work:

> Artificial intelligence has always been concerned primarily with building machines that are operating independently from humans. Most of AI is building machines that have nothing to do with human beings, they're just completely separate. Even a machine that plays chess, it doesn't care that it's playing against a human. It could be playing against another machine; it's got no model of the human. Same thing for these poker-playing robots. They're not modelling human feeling they're not modelling human anything; they're just modelling the game. They're just modelling inanimate objects and that's all…that's really weird when you think about it. There's no doubt that everybody must know that intelligence has a lot to do with other people. (P5)

As the quote above indicates, the professor is attuned to a real and imagined social presence of artificial intelligence, but the positioning of AI as a technical object is, as appropriate to the discipline, most attended to. This removal of the "human variable" is a more pronounced display of how the separation of social and technical aims to leverage the universality and "effectivity" of technology (Born 1997). At the same time, this universality provides space for technologies to be aligned with larger structural and

institutional goals, which contradicts the supposed separation that sources its universality. Such a view is not uncommon and follows the positioning of science external to the social to "protect the 'value neutrality' of the scientific process" (Douglas 2007, p.127; Liu 2017). While many of the practical, operational aspects of AI appear to be separated from humans with an emphasis on features like automation, for instance, the development of automation has always been fundamentally entangled and co-productive. Even in systematic categorizations of autonomy considering independent, agential action separate from a programmer's original input, there remain many scenarios in which developers must evaluate and re-adjust the machine's operational capacities (Warwick 2013; Richardson 2015). It is because technology is shaped by constraints or conditions in design and application that technical decisions made at one point in time can impact development made at another, or vice versa (Mosemghvdlishvili and Jansz 2013). This reconfirms that in the pathways of research and development, from acquiring datasets and programming algorithms, to designing user interfaces and eventual implementation, AI is in constant coproduction.

3.1 Making the social, technical

At the conference on Governance of Emerging Technologies where part of my observation took place, during a keynote speech the founder of the Center for Human-Compatible Artificial Intelligence [8] called to both "maximize human values" and manage risk in AI by accounting for the "biggest deviation of rationality" — our wants. He expressed that by learning to predict what people want, it will become easier to develop systems that are beneficial and will require "cultural work" to reach prediction. The call for cultural work seemed to suggest a holistic survey of societies globally for a shared set of wants, following research on the use of psychological and sociological modelling for artificial intelligence. For example, in the subfield of Affective Computing a major theoretical influence comes from Affect Control Theory (ACT). This sociological theory considers the relationship between emotion and culture, categorizing patterns of affective meaning that are socially shared (Rogers et. al 2014). One of my respondents is a professor who works with building such sociological models into AI solutions through this field of research. They explained that the modelling relies on "the sort of collective consciousness or collective nature of human intelligence", which in this case is associated with affect and emotion (P5). This is then mapped to cultural contexts through AI techniques. One such mapping is exemplified by a program the

[8] Center for Human-Compatible Artificial Intelligence https://humancompatible.ai/

professor has drawn influence from in his research, known as Interact. Available for download through Indiana University (2016):

> Interact is a computer program that describes what people might do in a given situation, how they might respond emotionally to events, and how they might attribute qualities or new identities to themselves and other interactants in order to account for unexpected happenings. Interact achieves its results by employing multivariate non-linear equations that describe how events create impressions, by implementing a cybernetic model that represents people as maintaining cultural meanings through their actions and interpretations, and by incorporating repositories of cultural meanings.

The repositories of cultural meanings are formatted as dictionaries of affective meaning. These contain set identities, behaviours, and settings. Categorized by place and date, some of the listed dictionaries include U.S.A.: Indiana 2003, Japan 1989-2002, Germany 2007, and Northern Ireland 1977. Data from these dictionaries then help to model interactions between actors and objects as events and determine the probable impressions each person holds after certain event actions. Cultures are depicted as totalities within Interact and fall within a normative process supported by philosophies of science that emphasize naturally embodied dispositions substantiated by a group, corroborated as "culture". Anthropologists, however, have problematized the definition of culture as a bounded concept. Emphasizing intragroup variations and movements, they argue that cultures are not homogenous entities. [9] The categorization in Interact of place-based identity, behaviour and setting meant to determine affect and impression reproduces the definition of culture as bounded and assumes a universality of emotions. It also places social experience as something that is rigidly patterned, based on its representation as static and deterministic. Instead, the codifying of emotions is already bound by cultural interpretations of emotion in the Interact program because it is influenced by the epistemological stance of ACT. In the representations of consistent "cultures", it also simultaneously erases and reifies various social and cultural elements due to a reliance on universality. Again, anthropologists have problematized universality and homogeneity both theoretically and methodologically. An added ontological viewing would further question the universality applied to social and cultural phenomena in Interact. These phenomena are brought into existence through their delineation in the first place, rather than being universally attributed, pre-existing conditions (Coopmans et al. 2014; Hoeppe 2015). In other words, the codifying of

[9] Definitions of culture have been problematized for many years (Gupta and Ferguson 1997; Hobart 2000; Clifford and Marcus 1986; Helmreich 2001)

culture and emotions in Interact is an embodied cultural interpretation — a phenomenon brought into the world through the activity of coding itself.

While this example from the professor is a plainly demonstrated site of coproduction, others are not as immediately discernible. As sociocultural factors are datafied, they become inscriptions: "visual/textual translations and extensions of scientific practice" that frame said factors as technical objects to legitimize their presence (Latour and Woolgar 1986, p.142). In making the social, technical, these essentialized, deterministic evaluations of sociocultural phenomena appear. Alternatively, going "back to the basics" in an anthropological or STS approach that calls attention to coproduction is not just a reminder, but an available strategy for interested specialists who find concern with the structuring of data or algorithm design. Within their work specialists do craft an understanding of the sociotechnical, where systems articulate with other forms of expertise and knowledge, all of which is value-laden. They balance a range of factors including technical operations, funding influences and design compatibility while ensuring that their work is adapted for other, already existing emerging technologies and the various contexts where AI is applied (Ekbia 2008; Johnson and Wetmore 2008). It is understandable that the keynote speaker mentioned a need to both maximize human values and deal with risk in AI. Exactly how our values are being handled still needs care-full, reflexive consideration and increased interdisciplinary collaboration, as many have already called for.

Importantly, it also asks us to confront the difficulties of making AI socially and technically sustainable. Programs like Interact may begin as an exploratory project in mapping moments of human sociality, but when implemented more widely, present worries similar to that seen in cases of algorithmic bias and imbalanced datasets. Concurrently, they call on the agential capacity of AI that generates a seemingly separate, yet impactful trajectory of more-than-human expansion. The sense of agency that AI evokes, especially when projecting affective qualities, is then heightened in social perception of its systems. Aligned with studies on anthropomorphism and technology, this suggests that specialists may unintentionally act to maintain their social worlds in research and development to more easily maneuver the unpredictable relation to more-than-human agencies (Eyssel et al. 2012; Picarra et al. 2016). Common exposure to such "affective algorithms" might long remain more speculative than practical, but as initially noted, seeing to sustainable sociotechnical relations at minimum requires us to acknowledge the messiness of coproduction, from conceptualization to application. Anthropologists can contribute with further analysis and ethnographic endeavors that showcase the situatedness of what it means to "do" AI, with and beyond human relationality, while offering tools of accountability

through critical reflection on the ontological and epistemological conditions in research and development.

4 CONSIDERING ETHICS AND REGULATION

In its most public standing, the ethics of artificial intelligence is an applied ethics. Among the focus on implications or consequences, academic inquiry has also specified a combination of theory and application through subfields such as roboethics and machine ethics (Wallach et al. 2008; Dougherty 2013; Vanderelst and Winfield 2018). In practice, discussions tend to privilege certain configurations or models of ethics, mainly those influenced by European moral philosophy that frame ethics as a complex form of decision-making (Torrance 2013; Englert et al. 2014; Cervantes et al. 2016). This was further confirmed by multiple respondents when the topic of AI ethics was raised, like a PhD student specializing in computer vision noted:

> If you're familiar with various philosophical theories of ethics, a lot of them involve either satisfying constraints based on rules, Kantian deontological ethics, or optimizing some function, Utilitarian like Mill or Bentham…now these sorts of optimization are actually very important in computer science in general, also in artificial intelligence. (P3)

One way this fits within experiences of AI ethics is through a framing of complexity and the popular narrative of innovation being inherently beneficial to humanity guiding research and development (Ekbia 2008). Both professors (P4, P5) mentioned a pattern in AI, like other STEM-related disciplines, where certain breakthroughs reach a level of visibility that sparks interest in the public. The current interest surrounds work on machine learning and deep neural networks, but they explained that this happens "once every 10 years" and that there have been at least "two of these hypes in the past" for AI. The professor whose research involves constraint programming described this as techno-optimism. They shared that the view of technology solving "all the problems and it's only a good thing" is a regular interpretation at their campus when students or colleagues discuss the social implications of artificial intelligence (P4).

This was further illustrated after a guest talk by the previous director of Microsoft Research Labs, Eric Horvitz. Speaking of the then-director's proposal about autonomous driving as the solution for deaths by drunk driving, the professor shared:

> There's easy technological fixes that prevent people from driving their cars when they're drunk…You don't have to go to autonomous driving to save 40,000 people, you can do it for a few hundred dollars. Autonomous driving will add thousands and thousands to the price of a car, so it's more of a 'I

> love technology' thing as opposed to a rational decision about what's the best way to prevent these deaths. (P4)

Here he suggests that there are already existing, commonplace fixes for current problems, but they are overshadowed by techno-optimism and to some extent, a fetishization of innovation. It is a sentiment that is similarly seen with the "black box" problem in AI, where the internal operations are mostly unknown, yet the output or outcomes — when they appear to be useful or harmless — can be left unchallenged. Though the black box problem is exacerbated by an amount of data and processing too complex for individual understanding, the complexity it engenders also motivates a simpler viewing of technology as isolated. Data is usually highlighted as the likely vessel for bias in this scenario, given its more direct connection to developer input and decision-making, but algorithms are no more separate from their sociotechnical makings than the data that feeds them. With algorithms learning from "either the human-trained input or the self-learned input" specialists aim to "identify what those outcomes are of the algorithm" (P6). But as previously shared, these outcomes can mirror social orders by the very act of their structuring whereby certain technical solutions become entrenched with choices determined and made available to groups with specific social, political or economic power. Combined, this has already translated to outcomes in facial recognition technology and predictive policing that reproduce existing inequalities, largely expanding on colonial makings that continue to place Black, Indigenous and racialized communities under directed surveillance by the state (Buolamwini and Gebru 2018; Benjamin 2019). Along this view, the black box of techno-optimism where technological success masks the intricacies of research and development prompts specialists to focus on those same tasks that create concern in the first place.

Still, to many of my respondents, these tasks do not intentionally fall into the black box of techno-optimism, they merely follow what it means to do work in artificial intelligence. Here between the hype things seem a bit more mundane, but are an important point of entry for discussions on ethics. Referring to students in the undergraduate courses he teaches, a doctoral candidate explained why such discussion is sometimes hard to find:

> Their jobs are not going to be 'how to design a comprehensive framework for running autonomous cars as a company, as a societal thing'; it's going to be 'can we solve this route planning problem for autonomous cars? Can we do image recognition accurately? And these are extremely important pieces of the puzzle, but it's not the part of the puzzle that touches on ethics. And so getting them interested in it would be difficult. (P2)

This does not mean that specialists have a lack of interest in ethics or ethical discussion. It instead confirms that work in AI is characterized according to structures of discourse that traditionally emphasize the technical prominence of the field, as was examined earlier. Again, this returns to the usefulness of identifying coproduction, particularly as the doctoral candidate's example introduces how ethics is positioned as something external to the technical. Both faculty and graduate students similarly suggested that ethical discussion is considered a "challenge outside of the curriculum", or is done in an "intentional way" through workshops outside of their research (P1). One of the professors additionally suggested that it may be because "people don't like to look too closely at what they're doing I guess, 'cause it's troubling sometimes, the role that we play" (P4). For some this translates into a question of understanding or competency:

> It's hard for me to talk about ethics because I don't really understand it that well to be quite honest with you; and that's probably the same for a lot of computer scientists, artificial intelligence researchers — that we're not too clear on what ethics is. I'm trying to learn, understanding it now at this kind of cultural consensus about things that we label as good vs bad essentially, but I know that there's other aspects to it. There's these 'whether you believe that all that matters are the consequences of things', what are these deontological ethics or consequential ethics. (P5)

> I don't know if I am qualified yet to really make professional thoughts about it. I don't have an ethics background. I have a computer science background which maybe gives me insight into some areas of it, but certainly does not give me the full picture. (P2)

In the above, ethics is discussed according to some form of formalized model of thought, either as philosophical theory, or as a professional background in ethics. There also exists a designation of authority for whom may discuss ethics and how it should be done that aligns with ethics as a delegated field of study. This is further supplemented by an underlying theme of uncertainty. For these AI specialists, uncertainty can be framed as both a challenge within the technical side of computer science and based on their responses, one that is ethical. On the technical side, there is the problem of "reasoning under uncertainty" that is and "has always been a key challenge in artificial intelligence" (P3). The other challenge is uncertainty that accompanies the ethical dimensions of emerging technologies and becomes normalized through the placement of ethics as external to practice, or as an add-on that specialists are not positioned to access (Akama et al. 2015). The dominant presence of ethics as an independent field of expertise and a major source of uncertainty, when taken up by specialists facilitates a detachment from ethical practice despite

being deeply implicated in ethics-focused structures of discourse, sensationalized or otherwise.

As this exploration of ethical practice comes with a partially normative approach, it is helpful to address a context where ethics enters narrow AI research and development more explicitly. In the subfield of machine ethics (ME) there is focus on ethical embodiment by intelligent machines (Brundage 2014; Vanderelst and Winfield 2018). Value judgments of morality are referenced here and often follow the two theories of ethics my respondents mentioned: deontological and teleological. While these models form top-down and bottom-up approaches, because of the functions they feature (e.g. utility function), there are issues with constraints and optimization where "some rather technical properties of the function" make it "very hard to find the best solution" (P3). Trade-offs between different group interests are one such complication depending on the functions used (P3). Also interesting to note, is that machine ethics recognizes the agency of AI and the extent to which it catalyzes ethical practice. AI agents are categorized here as implicit or explicit to indicate the "source" of ethics, either from the designer or from the machine's self-learning (Anderson and Anderson 2007; Veruggio and Abney 2011). Thus, in one way ME queries the performance of human ethics acting upon machine and in another it holds concern for AI as an independent, ethical agent. From both, it is as if human and machine intertwine through a sociotechnical ethic where the very relation to another entity designates an "implicit moral relationship" (Scheper-Hughes 1995). Extended to the broader discussion on ethics, this reintroduces questions of accountability when facing harmful AI outcomes and forwards action for a new set of sociotechnical, legal precedents, rights, and debates on the positionality of technologies by those deemed responsible for the public good (Ekbia 2008; Nota 2015).

In any event, whether for ethical AI or an ethics of AI, it is possible to tend to separations of ethical practice and recall coproduction by highlighting some of the ways that AI specialists configure the ethical in the everyday. As an anthropology of ethics this seeks to understand how morality is manifested and maintained in the range of experiences, contexts and interactions of individuals, agents and communities, for themselves and with others (Zigon 2010; Lambek 2010). It emphasizes how morality is not a closed system, but is relational and radically context-dependent. In this way, even uncertainty becomes an ethical relation. Given a gap in the literature on AI ethics and anthropology, further studies are needed to strengthen this approach, but a focus on ordinary, everyday practices and their ethical relations is one place to start. This can rely on ethnographic and participatory research in AI contexts, providing insights on how the ethical is situated in certain positionalities, where sites of coproduction emerge, and how this interlocks with features surrounding governance, public trust

in science or responsible design, to name a few broad examples. Engaging in this way could create and reconfigure choices in the negotiability of AI that expand access to the research and development of artificial intelligence, demonstrating transparency and building trust.

4.1 Fitting in regulation

When talking of ethics and regulation it is difficult to introduce weak/narrow AI without the influence of strong AI. Currently, the former is more practically delimited in discourse because application outcomes reach foundational structures and social institutions such as labour, security and healthcare.[10] The latter frequents sensationalized displays given its historicized presence in science fiction throughout literature and popular media, but still feeds into public communication of AI development, weak or strong. Together they foster anxious imaginaries in the public and include a perceived loss of agency where automation techniques are continually "becoming on-par or even better than human experts" (P6). Although P6 is referring to efficiency or accuracy in technical tasks, this "better than human" notion may in fact shift AI into positions of increased authority and affect how we orient ourselves with the world around us (Turner 2007; Muller 2014). Of course, among such sociotechnical barriers are opportunities to unsettle the conditions that arrange social, technical, or even ecological phenomena within hierarchies of value. Through the avoidance of "both social and scientific determinism", once more STS and anthropology supply space for and attention to alternative forms of regulation that might be viewed as a means of reconciling the many types of agency and artificial intelligence (Irwin 2008).

Frameworks and standards for ethical practice guiding narrow AI research and development are varied and localized. In speaking with respondents and from my field observations there are informal forms of best practice or documents employed through their local affiliations, but anything encompassing does not appear to be feasible. Here, themes of restraint and accountability emerge. A co-founder and CEO of an AI start-up in the talent acquisition industry preferred the term moderation rather than regulation, explaining that it would be better not to stop the "trajectory of technology" this way (P6). Techno-optimism appears again in his response, along with notions of technological determinism and isolation that were examined in previous sections. Simultaneously, specialists navigate the ethical urgency that emerges with AI. This comes out in conversations on regulation that are worn with uncertainty, especially at intersections of ethics and governance. Certain topics reasonably dominate

[10] Telehealth and artificial intelligence https://techcrunch.com/2017/04/19/ada-health/

due to their high-risk characterizations, like lethal autonomous weapons and technological unemployment:

> For some reason we're okay with people killing other people, but having an AI agent decide to kill a person people are less comfortable with (P2)

> In my opinion, AI is going to kill people. Not in the way that everyone thinks it's going to kill people, but people are going to die because of artificial intelligence. There is going to be job loss and it's going to be rapid and rampant. Now, the whole idea of people saying, well 're-skilling, re-training' that means the upending of an entire ecosystem called our current public education system which hasn't been revised since the first industrial revolution when it was generated (P6)

To implement regulation in these areas, many specialists outlined the importance of collaboration between politicians, legal scholars, application area experts and other AI specialists when creating frameworks or policies. A post-doctoral researcher interested in quantum computing and neural networks also showed how this compares with their regular interactions in the research community:

> It's also a very insular community right, like I personally know people at all of those companies, at high-up positions, and I'm like 3rd year post-doc I'm not a super senior person. My bosses personally know the founders of the groups at those companies right, so it's a very close-knit community that everybody knows everybody. So you're almost self-regulating just by the fact that the community is so small (P7)

This may indicate a slight disconnect between some research communities and larger planning for regulation, but as another respondent reminded "one thing that people often don't think of in the general public discourse is that somebody is going to have to actually write the programs that do these things"; that ultimately, those involved will have to listen to computer scientists about what is computationally possible (P3). It is an important consideration, but as this paper has also reminded, there are more than computational factors that will affect what is possible. Evidently, community and collaboration both influence what is possible in regulation and becomes a form of regulation itself. Collaboration may also act as ethical practice through its relationality. In North America, this has been visible in initiatives by organizations like The Canadian Institute for Advanced Research (CIFAR) and the IEEE. In the Canadian context, federally funded CIFAR is heralded for its interdisciplinary research and global network of commitments as it leads the country's national AI strategy. They produce policy recommendations and reports while creating special-interest workshops for programs analyzing AI and society. The American-based IEEE also has defined output through its Global Initiative

on Ethics of Autonomous and Intelligent Systems. Both strongly advocate a collaborative approach to ethics and regulation, emphasizing thought leadership across academia and industry.

Through basic textual analysis of Building an AI World[11], Rebooting Regulation[12], and Ethically Aligned Design[13], there appears to be limited inclusion of thought leaders in groups beyond dominant technical, legal, psychological and economic backgrounds. Little reference is given to contributions by historically underrepresented communities, researchers and practitioners who have already evoked many of the themes explored in this paper involving sociotechnical coproduction, situatedness and critical reflexivity (Gasparotto 2016; Winchester III 2019; Mohamed et al. 2020). It is here that the significant underrepresentation of Black, Indigenous, and racialized persons, women in particular, is once again apparent. Underrepresentation in STEM has been well-documented and despite equity initiatives continues to persist while certain ontological and epistemological conditions tolerate a critical lack of reflexivity (Morganson et al. 2010; Fontana et al. 2013). Certainly, the insular community referred to by the post-doc is not a revelation; neither is its presence being replicated in regulation. As mentioned, even in plans claiming "diversity and inclusion", peoples, knowledges and ways of being remain excluded because source institutions are not distanced from the underlying structures or discourses that background not only their DEI initiatives, but positionalities in leadership, e.g. related histories, identities, and practices (Ahmed 2007). Similarly, in the documents noted earlier, conventional expertise and disciplinary boundaries structure access and standards for ethical discussion. It reinforces certain definitions of ethics, questions of ethical practice, and what the major social implications of AI are (Reardon 2001). Choices and solutions are similarly narrowed and limited.

By "going back to the basics" in this way with an understanding of coproduction through critical inquiry, it may become easier to avoid such black-boxed conditions for AI ethics and regulation. The post-doctoral researcher shows how it might occur, even if limited, given their recognition of how insular the research community is and the way it translates to "self-regulation". Alternate action acknowledges space for critical collaboration, though this requires future analysis to substantiate. Finally, critical collaboration as regulatory practice includes multiple

[11] Building an AI World https://www.cifar.ca/docs/default-source/ai-society/buildinganaiworld_eng.pdf
[12] Rebooting Regulation https://www.cifar.ca/docs/default-source/ai-reports/rebooting-regulation-exploring-the-future-of-ai-policy-in-canada.pdf?sfvrsn=616c04f3_8
[13] Ethically Aligned Design https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

publics beyond thought leadership in academia or industry. It comes back to a request across disciplines, anthropology included, for engagement with non-specialized communities beyond research participation, although institutional and funding restrictions may hinder efforts in knowledge translation (Hayden 2007). The earlier discussed efforts by CIFAR and the IEEE do acknowledge this too, but it is not yet clear how accessible their feedback processes will be. Moving forward, the notion of critical collaboration presented here is just one of many considerations that have informed or acted alongside potential regulatory practices but may need reassessment throughout AI research and development.

5 CONCLUSION

Set in a Canadian context, this paper investigates coproduction and artificial intelligence from an anthropological perspective and is supplemented by foundational STS theory. Through noticeable examples of coproduction, first I introduce an anthropological approach to sociotechnical analysis of artificial intelligence, including the negotiability of technology. Interview responses and field observations specifically highlight the experiences of AI specialists and the ways in which sociocultural and technical elements entangle in the everyday. Next, AI ethics is situated with an anthropology of ethics through discussions on techno-optimism and conditions of uncertainty. Finally, accompanied by basic textual analysis of CIFAR and IEEE documents, regulation and ethical practice are addressed with the recommendation of critical collaboration that calls for additional reflexivity in public R&D practice.

It is important to note that this research is limited, particularly by a small sample size and reduced observation timeframe. As a result, the primary data can only represent a specific, localized Canadian context aligned with those already interested in the present topic. Despite such limitations, it acts as an initial return to a situated understanding of artificial intelligence and proposes further analysis from anthropological perspectives. It also indicates how STS can help to navigate the tensions that emerge when technical decisions are at odds with their wider social contexts. This is most noticeable in public perceptions of AI where imagined possibilities are complicated with the realities of technology. Again, additional study is surely required to go beyond my brief focus on a small grouping of specialist experiences in artificial intelligence, to the great variety of communities, discourses and processes that continue to emerge. It would be encouraging to see future works include ethnographies of applied AI and public knowledge settings, feminist analysis of AI systems in healthcare, or perhaps participatory action research on globalizing AI governance. A digital ethnographic study of machine ethics, the field

focusing on ethical embodiment by intelligent machines, might also be of interest (Anderson and Anderson 2007). In our attempts to secure both equitable and non-harmful outcomes from artificial intelligence, returning to a basic, but critical understanding of sociotechnical coproduction, along with how we reach this understanding is important.

FUNDING STATEMENT AND ACKNOWLEDGMENTS

This paper draws on research supported by the Social Sciences and Humanities Research Council (Canada).

REFERENCES

Ahmed, S. (2007). ”You end up doing the document rather than doing the doing”: Diversity, race equality and the politics of documentation. Ethnic & Racial Studies, 30(4), pp. 590-609. https://doi.org/10.1080/01419870701356015

Anderson, M., & Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. AI Magazine, 28(4), pp. 15-15. ttps://doi.org/10.1609/aimag.v28i4.2065

Akama, Y., Pink, S., & Fergusson, A. (2015, April). Design+ Ethnography+ Futures: Surrendering in Uncertainty. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 531-542.
https://doi.org/10.1145/2702613.2732499

Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code (1st ed.). Polity.

Bille, M., & Sørensen, T. F. (2007). An anthropology of luminosity: The agency of light. Journal of Material Culture, 12(3), pp. 263-284. https://doi.org/10.1177/1359183507081894

Boellstorff, T., Nardi, B., Pearce, C., & Taylor, T. L. (2012). Ethnography and virtual worlds: A handbook of method. Princeton University Press.

Born, G. (1997). Computer software as a medium: Textuality, orality and sociality in an artificial intelligence research culture. In Rethinking Visual Anthropology, pp. 139-69.

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In The Cambridge Handbook of Artificial Intelligence, 1, pp. 316-334.

Brundage, Miles. (2014). "Limitations and risks of machine ethics." Journal of Experimental & Theoretical Artificial Intelligence, 26, pp. 355-372. https://doi.org/10.1080/0952813X.2014.895108

Cervantes, J. A., Rodríguez, L. F., López, S., Ramos, F., & Robles, F. (2016). Autonomous agents and ethical decision-making. Cognitive

Computation, 8(2), pp. 278-296. https://doi.org/10.1007/s12559-015-9362-8

Cerwonka, A., & Malkki, L. H. (2008). Improvising theory: Process and temporality in ethnographic fieldwork. University of Chicago Press.

Clifford, J., & Marcus, G. E. (Eds.). (1986). Writing culture: the poetics and politics of ethnography: a School of American Research advanced seminar. Univ of California Press.

Combi, M. (1992). The imaginary, the computer, artificial intelligence: A cultural anthropological approach. AI & Society, 6(1), pp. 41-49. https://doi.org/10.1007/BF02472768

Coopmans, C., Vertesi, J., Lynch, M. E., & Woolgar, S. (Eds.). (2014). Representation in scientific practice revisited. MIT Press.

Dougherty, M. (2013). Something Old, Something New, Something Borrowed, Something Blue Part 2: From Frankenstein to Battlefield Drones; A Perspective on Machine Ethics. Journal of Intelligent Systems, 22(1), pp. 1-7. https://doi.org/10.1515/jisys-2013-001

Douglas, H. (2007). Rejecting the Ideal of Value-Free Science. In Value-Free Science? Ideals and Illusions. Oxford: Oxford University Press.

Ekbia, H. R. (2008). Artificial dreams: the quest for non-biological intelligence. Cambridge University Press.

Englert, M., Siebert, S., & Ziegler, M. (2014). Logical limitations to machine ethics with consequences to lethal autonomous weapons. arXiv arXiv:1411.2842

Eyssel, F., Kuchenbrandt, D., Hegel, F., de Ruiter, L. (2012). "Activating Elicited Agent Knowledge: How Robot and User Features Shape the Perception of Social Robots." Robot and Human Interactive Communication, pp. 851-857. doi: 10.1109/ROMAN.2012.6343858.

Fontana, M., Wells, M. A., & Scherer, M. C. (2013). A holistic approach to supporting women and girls at all stages of engineering education. Proceedings of the Canadian Engineering Education Association (CEEA). http://ojs.library.queensu.ca/index.php/PCEEA/article/view/4873

Forsey, M. G. (2010). Ethnography as participant listening. Ethnography, 11(4), pp. 558-572. https://doi.org/10.1177/1466138110372587

Forsythe, D. E. (1993a). Engineering knowledge: The construction of knowledge in artificial intelligence. Social Studies of Science, 23(3), pp. 445-477. https://doi.org/10.1177/0306312793023003002

Forsythe, D. E. (1993b). The construction of work in artificial intelligence. Science, Technology, & Human Values, 18(4), pp. 460-479. https://doi.org/10.1177/016224399301800404

Franklin, S., & Roberts, C. (2006). Born and made: An ethnography of preimplantation genetic diagnosis. Princeton University Press.

Gasparotto, M. (2016). Digital colonization and virtual indigeneity: Indigenous knowledge and algorithm bias. https://doi.org/doi:10.7282/T3XG9TFG

Gupta, A., & Ferguson, J. (Eds.). (1997). Culture, power, place: Explorations in critical anthropology. duke University press.

Hacking, I. (1999). The social construction of what? Harvard university press.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. Feminist Studies, 14(3), pp. 575-599. https://www.jstor.org/stable/3178066

Hayden, C. (2007). Taking as giving: Bioscience, exchange, and the politics of benefit-sharing. Social Studies of Science, 37(5), pp. 729-758. https://doi.org/10.1177/0306312707078012

Helmreich, S. (2001). After culture: reflections on the apparition of anthropology in artificial life, a science of simulation. Cultural Anthropology, 16(4), pp. 612-627. https://www.jstor.org/stable/656650

Hobart, M. (2000). After culture: Anthropology as radical metaphysical critique. Duta Wacana University Press.

Hoeppe, G. (2015). Representing Representation. Science, Technology, & Human Values, 40, pp. 1077-1092. https://doi.org/10.1177/0162243915594025

Indiana University. (2016). "Interact" http://www.indiana.edu/~socpsy/ACT/interact.htm.

Ingold, T. (2008). When ANT meets SPIDER: Social theory for arthropods. In Material agency (pp. 209-215). Springer.

Irani, L. (2015). Justice for 'data janitors'. Public Culture, 15. https://www.publicbooks.org/justice-for-data-janitors/

Irwin, A. (2008). STS Perspectives on Scientific Governance. In The handbook of science and technology studies, 24, pp. 583.

Jasanoff, S. (Ed.). (2004). States of knowledge: the co-production of science and the social order. Routledge.

Jasanoff, S. (2016). The ethics of invention: technology and the human future. WW Norton & Company.

Johnson, D. G., & Wetmore, J. M. (2008). STS and Ethics: Implications for Engineering Ethics. In The handbook of science and technology studies, 23, pp. 567.

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. Communications of the ACM, 57(3), pp. 78-85. https://doi.org/10.1145/2500873

Kay, J., Barg, M., Fekete, A., Greening, T., Hollands, O., Kingston, J. H., & Crawford, K. (2000). Problem-based learning for foundation

computer science courses. Computer Science Education, 10(2), pp. 109-128. https://doi.org/10.1076/0899-3408(200008)10:2;1-C;FT109

Lambek, M. (Ed.). (2010). Ordinary ethics: Anthropology, language, and action. Fordham University Press.

Latour, B., & Woolgar, S. (1986). Laboratory life: The construction of scientific facts. Princeton University Press.

Latour, B. (1991). We have never been modern. Harvard University Press.

Latour, B. (1999). Pandora's hope: essays on the reality of science studies. Harvard University Press.

Liu, Jennifer A. (2017). Situated stem cell ethics: beyond good and bad. Regenerative Medicine, 12, pp. 587-591. https://doi.org/10.2217/rme-2017-0059

Manderson, L., Davis, M., Colwell, C., & Ahlin, T. (2015). On secrecy, disclosure, the public, and the private in anthropology: an introduction to supplement 12. Current Anthropology, 56(S12), pp. 183-190. https://doi.org/10.1086/683302

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. Philosophy & Technology, pp. 1-26. https://doi.org/10.1007/s13347-020-00405-8

Morganson, V. J., Jones, M. P., & Major, D. A. (2010). Understanding women's underrepresentation in science, technology, engineering, and mathematics: The role of social coping. The Career Development Quarterly, 59(2), pp. 169-179. https://doi.org/10.1002/j.2161-0045.2010.tb00129.x

Mosemghvdlishvili, L., & Jansz, J. (2013). Negotiability of technology and its limitations: The politics of App development. Information, Communication & Society, 16(10), pp. 1596-1618. https://doi.org/10.1080/1369118X.2012.735252

Muller, Vincent C. (2014). Risks of general artificial intelligence. Journal of Experimental & Theoretical Artificial Intelligence, 3, pp. 297-301. https://doi.org/10.1080/0952813X.2014.895110

Nardi, B. (2010). My life as a night elf priest: An anthropological account of World of Warcraft. University of Michigan Press.

Nota, C. (2015). AGI Risk and Friendly AI Policy Solutions. Retrieved from https://cpnota.github.io/nota_agi_risk.pdf

Picarra, N., Giger, J.C., Pochwatko, G., and G. Goncalves. (2016). Making sense of social robots: A structural analysis of the layperson's social representation of robots. Revue europeenne de psychologie appliquee, 1-pp. 13.

Reardon, J. (2001). The human genome diversity project: a case study in coproduction. Social Studies of Science, 31(3), pp. 357-388. https://doi.org/10.1177/030631201031003002

Richardson, K. (2015). An anthropology of robots and AI: Annihilation anxiety and machines. Routledge.

Robertson, J. (2010). Gendering humanoid robots: Robo-sexism in Japan. Body & Society, 16(2), pp. 1-36. https://doi.org/10.1177/1357034X10364767

Rogers, K. B., Schröder, T., & von Scheve, C. (2014). Dissecting the sociality of emotion: A multilevel approach. Emotion Review, 6(2), pp. 124-133. https://doi.org/10.1177/1754073913503383

Scheper-Hughes, Nancy. (1995). The Primacy of the Ethical: Propositions for a Militant Anthropology. And Responses. Current Anthropology, 36, pp. 409-440. https://doi.org/10.1086/204378

Seaver, N. (2018). What should an anthropology of algorithms do? Cultural Anthropology, 33(3), pp. 375-385. http://orcid.org/0000-0002-3913-1134

Solomon, M. (2008). STS and Social Epistemology of Science. In The handbook of science and technology studies, 10, pp. 241.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous And Intelligent Systems, Version 2. IEEE. https://standards.ieee.org/industry-connections/ec/autonomous-systems/index.html

Torrance, S. (2013). Artificial agents and the expanding ethical circle. AI & Society, 28(4), pp. 399-414. https://doi.org/10.1007/s00146-012-0422-2

Turner, Bryan S. (2007). Culture, technologies and bodies: the technological Utopia of living forever. The Editorial Board of the Sociological Review, pp. 19-36. https://doi.org/10.1111/j.1467-954X.2007.00690.x

Vanderelst, D., & Winfield, A. (2018, December). The dark side of ethical robots. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (pp. 317-322). https://doi.org/10.1145/3278721.3278726

Veruggio, G., & Abney, K. (2011). Roboethics: The Applied Ethics for a New Science. In Robot ethics: The ethical and social implications of robotics, 22, pp. 347.

Wallach, W., Allen, C., & Smit, I. (2008). Machine morality: bottom-up and top-down approaches for modelling human moral faculties. AI & Society, 22(4), pp. 565-582. https://doi.org/10.1007/s00146-007-0099-0

Warwick, K. (2013). Artificial intelligence: the basics. Routledge.

Winchester III, W. W. (2019). Engaging the Black Ethos: Afrofuturism as a Design Lens for Inclusive Technological Innovation. Journal of Futures Studies, 24(2), pp. 55-62. https://jfsdigital.org/articles-and-essays/vol-24-no-2-december-2019/engaging-the-black-ethos-

afrofuturism-as-a-design-lens-for-inclusive-technological-innovation/

Zigon, J. (2010). Moral and ethical assemblages. Anthropological Theory, 10(1-2), pp. 3-15. https://doi.org/10.1177/1463499610370520