Beneath the Hype: Engaging the Sociality of Artificial Intelligence

by

Leah Govia

A thesis

presented to the University of Waterloo

in the fulfillment of the

thesis requirement for the degree of

Master of Arts

in

Public Issues Anthropology

Waterloo, Ontario, Canada, 2018

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Abstract

Artificial intelligence (AI) is highly visible in today's public media. With potential uses across domains such as healthcare, labour and transportation, its capacity to impact human lives is widely apparent. As it continues to enter into public view, concerns surrounding its research and application also arise. Here, narratives of techno-optimism, technological determinism, and dystopia often shape the AI imaginary with sensationalist displays of super-intelligence and existential concern. Counterpoised to these representations, this thesis investigates the sociality that inheres in everyday practices within artificial intelligence as emerging technology and as a field of study. Drawing on methods and scholarship from STS and socio-cultural anthropology, I explore the attitudes and experiences of specialists to analyze how entanglements of the socio-cultural, ethical and technical appear within more mundane, everyday practices of AI. Often overshadowed by popular, sensationalized understandings of technology, the focus on such experiences and practices allows for an initial view into a situated understanding of AI beneath the hype.

Acknowledgements

Thank you to friends and faculty from the Department of Anthropology at the University of Waterloo for the support during my time at the university. Deepest thanks to my supervisor, Jennifer Liu. I will always be grateful for what you have taught me and thankful for your patience and guidance throughout this process. Thank you to my committee members, Adrienne Lo and Götz Hoeppe for their guidance and support too. Best wishes to the graduate cohort. I am glad to have shared this time with you all. Thank you to my interlocutors; to the people, places and experiences that propel my love for anthropology. Finally, to my parents, brothers, and closest friends — thank you always.

This research was supported by the Social Sciences and Humanities Research Council of Canada.

Table of Contents

List of Figures

Chapter 1

Artificial Intelligence in the Realm of Public Issues Anthropology

Towards the end of 2016, six North American tech industry giants — Amazon, Apple, Google, Facebook, IBM and Microsoft — announced a collaboration to study the impacts of artificial intelligence and create best practices for its management. Now established as the *Partnership on AI*[1], this consortium is only one influential example of the increased attention brought to emerging technologies within the last five years. Artificial intelligence (AI) in particular has been given more vigorous focus due to its pronounced presence as a topic of public concern. Despite its often sensationalized media portrayal, as both a field of study and emerging technology, artificial intelligence has had a historicized presence in the public. In the North American setting, it falls within a rich history of science fiction in literature, but also in movies, video games and other popular media. From classics such as Asimov's (1950) *I, Robot* to blockbusters such as the *Terminator* series, AI has engendered a view that balances fantasy and anxiety in the public.

Understanding of the public's perception of AI — that which is often inspired by the storied mediums previously mentioned — is complicated with the realities of the technology's capabilities. In the case of AI, it appears that some of the worries imagined through the fictional setting have begun to bleed into reality. Multiple leading experts in AI-related industries and disciplines have provided public statements detailing how AI will become a cause for concern in many domains of our societies. While still often limited to specific tasks, technological systems have generally surpassed the capabilities of human action, and now more apparently surpass that of human thought (Cellan-Jones 2014). As the systems are applied to infrastructural aspects of

[1] Partnership on AI https://www.partnershiponai.org/

society — transportation, healthcare, security, labour, welfare — AI will become more embedded in the everyday, ordinary experiences with noticeable public interaction. How the public perceives these developments will depend on many factors, but one most central is trust. It is unsurprising, then, that as these technologies continue to emerge and are publicized, questions surrounding their design, application and impact bring AI's social and ethical implications to the forefront as a public issue.

In detailing the significance of artificial intelligence and its potential impacts on the public, sociality and ethics becomes a central point of discussion. As this thesis investigates the sociality of artificial intelligence, it considers how the socio-cultural, ethical and technical entangle in its practice. Through the attitudes and experiences of AI specialists, it focuses on articulations in more, mundane, everyday contexts beneath many of the popular views of AI. From here, it steps into a more situated understanding of AI and encourages work that closes the gap in engagement between multiple stakeholders — including the public — that are impacted by its research and technologies.

For artificial intelligence to be truly trustworthy, its technologies must be constructed and applied in ways that are suitable for a range of experiences and contexts, not only in those that reproduce hegemonic, normative subscriptions of being. By acknowledging the need for negotiability at each stage of research, development and application, it may become easier to account for the potential outcomes of the socio-technological existence that is artificial intelligence (Mosemghvdlishvili and Jansz 2013). Again, as artificial intelligence is increasingly brought into public view, questions about its various designs, applications and impacts are also publicized. While we may not be able to predict how intelligent technologies will act — even if

they are modelled on human behavior and researcher values — it is important to guarantee that they act in a non-harmful way given that they must interact with the public.

Finally, the intended publication venue for this research is the peer-reviewed journal *AI & Society: Knowledge, Culture and Communication*. This multidisciplinary journal appeals to discussion on "societal issues" surrounding the design, development, application, and regulation of emerging technologies. In addition, it brings attention to features such as the socio-cultural and ethical implications of emerging technologies while promoting the diverse voices who engage with their outcomes. Previous topics explored in the journal include AI and governance, security, identity, and welfare.

Chapter 2

Beneath the Hype: Entangled in the Sociality of Artificial Intelligence

2.1 Introduction

> *"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"*

While such a prediction from an industry CEO represents just one of many possibilities, it is difficult to deny the potential this emerging technology has to impact humanity in ways that will eventuate within our lifetimes. In current discussions on technology, artificial intelligence (AI) has become a media buzzword generating both hope and fear. Underlain by a long history within science fiction, it has garnered the interest of both techno-optimists and dystopic alarmists. Discourses of the technological shared and reproduced in these ways — including contemporary speculations by prominent figures Elon Musk and the late Stephen Hawking — have helped to shape public imaginations of artificial intelligence. As a counterbalance to sensationalist representations of AI, and also opposed to the technological determinism that such views often assume, this thesis investigates the sociality that inheres in everyday practices within artificial intelligence. More specifically, drawing on scholarship from STS and socio-cultural anthropology, I explore the attitudes and experiences of specialists in order to analyze how entanglements of the sociocultural, ethical and technical emerge within more mundane practices of AI research and technology. A focus on everyday practices allows for an initial view into a situated understanding of AI that is often overshadowed by the popular or sensationalized understandings.

AI in the Literature

As a field of study, AI may be described as the research of technological agents that operate with intelligence. While not entirely agreed upon, in the field intelligence has been generally defined as an agent's "ability to [efficiently] optimize the world according to their preferences" (Muehlhauser and Salamon 2012, 2). In other words, it is an agent's — human or not — ability to achieve goals based on the resources available to them across multiple domains. As emerging technology, these AI agents typically operate with problem-solving formulas known as algorithms, which interpret large quantities of information or data to reach the preferred outcome set by a programmer (Warwick 2012). Intelligent agents and systems have been categorized in different ways to distinguish how capable they are, with terms such as weak and strong AI being used, although these categories overlap and are not consistently taken up by technologists (Warwick 2012). This thesis features more of the former, also known as narrow AI. Technologies in this category are described as task-specific and "goal-directed" as they typically have "competence…in a single, restricted domain" and are "deliberately-programmed" this way (Bostrom 2014). Popular examples of narrow AI include virtual assistants such as Apple's Siri and autopilot systems found in transportation such as planes (Markoff 2015; Shankland 2017).

The literature on artificial intelligence within anthropology is still developing and less established than in other fields of STS, but topics including post-humanism, virtual worlds, and human-machine interaction offer related insights (Born 1997; Robertson 2007; 2010; Nardi 2010; Boellstorff 2012; Richardson 2015). In an earlier inquiry, Mariella Combi (1992) focuses on the AI imaginary in relation to how problems and solutions — both technical and social — are constructed through human-computer relations. This contributes to ideas on socio-technological co-production through the AI lens. Similarly, Diana Forsythe's work in the early 1990's involved

an ethnographic account of knowledge-making in an AI scientific community. In this, she investigates shared practices and meanings, accompanied by an anthropological intervention that highlights how knowledge is localized rather than representative of a universal commonsense (Forsythe 1993ab). By examining knowledge-making amongst AI experts in their practices at work — for example, where certain industry traditions influence technical decisions — Forsythe's work interacts with the STS approach at focus in this thesis and evidences how the social inheres in the technological, and vice versa.

Consideration of how socio-cultural, ethical and technical elements entangle within artificial intelligence requires some understanding of socio-technological co-production. This co-production is already seen through endeavors in user experience design and telehealth, where social and technical components are made explicitly co-dependent and simultaneously occurring, from stages of product conceptualization to application[2]. But there also already exists an established body of literature within STS that shows how technologies and social orders co-produce one another (Latour 1999; Solomon 2008). Foundational scholarship includes the work of Ian Hacking in *The Social Construction of What?* (1999), along with the seminal work of Sheila Jasanoff, *States of Knowledge: The Co-Production of Science and the Social Order* (2004). They provide evidence for theory that explains how science and society are "underwriting the other's existence" and how with such understanding, we should avoid "both social and scientific determinism" (Irwin 2008, 590). Expanding upon this in her recent contribution, Jasanoff's analysis in *The Ethics of Invention: Technology and the Human Future* (2016) challenges dominant views of socio-technological interaction to further encourage an understanding that technology is not apolitical or amoral. This includes discussion on topics such as technological determinism, technocracy and unintended consequences that can damage people and places

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

interacting with technology. Again, these works explain that technology is co-productive with the social and their intersection should be considered when discussing, developing and applying the 'products' of science and technology.

Accompanying co-production is scholarship on the agency of 'things' that would position emerging technologies like AI as social agents (Latour 2000; Bille and Sorensen 2007; Ingold 2009; Edensor 2011; Olsen 2012). This means that 'things' also act in ways typically associated with human and non-human animals, being capable of relational, autonomous action within an environment. Thus, based on the literature it can be said that artificial intelligence manifests both direct physical and symbolic material agency. As technology and as a field of thought or practice, it has the ability to transform spaces, facilitate experience with various agential actors, and can create different kinds of relations — be they social, cultural, ethical, or technical (Bille and Sorensen 2007; Ingold 2009). AI can re-frame what it means to be human, brings people into the realm of human-machine interaction and produces new social orders as a point of contact for entanglement (Latour 1991). In these ways agency becomes another indicator of co-production and even helps to identify sites of co-production.

2.2 Methodology

This thesis utilizes anthropological methods, including unobtrusive observation, semi-structured interviews, archival research, and media analysis to investigate entanglements of the socio-cultural, ethical and technical in artificial intelligence. What are some of the ways that technologists discuss the social and technological in AI? What is their making of the cultural? How do they engage with concepts of the ethical? Such questions guide my analysis of experiences shared by individuals working with narrow AI. In this context, AI specialists are defined as individuals from both industry and academic domains whose work or research has a concentrated focus on artificial intelligence. This includes a range of individuals and examples such as tech industry professionals applying AI techniques to business strategies, university professors researching machine learning, and graduate students who specialize in AI-related research.

Observation occurred over three days at The Fifth Annual Conference on Governance of Emerging Technologies: Law, Policy and Ethics at Arizona State University. This event was attended by a multi-disciplinary group whose work focuses on emerging technologies including AI. In this research, such observation allows a "means of doing so"; the reflexive-consciousness of being in the field and both learning and observing enables analysis that is adaptive to emergence and process (Franklin and Roberts 2006). This makes it particularly suitable in an AI context. By "becoming the phenomenon" and attending the conference, or simulating the position of the specialists attending, it is possible to access the epistemological processes involved in the experience of being a member in an AI-related community (Franklin and Roberts 2006). This interpretive process is informed by the notion of ethnography as embodied practice

and highlights the dynamic activity of the field (Cerwonka and Malkki 2008; Bernard 2011; Boellstorff 2012).

Along with unobtrusive observation, semi-structured interviews were a focal methodological component. Seven participants were interviewed in this study — two university professors, two PhD students, one MSc student, one post-doctoral researcher, and one industry professional. Each of these individuals work within AI communities or related spaces, but with particular emphasis on those situated at the University of Waterloo, Canada. The faculty of computer science at this institution is renowned for its connections to the tech industry, with graduates often finding placement in positions at companies such as Google, Facebook, and Apple. Existence of the Artificial Intelligence Group further evidences a concentration of AI research at the university, adding to its appeal as a source of data.

Semi-structured interviews allow for discussion beyond a strict discursive procedure and accommodates the interactive choices of participants (Briggs 1986). It is through discussion with the specialists — for example, in the thoughts and predictions for AI that they voice — that information on what constitutes the socio-cultural, ethical and technical appears. Finally, the interviews were approached with the concept of 'engaged listening' and an ethnographic imaginary that allows the researcher to gain ethnographic insight similar to that of participant observation (Forsey 2010). This is evidenced by interview questions focused on the biographies of participants and unanticipated changes in questions throughout each interview, according to each individual's utterances.

2.3 Engaging AI's sociality

Studies of the sociality of science and technology have long been explored in the social sciences and humanities, yet studies focusing on the sociality of AI remain underdeveloped within anthropology. Here, I draw upon the suggestion that distinctions between the social and technical are often fabricated rather than actual (Latour and Woolgar 1986). Further, Johnson and Wetmore assert that technology is not constructed in isolation, but is co-produced with society. Here, it "both embeds and is embedded in social practices, identities, norms, conventions, discourses, instruments and institutions—in short, in all the building blocks of what we term the *social*" (Jasanoff, 2004, 3; Johnson and Wetmore 2008). Addressing such a socio-technological co-production of artificial intelligence exposes how its features are in constant entanglement while simultaneously emphasizing the features and how they hold potential, contingent configurations specific to the AI context, still removed from a narrative of technological determinism.

How does artificial intelligence illustrate socio-technological co-production? Popular media examples consist of goals interacting directly with society, such as building machines with human-like minds or creating androids; but these are only one popularized subset of the socio-technical relationship AI contains. Much of the ordinary, technical decisions within the everyday practices of AI work also articulate with factors categorized as 'social' or 'cultural' and the relation to being human. But in talking with one professor of computer science at the University of Waterloo, such articulation is not always realized:

> "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. You spend a lot of your time thinking about other human beings*"*

This quote reflects on how AI systems are sometimes framed simply as "technical objects", where humans may be excluded from the programming despite potential human-system interactions in application. This is especially noticeable in news of driverless cars and a focus on their efficiency through the reduction of human mistakes with the removal of physical, human direction of the machines. In removing the human variable, it may also act as a separation of social and technical to "establish the universality and effectivity" of the technology (Born 1997, 142). This then aligns with larger, structural and institutional goals in industry and public domains where AI technologies are increasingly implemented for social and economic solutions. Here it is also possible that a notion of technology in isolation partially explains a perceived distance of AI systems from the social. This 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, 127; Thorpe 2008; Liu 2017). It is a view that has been analyzed in more historical contexts, such as in Latour's *We Have Never Been Modern* (1991). For example, in one chapter he discusses Shapin and Schaffer's *Leviathan and the Air-Pump* (1985) with a discussion of topics including political and scientific representations, fact-making and the boundaries of science. And while many of the operational aspects of intelligent machines can be separated from humans in practice when given emphasis on automation, the development or "building" of

these machines is fundamentally entangled in and co-productive with human-machine interaction (Robertson 2010; Richardson 2015). Even in technical categorizations of autonomy — independent agential action separated from the programmer's original input — there are instances in which developers must evaluate and re-adjust the machine's operational capacities (Warwick 2012). This follows assertions that technology is "patterned by the conditions of its creation and use" and technical decisions made at one point in time can impact development made at another and vice versa (Mosemghvdlishvili and Jeroen 2013). Thus, from acquiring datasets to programming algorithms and designing user interfaces, AI produces and is produced by entanglements of the socio-cultural and technical, amongst the many other factors commonly attributed to human sociality.

At the conference on Governance of Emerging Technologies where my observation took place, in a keynote speech on the responsible development of AI, a well-known figure in the field called to both 'maximize human values' and manage risk in AI by accounting for the "biggest deviation of rationality" — our wants. They expressed that by "learning to predict what people want", it will become easier to develop systems that are beneficial. To reach prediction they explained that it would require "cultural work", although the details of this were not specified. It was not made clear who the 'people' referenced would be, although the call for 'cultural work' seemed to suggest a holistic survey of societies globally. Despite this, it was not discussed how people's 'wants' could come into conflict, which instead assumes a shared set of 'wants' across the referenced 'people'. This omits a common reality where wants are often negotiated and in tension between multiple stakeholders, across and within societies, communities, groups and individuals. Similarly, research on the usefulness of psychological and sociological models underpins some approaches in in artificial intelligence. For example, in a subfield of computer

science, Affective Computing, a major theoretical influence comes from Affect Control Theory (ACT). This sociological theory considers the relationship between emotion and 'culture' through the categorization of "socially shared patterns of affective meaning" (Rogers et. al 2014). One of the interviewed professors works with building sociological models into AI solutions through this field of research. They explain that the models rely on "the sort of collective consciousness or collective nature of human intelligence" which in this case, becomes associated with affect and emotion. How this is mapped to culture through AI techniques is exemplified in a program mentioned by the professor 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.

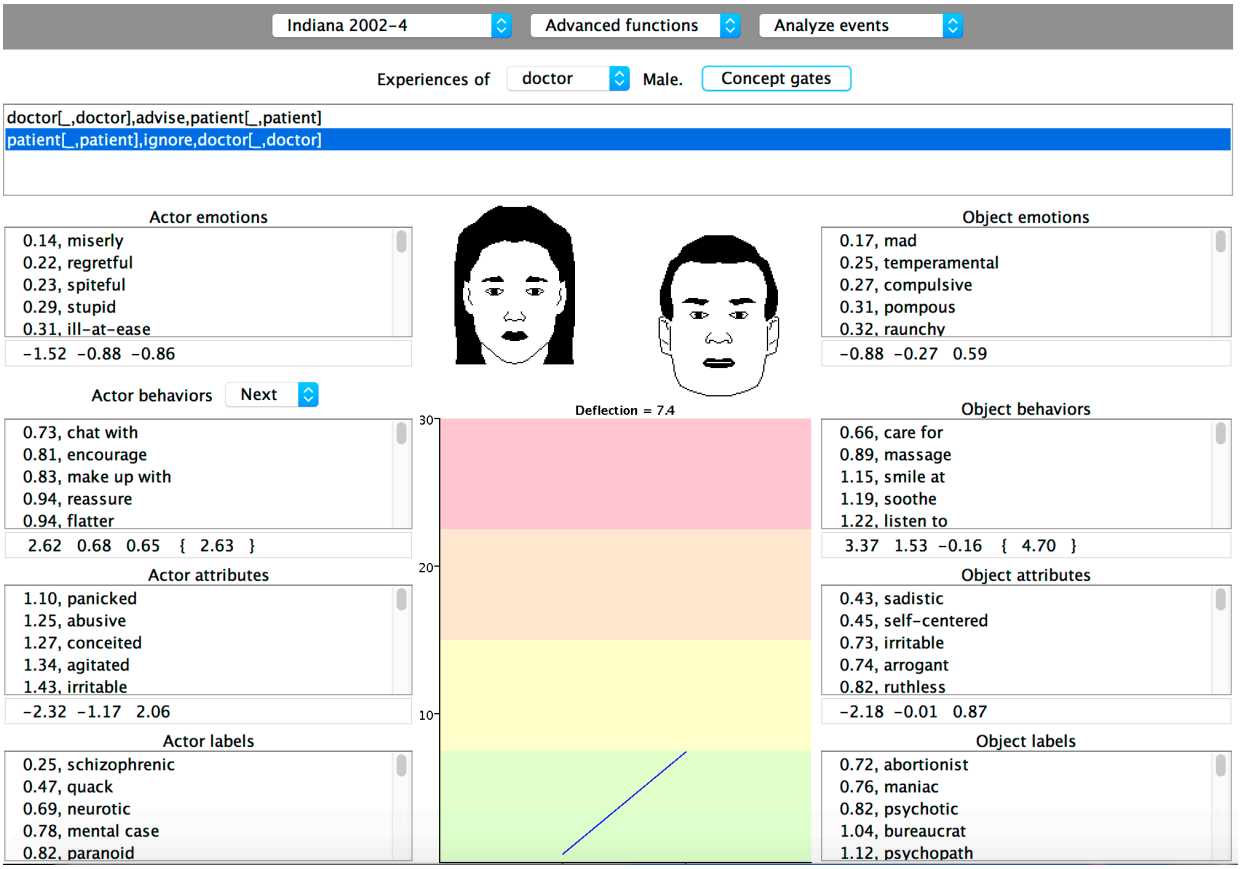

Figure 1 Interact model of interactions[3]

In *Interact*, cultures are depicted as totalities that fall within a normative process of STEM-related description. Philosophies of science often support this positioning that, rather than privileging individualism, emphasizes naturally embodied dispositions substantiated by a group, corroborated as "culture" (Solomon 2008). Anthropologists, however, have problematized the definition of culture as a bounded concept[4]. Emphasizing intragroup variations and movements (e.g. migrations), they argue that cultures are not homogenous entities (Clifford and Marcus 1986; Helmreich 2001). The categorization in *Interact* of place-based identity, behavior and setting meant to determine affect and impression reproduces the definition of culture as bounded and

[3] Interact http://www.indiana.edu/~socpsy/ACT/interact.htm
[4] Definitions of culture have been problematized for many years (Gupta and Ferguson 1997; Hobart 2000)

assumes a universality of emotions. It also places social experience as something that is rigidly patterned, based on its representation as static and deterministic.

For one, 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 its reliance on universality. Again, anthropologists have problematized universality and homogeneity both theoretically and methodologically. In more recent trends, an ontological approach would question the universality applied to social and cultural phenomena through a program like *Interact*. This would suggest that the phenomena are brought into existence through their delineation in the first place, rather than being universally attributed and pre-existing conditions (Coopmans et al. 2014; Hoeppe 2015). In other words, the codifying of 'culture' and emotions in this way is itself an embodied 'cultural' interpretation — a phenomenon brought into the world through the activity of codifying itself.

Along this view, the reifying of culture as a set of specific variables also frames social and cultural elements as technical. As previously discussed, AI is sometimes framed as a set of technical objects, separating the social and technical to consolidate the universality of its systems and technologies. Those transforming intangible, non-technical socio-cultural factors into representations of data produce inscriptions — that is, "visual/textual translations and extensions of scientific practice" (Latour and Woolgar 1986, 142). These inscriptions frame such factors as technical objects and aim to legitimize their representation as 'technical' (Latour and Woolgar 1986). This then has the effect of reifying a universality in these social and technical elements. But as this analysis has shown, instead it results in an essentialist and reductionist approach to the socio-cultural elements and emotions highlighted in the computer simulations of *Interact*.

Again, this is antithetical to current anthropological approaches to “cultural” analysis, but nonetheless, is also extended to other contexts of AI application as explored below.

Managing co-production

Noted above, through programs and development using notions of culture and identity in certain ways, the technologies too, become entangled with certain makings of the socio-cultural. How this co-production affects users has already been framed as yielding both positive and negative results for public contexts. This is related to ways that power and governance emerge through AI’s socio-technological framework. Venn (2007) suggests that current Western governmentality is increasingly defined by bio-political modes of governance that are "mediatized and informationalized" in the attempts to 'improve our lives' through additional surveillance. Here, technologies are viewed as tools to be employed for enhanced efficiency and accuracy in domains such as healthcare, labour and security. A professor of computer science at the University of Waterloo presents an example that shows how this might work in practice. They explain that application of AI techniques for "risk assessment in the justice system" is an "obvious ethical concern" and "when it involves human rights there needs to be oversight". At this point, the ethical becomes apparent for the professor in the context of human rights discourse. They then further explained:

> “Judges are deciding how long to sentence somebody based on their risk, a risk of re-offending. And if this is being done by black box machine learning techniques, where there’s no human in the loop that is held accountable for the recommendation, clearly that’s not right...”

With the increasing emphasis on information and data in the Information Age, artificial intelligence has shifted into a position of increased authority. Here, some scholars note that

technologies hold an authoritative position because they allow for a transcendence of humanity and nature (Turner 2007; Yampolskiy and Fox 2013; Muller 2014). As mentioned above, certain AI-related technologies have proven enhanced surveillance capabilities which have been employed within policing. Through systems using machine learning techniques, the New York Police Department has attempted "predictive policing" which tries to identify "where and when crimes are more likely to happen and who may commit them" (Stanford University 2016; Newcombe 2016). This can be problematized for issues with bias and a concern that when obtaining predictions from policing machines, the socio-economic factors that contribute to perceived crime 'hot-spots' are not taken into consideration, although they could potentially be considered in the system design process by developers. It is an example of how AI systems can reproduce bias despite their representation as 'neutral' and unbiased. This is accompanied by research explaining how incarceration rates for certain minority groups are already statistically higher due to racial profiling and other forms of systemic inequality resulting from residual imperialist thought that places certain bodies under directed surveillance by the state (Miller 2014). These narrow AI techniques begin to "define how we are seen, by providing a digital lens, tailored by statistics and other biases" and populations are represented as data (van Otterlo 2014, 258).

An example of how populations are represented as data comes from an industry context where AI is used to account for forms of bias. One participant, co-founder and CEO of a company in the talent acquisition industry discussed how AI works against unconscious human bias in their services:

> *"*One of the things that [cognitive recruiting assistant AI] doesn't do is review resumes based on name. Name is one of the first lines of defence that we use against that type of bias. Names can be racially imbued, they can be engendered,

> so forth and so forth. Within our system the actual matching algorithms that occur, from personality fit to concept analysis, do not involve name whereas the human recruiter always involves name*"*

Here, the major focus is the reduction of gendered and racialized bias in human decision-making processes. They further explained the unbiased status of the underlying programming of AI techniques employed in this:

> *"*A machine learning algorithm in and of itself is not biased. It doesn't have an ethical disposition one way or another. As the algorithm learns from either the human-trained input or the self-learned input, we have to identify what those outcomes are of the algorithm. We have to be able to identify whether or not it is actually causing the ideal resultant, otherwise it becomes a useless piece of software. What we want within our datasets— that's the piece that can start to cause particular types of bias, potentially in the future. That's where audits are required, internal audits: 'are we representing a broad enough dataset or broad enough database for x type of job position, x sector of job etcetera, etcetera?' So we do take it on ourselves to see, are we getting a good cross-section or representation of those that are applying to jobs?"

Data is highlighted as the likely vessel for bias given its more direct connection to human input and decisions. What data is used, how diverse it is, where it was acquired — these sorts of questions were also mentioned in discussion with the CEO as they apply to algorithmic data bias. In application, the AI is meant to correct for human bias by editing out certain social factors. Working with representations of identity as data, there is a return to the social becoming an object. And while the AI specialist addresses matters of representation and systemic bias critically, they also seem to position bias as a technical problem to be dealt with through a wide held understanding that algorithms are inherently asocial.

Those developing algorithms are influenced by multiple factors that are social — the algorithms too, are already social, as the decisions of human developers. And within their work these specialists craft an understanding of how the technical components articulate with other expertise and forms of knowledge, all of which are value-laden. They must balance factors ranging from technical operations, funding influences, compatibility with user design while ensuring that the algorithms are adapted for other, already existing and emerging technologies and so on (Johnson and Wetmore 2008; Ekbia 2008; Warwick 2012). In other words, when an AI researcher or developer makes a decision it is "not simply a detached technical decision but has ethical and value content and implications" (Johnson and Wetmore 2008, 572). Added to this, recent work suggests that humans prefer to anthropomorphize objects in order to establish their relation to non-human objects. Thus, what we represent as identity and personality is projected onto objects and is more heightened when they "evoke a sense of agency", such as in AI technologies. In other words, the designer or developer may consciously or unconsciously aim to maintain their "social world" in the design of expert systems (Sherr 1995; Eyssel et al. 2012; Picarra et al. 2016).

Finally, extended to the discussion on governance, the concept of negotiability in technology "points to the existence of a surplus of workable solutions or choices for any given technology, both in the process of design and in its utilization", negotiated by different groups with varying levels of social, political and economic power or authority. In AI, this multitude of actors and their decisions also impacts the range of choices that affect development, meaning that "certain solutions may become entrenched in the technical or social infrastructure of existing technologies" (Mosemghvdlishvili and Jansz 2013, 1599). Thus, with the above considered — of co-production, an ontological approach, and the negotiability of choices and action — I suggest

that the social already inheres in practices of AI. In other words, and while it may not be acknowledged, there is no designing "out" the social from algorithms because technical practices are already social too.

2.4 Beneath the Hype

Emerging technologies often evoke a narrative that frames innovation and development as inherently beneficial for humanity (Ekbia 2008; Stanford University 2016). A computer science professor specializing in optimization specifically noted this innovation discourse and referred to it as 'techno-optimism' — a notion that technology "will solve all the problems and it's only a good thing". They described an ethos of techno-optimism at their campus amongst students and colleagues. This was also positioned alongside current sensationalized research and popular narratives. Both of the professors interviewed mentioned that there is a pattern in AI, like other STEM-related disciplines, wherein certain breakthroughs reach a level of visibility that even sparks interest in the general 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. Such 'hypes' are not uncommon for academia regardless of discipline, but with technological development there is also emphasis on how it is an indicator of intellectual and societal 'progress' (Genus 2006).

On the discourse surrounding autonomous driving, one professor referred to the director of Microsoft Research who argues that autonomous driving prevents deaths from drunk driving. This professor provided a counter-argument:

> *"*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"

In this statement, the professor 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. As well, those working in AI recognize that public perceptions of AI hype do not necessarily reflect actual spaces and practices of ordinary, everyday work in AI. In many instances, beneath the hype, the reality of the expected capabilities of AI are much more ordinary. For example, one PhD student explained that the "development as of two years ago" was "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". They also suggested that in many of the industry jobs that University of Waterloo computer science graduates pursue, work is configured as technical tasks instead of societally transformative undertakings:

> "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?'"

This suggests that for those working in AI the focus is not always aimed directly at the sensational aspect of what the work entails. Instead, it is the technical frameworks and solutions to specific problems that are of most interest. For some it is in the specialized forms of AI — formed from these more basic, everyday practices often based on "statistics and reasoning", as one PhD student states — that create the most concern in practice.

A PhD student specializing in computer vision indicated that the repercussions of narrow AI seen with special purpose algorithms have an immediacy that is more worrying than the popular displays of general-purpose systems: "what people will do with the special-purpose algorithms, that's more of a concern from my perspective because it's more immediate and even in some cases already occurring". Here, AI is positioned as a neutral tool that, in its use, becomes

embedded with moral value. Emergent technologies are neither inherently 'good' nor 'bad', but neither are they neutral or value-free. Special-purpose algorithms have already been noted with machine learning and deep learning application, like that of the talent acquisition industry and predictive policing examples earlier discussed. They are also prevalent in commonly used, online search engines and social media applications that have become part of the everyday for many[5]. Thus, rather than the concern being what people will do with these algorithms, it is how these algorithms already impact developers, users and environments and "changes the way people interact with the world", as a professor of computer science explained.

These concerns are additionally compacted with the black box problem in artificial intelligence. This has been framed in the technical sense, as systems with mysterious or unknown internal operations (Hackett et al. 2008). But in science studies, the black box problem can refer to "the way scientific and technical work is made invisible by its own success. When a machine runs efficiently, when a matter of fact is settled, one need focus only on its inputs and outputs and not on its internal complexity" (Latour 1999, 304). For AI, the black box problem has been exacerbated by advancements in those special-purpose systems with machine learning and deep neural networks that involve an amount of data and processing too complex for human capabilities (Ekbia 2008). In other words, researchers and developers aren't always sure what computations occur when a system is operating because it works beyond the understandings and predictions of the programmers themselves. But besides an issue with understanding it also reinforces notions of technology in isolation and perhaps guides researchers into focusing on those technical tasks that created concern in the first place. The black box problem with narrow

[5] Google's use of deep learning https://www.forbes.com/sites/bernardmarr/2017/08/08/the-amazing-ways-how-google-uses-deep-learning-ai/#3346dd023204

AI further leads to questions of ethical and social implications under a frame of uncertainty, explored in the next section.

AI and the ethical

Examining some of the ways that specialists configure the ethical in everyday practices of AI fits alongside insights on the ways that sociality entangles with artificial intelligence. Ethics has been a topic of inquiry in STS with extensive work on bioethics and technoethics in particular. In anthropology, it has been approached through two major tracks — as value models guiding research and as a topic of study exploring the ways that morality is manifested and maintained in the range of everyday experiences, contexts and interactions of individuals and communities (Scheper-Hughes 1995; D'Andrade 1995; Nader 2002; Laidlaw 2002; Lambek 2010; Robbins 2016). This thesis follows the latter. There are various means to consider how ethics enters the everyday. For example, Zigon (2010) uses a phenomenological approach that considers assemblage — articulations of components, be they institutional or public discourses and embodied dispositions — and highlights practice across multiple agential capacities. He emphasizes that morality is not a closed system, but radically context-dependent. Other more critical approaches require both empirical and theoretical investigation that accounts for epistemological processes (Fassin 2008). In any such approach, the emphasis on ordinary activities and decisions in daily experiences remains the central sites for analysis in this thesis (Lambek 2010).

Questions on ethics in relation to artificial intelligence have received increased public attention during the last few years. It has also been an established feature in academic spaces through literature on roboethics and machine ethics (Allen et al. 2006; Moor 2006; Anderson and

Anderson 2007; Anderson 2008; Dougherty et al. 2013; Brundage et al. 2014; Vanderelst and Winfield 2016). In AI scholarly contexts, discussions have tended to privilege certain configurations or models of ethics, mainly those influenced by Western moral philosophy that frame ethics as a "complex form of decision-making" (Wallach et al. 2008; Torrance 2013; Englert et al. 2014; Cervantes et al. 2016).

Studies in educational settings show that AI is a subfield of computers science and that it is a considerably practice-oriented discipline. For example, students learn various coding languages, typically accessed through computer interaction. Thus, students must physically code using a machine in order to understand how the user's input influences the functions of the system (Kay et al. 2000; Helmreich 2001). With primary actions typically facilitated by a computer then, to "AI specialists the central meaning of work may be writing code and building systems" (Forsythe 1993a, 470; Kay et al. 2000; Helmreich 2001). This was similarly noted by a professor of computer science who explained that "at the research level it's just studying algorithms", giving the 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".

Thus, even as AI entangles social and technical elements, there are moments in its everyday practice that distance specialists from the social elements, including those related to ethical features. With the characterization of work in AI assigned to certain structures of discourse, other topics can be sidelined and positioned as play or as a "waste of time" while the technical, algorithmic aspects of AI become the focus (Forsythe 1993ab). This came into discussion with a graduate student specializing in multi-agent systems when talking about AI ethics:

> *"*I think ethical discussion is good, I mean that's definitely my background. I'm more interested in arts and social sciences and junk like that than most people in computer science*"*

The association of ethical discussion with humanistic disciplines — in this example somewhat stigmatized as "junk like that" — distances ethics from AI as something that may be external to it. Placing ethics in a position alongside technical tasks may also mask other considerations and consequences of the technologies at work. When asked about discussion of ethics in their education, the graduate student explained that if ethics was spoken about it was done in an "intentional way" such as in special classes and application examples. This was further confirmed by one of the PhD students who was familiar with the more administrative side of the program which the Master's student is enrolled in. To another PhD student working in machine learning and computer vision, the ethical was discussed through direct reference to Western philosophical theories of ethics important to AI and computer science. They described the relation ethics has to theoretical computer science and problem-solving methods used for algorithms depending on the class of complexity of a problem.

For some it is a question of understanding. The university professor working with sociological models explained:

> *"*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*"*

For others, like a PhD student studying multi-agent systems, there was more concern about their competency in discussing the ethical:

> "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"

In the above examples from AI specialists, ethics is discussed according to some form of formalized model of thought — either as philosophical theories and problem-solving methods, or as a qualified background in ethics. There appears to be a designation of authority on who may discuss ethics and how it should be done. In this way ethics may be framed as a technical problem that unintentionally positions it as separate from the scientific practice. While this returns to the idea of technology in isolation, it is only one making of ethics for some AI specialists. In another context beyond the defined, categorized 'ethics', engagement with the ethical as a broader condition was an apparent point of analysis. The professor of computer science specializing in constraint programming explained that ethical discussion is considered more of a "challenge outside of the curriculum" as "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*"*.

This participant continued with their explanation, speaking about issues of economic inequality in North America where individuals employed through fields such as computer science are part of a "select group that are all the wealth". This was reiterated in the common mention of technological unemployment by participants and other major issues associated with AI in the popular discourses, but with 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 one that is ethical. On the technical side, there is a problem of implementing ethical principles within systems affected by "reasoning under uncertainty" that, as one PhD student explained "is and has always been a key challenge in artificial intelligence". This concern is also popular in the machine ethics literature and testing of ethical implementation has been discussed, but remains more speculative than evidential at this point in time (Anderson and Anderson 2007). The other challenge of uncertainty — as something accompanying the ethical dimensions of emerging technologies — becomes normalized through these specialists' approaches to the social implications of AI, as previously explored (Akama et al. 2015). By conceptualizing ethics and ethical challenges as things that compose an independent field of expertise like that of artificial intelligence, along with the uncertainty associated with both AI's implications and a defined ethics, the two become simultaneously comparable and appear to be disentangled. As a result, when an AI specialist explains their position working in AI and actively states their detachment from the knowledge that an ethicist has, despite being implicated in ethical items nonetheless, it often goes unchallenged.

2.5 Discussion: Expanding Regulation and Expertise

Currently there are no official standards of ethical practice that guide the development and application of artificial intelligence. In speaking with my interlocutors and from my observations at the conference on emerging technologies, there are more informal forms of best practice and documents employed by researchers and developers through their local affiliations, but anything encompassing AI in general has met resistance due to difficulties addressing its varied application across multiple domains. Here, themes of moderation and accountability emerged in the responses by interlocutors. The industry professional for example, preferred the term "moderation" rather than regulation and explained that it would be better not to stop the "trajectory of technology" this way. This followed the previously described techno-optimism and reinforces the idea of technology in isolation in that it suggests that technology has its own trajectory untouched by other factors, including human influence. In this notion of moderation, there was emphasis on a need for collaboration between multiple stakeholders and actionable insights to come from this.

For others like the PhD student in computer vision and machine learning, there could be regulation, and it would imply collaboration too:

> *"*I think the primary groups that you would need are the politicians and legal scholars, application area experts, people who are experts in the problem that the AI system is being applied to, and experts in the development of artificial intelligent systems. I think good reasonable laws that allow AI systems to serve the public interest and move society forward, I think they could be created, but I think it would have to involved a group like that to design them where all of the parties take all of the others seriously*"*

The PhD student further explained that "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" and at that those other parties involved will have to listen to computer scientists about what can occur, to make sure that it is computationally feasible in the first place. How these laws will balance social and ethical requirements with technical abilities suggests that there is a perspective of compromise and imperfection, which may also stem from conditions of uncertainty. While this is a concern to seriously consider in moderating AI technologies, this also leaves a lack of inclusion of other kinds of expertise from groups that act outside of traditional legal and economic backgrounds.

The considerable underrepresentation of women and people of colour in STEM fields is well recognized (Morganson et al. 2010; Good et al. 2012; Fontana et al. 2013). This is also evident in initiatives such as the IEEE's *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*, which focuses on "empowering technologists to prioritize ethical considerations in the creation of Artificial Intelligence and Autonomous Systems". The IEEE describes the document as representing "the collective input of over one hundred global thought leaders in the fields of Artificial Intelligence, law and ethics, philosophy, and policy from the realms of academia, science, and the government and corporate sectors" (IEEE 2016). But in the document itself there is little reference to "thought leaders" in anthropology, STS, women's studies, sociology or other areas of thought that have thoroughly evidenced, established, valuable input about technology and society that 'technologists' can use. It is also here that the underrepresentation of women and people of colour re-appears. If moderation is to occur amongst an assemblage of stakeholders, it is

necessary to acknowledge that the stakes are high for everyone and that there are certain voices and areas of inquiry that are neglected by other epistemological traditions.

Here, inclusion can also act as ethical practice, just as collaboration may be a method of regulation. As suggested, even in those initiatives claiming diversity, certain peoples and methods of inquiry are ignored. Those claiming diversity cannot distance themselves from the underlying structures and discourses that background not only the initiatives, but their own histories, identities, practices, and so on. It is these things that may allow for a lack of inclusion, as noted earlier. Influencing this understanding, Jenny Reardon's work on the Human Genome Diversity Project uses an analysis of coproduction to address issues of "equity, participation and subjects' rights" and comments on inclusion (2001, 371). She attributes the failure of the HGDP, in part, to a lack of appreciation for the full scope of ethical considerations. Extended to *Ethically Aligned Design*, the expertise and backgrounds noted for collaboration become the principle communities and standard contexts in the document's discussion of ethical design in AI. The emphasis on STEM, government and industry caters to certain questions and reinforces certain definitions of ethics, methods of ethical action, as well as what the major social implications of AI are. For example, the concern about technological unemployment mentioned in previous sections is mainly positioned with economic issues. This misses an abundance of other factors implicated in technological unemployment that are often addressed and represented by voices not within STEM, for instance. Including additional collaborators and perspectives is necessary to ask more inclusive questions that are cognizant of experiences and discourses which prove that social, ethical, and scientific/technical issues are "inextricably interconnected and come into being together" (Reardon, 2001, 381). Thus, while *Ethically Aligned Design* and similar initiatives such as the *Partnership on AI* aim to build an ethical element into development from

the beginning, to allow for more socially responsible AI, there is a need for increased representation and inclusion.

Finally, briefly mentioning collaboration as regulation — considering the previously mentioned concept of negotiability in technology, if certain expertise is considered and others are not, regulation or 'moderation' can become black-boxed too (Mosemghvdlishvili and Jansz 2013). Other expertise should also be accompanied by another important factor — public input. There is a need for integrated publics in collaborative efforts of regulation as there is a gap in both localized and structural engagement with the public in these initiatives. This is no fault of computer scientists and is often due to institutional restrictions and lack of encouragement for knowledge dissemination accompanied by a public image of AI that is more fantastical than realistic. Importantly, representation and interaction in collaborative efforts such as *Ethically Aligned Design* could promote education in AI and foster public trust, but it must not act as a placeholder for ethical action as it has in other domains, but one of many techniques in the emergent environment of AI (Hayden 2007). For these reasons, collaboration beyond the norm and dominant epistemological practices is most beneficial. How this works in action is enough discussion for separate additional study, but the present suggestion of a more critical view in regulation acts as an initial intervention.

2.6 Conclusion

While much of the public media is concerned with the thrills of artificial intelligence, this work brings to light some of the more ordinary discourses and practices around artificial intelligence. First, I investigate how socio-cultural phenomena are entangled in AI through co-production. This refutes notions of technology in isolation that divorces technical results from social influences. Exploring makings of the social-cultural in examples such as bias in AI design and application, I highlight how technical decisions are entangled with sociality. This further confirms how technology is not neutral as the personal beliefs and values of designers, developers and researchers quite often enter the systems they create (Sherr 1995). Next, as additional insight into the socio-technological co-production of artificial intelligence, I uncover some of the ways that popular ideologies including techno-optimism and innovation interact with the entanglements of social, cultural, and technical in AI. In this discussion I also note how the ethical is negotiated through normalized uncertainty and the ways that it is formed in special-purpose systems, beneath the hype. Finally, I provide discussion on the regulation of AI and the need for increased pubic integration and inclusive expertise. This is meant to encourage a more critical approach to AI's capabilities and potential implications so that it is more beneficial to the heterogeneity of publics that will interact with the emerging technology.

This thesis adds some perspective to the ways that artificial intelligence entangles social, cultural, ethical and technical factors. Additional study is surely required to go beyond my brief focus on a small group of AI specialists, to the great variety of communities, discourses and processes involved in the various emerging contexts of artificial intelligence. Other scholarly endeavours could include ethnographies on applied AI and public knowledge, feminist STS analysis of AI systems in healthcare, or case-studies in globalizing artificial intelligence. An

anthropological study of machine ethics — the field focusing on ethical embodiment by intelligent machines — would also be of interest (Anderson and Anderson 2007). In order to reach the goal of technological systems that enact globalizing diversity and reduce prevalent ideological and social biases, understanding that technology is co-productive with society and how we reach this understanding is important. It is in all of our best interests to guarantee non-harmful outcomes from these technologies.

References

Akama, Yoko, Pink, Sarah, and Annie Fergusson. 2015. “Design + Ethnography + Futures: Surrendering in Uncertainty”. doi.10.1145/2702613.2732499.

Allen, Colin, Wallach, Wendell, and Iva Smit. 2006. "Why machine ethics?" *IEEE Intelligent Systems,* 21:12-17.

Anderson, Michael, and Leigh Anderson. 2007. "Machine Ethics: Creating an Ethical Intelligent Agent" *AI Magazine,* 28:15-26.

Anderson, Susan Leigh. 2008. "Asimov’s “Three laws of robotics” and machine metaethics." *AI & Society,* 22:477-493.

Asimov, Isaac. 1950. *I, Robot*. Greenwich, CT: Fawcett Publications.

Bernard, Russell H. 2011. *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. Lanham, MD: Altamira Press.

Bille, Mikkel, and Tim Flohr Sørensen. 2007. "An anthropology of luminosity the agency of light." *Journal of Material Culture,* 12:263-284.

Boellstorff, Tom. 2012. *Ethnography and virtual worlds: A handbook of method*. Princeton, NJ: Princeton University Press.

Born, Georgina, Banks, Marcus, and Howard Morphy. 1997. "Computer software as a medium: Textuality, orality and sociality in an artificial intelligence research culture." In *Rethinking visual anthropology.* New Haven, CT: Yale University Press.

Bostrom, Nick, and Eliezer Yudkowsky. 2014. "The ethics of artificial intelligence." In *The Cambridge handbook of artificial intelligence*. Cambridge, MA: Cambridge University Press.

Briggs, Charles L. 1986. *Learning How to Ask. A Sociolinguistic Appraisal of the Role of the Interview in Social Science Research*. Cambridge, MA: Cambridge University Press.

Brundage, Miles. 2014. "Limitations and risks of machine ethics." *Journal of Experimental & Theoretical Artificial Intelligence,* 26:355-372.

Cellan-Jones, Rory. 2014. “Stephen Hawking warns artificial intelligence could end mankind” *BBC*, December 2. http://www.bbc.com/news/technology-30290540.

Cervantes, Jose-Antonio, Rodriguez, Luis-Felipe, Lopez, Sonia, Ramos, Felix, and Francisco Robles. 2016. "Autonomous Agents and Ethical Decision-Making." *Cognitive Computing,* 8:278-296.

Cerwonka, Allaine, and Liisa H. Malkki 2008. *Improvising theory: Process and temporality in ethnographic fieldwork*. Chicago, IL: University of Chicago Press.

Clifford, James, and George E. Marcus. 1986. *Writing culture: The poetics and politics of ethnography*. Berkeley, CA: University of California Press.

Combi, Mariella. 1992. "The imaginary, the computer, artificial intelligence: A cultural anthropological approach." *AI & society,* 6:41-49.

Coopmans, Catelijne, Vertesi, Janet, Lynch, Michael E. and Steve Woolgar. 2014. *Representation in scientific practice revisited*. Cambridge, MA: The MIT Press.

D’Andrade, Roy. 1995. “Moral Models in Anthropology.” *Current Anthropology* 36:399-408.

Dougherty, Mark. 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,* 22:1-7.

Douglas, Heather. 2007. “Rejecting the Ideal of Value-Free Science.” In *Value-Free Science? Ideals and Illusions.* Oxford: Oxford University Press.

Edensor, Tim. 2011. “Entangled agencies, material networks and repair in a building assemblage: the mutable stone of St Ann’s Church, Manchester.” *Transactions of the Institute of British Geographers,* 36:238–252.

Ekbia, Hamid R. 2008. *Artificial Dreams.* Cambridge: Cambridge University Press.

Englert, Matthias, Siebert, Sandra, and Martin Ziegler. 2014. "Logical limitations to machine ethics with consequences to lethal autonomous weapons." *arXiv preprint 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,* 851-857.

Fassin, Didier. 2008. "Beyond Good and Evil? Questioning the anthropological discomfort with morals." *Anthropological Theory,* 8:333-344.

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

Forsey, Martin G. 2010. "Ethnography as participant listening". *Ethnography,* 11:558-572.

Forsythe, Diana E. 1993a. "Engineering knowledge: The construction of knowledge in artificial intelligence." *Social studies of science,* 23:445-477.

Forsythe, Diana E. 1993b. "The construction of work in artificial intelligence." *Science, Technology & Human Values,* 18:460-479.

Franklin, Sarah, and Celia Roberts. 2006. "Studying PGD." In *Born and Made: An ethnography of preimplantation genetic diagnosis.* Princeton, NJ: Princeton University Press.

Genus, Audley. 2006. "Rethinking constructive technology assessment as democratic, reflective, discourse." *Technological Forecasting and Social Change,* 73:13-26.

Good, Catherine, Rattan, Aneeta, and Carol S. Dweck. 2012. Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology,* 102:700-717. doi:10.1037/a0026659

Gupta, Akhil, and James Ferguson. 1997. *Culture, power, place: Explorations in critical anthropology*. Durham, NC: Duke University press.

Hackett, Edward J., Amsterdamska, Olga, Lynch, Michael, and Judy Wajcman. 2008. *The handbook of science and technology studies*. Cambridge, MA: The MIT Press.

Hacking, Ian. 1999. *The social construction of what?*. Cambridge, MA: Harvard University Press.

Hayden, Cori. 2007. "Taking as giving: Bioscience, exchange, and the politics of benefit-sharing." *Social Studies of Science,* 37:729-758.

Helmreich, Stefan. 2001. After culture: reflections on the apparition of anthropology in artificial life, a science of simulation. *Cultural Anthropology,* 16:612-627.

Hobart, Mark. 2000. *After culture: Anthropology as radical metaphysical critique*. Duta Wacana University Press.

Hoeppe, Götz. 2015. Representing Representation. *Science, Technology, & Human Values,* 40:1077-1092.

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

Ingold, Tim. 2009. "When ANT meets SPIDER: Social Theory for Anthropods." In *Material Agency: Towards a Non-Anthropocentric Approach*, edited by Carl Knappett and Lambros Malafouris, 209-215. New York: Springer.

Irwin, Alan. 2008. "STS Perspectives on Scientific Governance." In *The Handbook of Science and Technology Studies*. Cambridge, MA: The MIT Press.

Jasanoff, Sheila. 2004. *States of knowledge: the co-production of science and the social order.* New York, NY: Routledge.

Jasanoff, Sheila. 2016. *The Ethics of Invention: Technology and the Human Future*. New York: WW Norton & Company.

Johnson and Wetmore. 2008. "STS and Ethics: Implications for Engineering Ethics" In *The Handbook of Science and Technology Studies*. Cambridge, MA: The MIT Press.

Kay, Judy, Barg, Michael, Fekete, Alan, Greening, Tony, Hollands, Owen, Kingston, Jeffrey H., and Kate Crawford. 2000. Problem-based learning for foundation computer science courses. *Computer Science Education,* 10:109-128.

Laidlaw, James. 2002. "For an anthropology of ethics and freedom." *Journal of the Royal Anthropological Institute,* 8:311-332.

Lambek, Michael. 2010. *Ordinary ethics: anthropology, language, and action.* New York: Fordham University Press.

Latour, Bruno and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts.* Princeton, NJ: Princeton University Press.

Latour, Bruno. 1991. *We have never been modern*. Cambridge, MA: Harvard University Press.

Latour, Bruno. 1999. *Pandora's hope: essays on the reality of science studies*. Cambridge, MA: Harvard University Press.

Latour, Bruno. 2000. "The Berlin key or how to do words with things." In *Matter, materiality and modern culture*. New York, NY: Routledge.

Liu, Jennifer A. 2017. "Situated stem cell ethics: beyond good and bad." *Regenerative Medicine,* 12:587-591.

Markoff, John. 2015. "Planes without Pilots" *The New York Times*, April 6. https://www.nytimes.com/2015/04/07/science/planes-without-pilots.html

Marr, Bernard. 2017. "The Amazing Ways Google Uses Deep Learning AI" *Forbes,* August 8. https://www.forbes.com/sites/bernardmarr/2017/08/08/the-amazing-ways-how-google-uses-deep-learning-ai/#3346dd023204.

Miller, Reuben J. 2014. "Devolving the carceral state: Race, prisoner reentry, and the micro-politics of urban poverty management." *Punishment & Society,* 16:305-335.

Moor, James. 2006. "The Nature, Importance, and Difficulty of Machine Ethics." *IEEE Intelligent Systems,* 18-21.

Morganson, Valerie J., Jones, Meghan P., and Debra A. Major. 2010. Understanding women's underrepresentation in science, technology, engineering, and mathematics: The role of social coping. *The Career Development Quarterly,* 59:169-179. doi:10.1002/j.2161-0045.2010.tb00060.x

Mosemghvdlishvili, Lela, and Jeroen Jansz. 2013. "Negotiability of technology and its limitations: The politics of App development." *Information, Communication & Society,* 16:1596-1618.

Muehlhauser, Luke, and Anna Salamon. 2012. "Intelligence Explosion: Evidence and Import." In *Singularity Hypotheses: A Scienti c and Philosophical Assessment*, edited by Amnon Eden, Johnny Soraker, James H. Moor, and Eric Steinhart. Berlin: Springer.

Muller, Vincent C. 2014. "Risks of general artificial intelligence." *Journal of Experimental & Theoretical Artificial Intelligence,* 3:297-301.

Nader, Laura. 2002. "Breaking the silence-politics and professional autonomy." *Anthropological quarterly,* 75:160-169.

Nardi, Bonnie 2010. *My life as a night elf priest: An anthropological account of World of Warcraft*. Ann Arbor, MI: University of Michigan Press.

O'Hear, Steve. 2017. "Ada is an AI-powered doctor app and telemedicine service" *Techcrunch,* April 19. https://techcrunch.com/2017/04/19/ada-health/.

Olsen, Bjørnar. 2012. "Symmetrical Archaeology." In *Archaeological Theory Today*, edited by Ian Hodder, 208-228. Cambridge: Polity Press.

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,* 1-13.

Reardon, Jenny. 2001. "The human genome diversity project: a case study in coproduction." *Social studies of science,* 31:357-388.

Richardson, Kathleen. 2015. *An anthropology of robots and AI: annihilation anxiety and machines*. New York, NY: Routledge.

Robbins, Joel. 2016. "What is the matter with transcendence? On the place of religion in the new anthropology of ethics*." *Journal of the Royal Anthropological Institute,* 22:767-808.

Robertson, Jennifer. 2007. Robo sapiens japanicus: Humanoid robots and the posthuman family. *Critical Asian Studies,* 39:369-398.

Robertson, Jennifer 2010. Gendering humanoid robots: robo-sexism in Japan. *Body & Society,* 16:1-36.

Rogers, Kimberly B., Schröder, Tobias and Christian von Scheve. 2014. "Dissecting the Sociality of Emotion: A Multilevel Approach." *Emotion Review*, 124-133.

Scheper-Hughes, Nancy. 1995. "The Primacy of the Ethical: Propositions for a Militant Anthropology." And "Responses." *Current Anthropology,* 36:409-440.

Shankland, Stephen. 2017. "How Apple uses AI to make Siri sound more human" *CNET*, August 23. https://www.cnet.com/news/apple-ai-machine-learning-makes-siri-sound-human-on-ios-11/.

Shapin, Steven, Schaffer, Simon, and Thomas Hobbes. 1985. *Leviathan and the air-pump*. Princeton, NJ: Princeton University Press.

Sherr, Leslie. 1995. Is technology neutral? *Print,* 49:154.

Solomon, Miriam. 2008. "STS and Social Epistemology of Science". In *The Handbook of Science and Technology Studies*. Cambridge, MA: The MIT Press.

Stanford University. 2016. "One Hundred Year Study on Articial Intelligence (AI100)." *https://ai100.stanford.edu*.

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. 2016. "Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 1." *IEEE*.

Thorpe, Charles. 2008. "Political Theory in Science and Technology Studies." In *The handbook of science and technology studies*. Cambridge, MA: The MIT Press.

Torrance, Steve. 2013. "Artificial agents and the expanding ethical circle." *AI & society,* 28:399-414.

Turner, Bryan S. 2007. "Culture, technologies and bodies: the technological Utopia of living forever." *The Editorial Board of the Sociological Review,* 19-36.

Vanderelst, Dieter, and Alan Winfield. 2016. "The Dark Side of Ethical Robots." *arXiv preprint arXiv:1606.02583.*

van Otterlo, Martijn. 2014. "Automated Experimentation in Walden 3.0: The Next step in Profiling, Predicting, Control and Surveillance." *Surveillance & Society,* 12:255-272.

Venn, Couze. 2007. "Cultural Theory, Biopolitics, and the Question of Power." *Theory, Culture & Society,* 24:111-124.

Wallach, Wendell, Allen, Colin, and Iva Smit. 2008. "Machine morality: bottom-up and top-down approaches for modelling human moral faculties." *AI & Society,* 22:565-582.

Warwick, Kevin. 2012. *Artificial Intelligence the basics*. New York, NY: Routledge.

Yampolskiy, Roman and Joshua Fox. 2013. “Safety Engineering for Artificial General Intelligence” *Topoi,* 32:217-226.

Zigon, Jarrett. 2010. “Moral and ethical assemblages: A response to Fassin and Stoczkowski.” *Anthropological Theory,* 10:3-15.