The I in AI

Artificial Intelligence as a Way of Seeing, Humanist Framework, and Metadiscourse

Andrew Berger

Department of English

McGill University, Montreal

April 2021

A thesis submitted to McGill University in partial fulfillments of the requirements of the degree of Master of Arts

Table of Contents

Abstract

In this thesis, I argue that Artificial Intelligence is at its core a way of seeing that demands a humanist approach. My analysis is grounded in the recent technological breakthroughs that have operationalized AI at scale, situating this work within an emerging field of humanist-AI research. In Chapter One, I analyze AI discourse in journalism using the tools afforded by AI (including topic modelling and word vectors), which I employ to both create a metadiscursive space and show how generating meaning from the patterns AI detects still requires human analysis. I find in the data that even the mostly surface-level AI discourse of journalism embeds questions that are humanist by nature, and the first section ends by illuminating the language through which humanness emerges across the political spectrum.

In Chapter Two, I change the lens of analysis to science fiction and contrast the arguments made about AI by scholars in the past with how AI is portrayed in sci-fi today. My analysis reveals that these contemporary texts are—like all current AI discourse—fundamentally about what it means to be human. I draw on theories of surveillance and social control to situate my textual analysis within our current digital climate, focusing on how AI affords a new theoretical space that collapses the distance between the theory-dense world of humanities research and the mundane, everyday experience of living with AI.

In Chapter Three, I analyze the implications of AI on human creativity more broadly, and I suggest that the tools of humanist inquiry are essential for building a world increasingly shaped by AI. I structure my analysis around data and describe how AI provides a new level of mediation to understand the human condition, and I describe the value of this new rhetorical space in relation to bias and automation. I then suggest that AI's societal impact may be most clear in how it is leading to a novel model of the self, the profile-self, which I provide evidence

for from a variety of journalistic sources, narrative accounts, and personal experiences. I conclude that from the most distanced perspective, this AI-led global transformation is a process of compression, automation, and datafication that threatens our most innate cognitive frameworks and may lead us to a world void of the depth that defines the human condition.

Résumé

Dans cette thèse, je soutiens que l'intelligence artificielle (IA) est essentiellement une façon de voir qui exige une approche humaniste. Mon analyse est fondée sur les récentes percées technologiques qui ont permis d'opérationnaliser l'IA à grande échelle, situant ce travail dans un domaine émergent de la recherche humaniste-IA.

Dans le premier chapitre, j'analyse le discours sur l'IA dans le journalisme à l'aide des outils offerts par l'IA, que j'utilise pour créer un espace métadiscoursif et aussi pour dépeindre les limites inhérentes de l'IA en tant que façon de voir.

Je trouve que même les discours journalistiques les plus superficiels sur l'IA intègrent des questions qui sont fondamentalement humanistes, et la première section se termine en éclairant le langage à travers lequel une certaine humanité émerge sur le spectre politique.

Le deuxième chapitre analyse à travers l'optique de la science-fiction, et je mets en contraste les arguments avancés par les universitaires sur l'IA avec la façon dont l'IA est dépeinte dans la science-fiction d'aujourd'hui pour montrer que ces textes contemporains — comme tous les discours actuels sur l'IA — traitent fondamentalement de ce que cela signifie d'être humain.

Je m'appuie sur les théories de la surveillance et du contrôle social pour situer mon analyse textuelle dans notre climat digital actuel, en me concentrant sur la façon dont l'IA offre un nouvel espace théorique qui diminue la distance entre le monde dense en théories de la recherche en sciences humaines et l'expérience banale de la vie quotidienne avec l'IA.

Dans le troisième chapitre, j'analyse plus généralement les implications de l'IA sur la créativité humaine, et je suggère que les outils de la recherche humaniste sont essentiels pour construire un monde de plus en plus façonné par l'IA. Je termine en reliant la rhétorique

technique au sein du discours sur l'IA aux théories les plus fondamentales de la psychologie et de la philosophie contemporaines pour révéler comment l'IA est déjà en train de refaire l'humanité.

Je conclus que voir le monde à travers l'IA offre un cadre nouveau pour articuler ce que signifie être humain, et c'est seulement en analysant les possibilités offertes par le discours sur l'IA que nous pouvons commencer à sillonner les implications existentielles de cette technologie opaque.

Acknowledgments

Writing this thesis has been a thoroughly pedagogical experience, particularly in how it has given me perspective on what matters most in life, which is people. We are all the product of the people we love, and this thesis—as an extension of me—is no different.

To start then, I would like to thank my family, whose unconditional love has taught me more about life than any lesson. Thank you for your patience, kindness, and support throughout my life. Thank you, too, for instilling in me a curiosity for the world and the belief that I can shape it.

Thank you to my supervisor, Ned Schantz, whose conversations and classes have played a significant role in framing how I see the world. Thank you, too, for helping me find my moorings in graduate school and for your guidance throughout this process.

Thank you to Richard So, who first introduced me to coding through text analytics. Your influence on my life's trajectory is already clear, and I look forward to working together again in the future.

Thank you to my friends, whose support system I can always rely on when my batteries are running low. In particular, thank you to my roommates, whose kindness and conversations provided the power source I have relied on all year, and whose friendship always nudges me in the right direction.

Finally, thank you to whoever is reading this. Writing, at its best, is like a good conversation, so thank you for taking the time to connect with me.

Introduction

How to See AI

> Seeing comes before words. The child looks and recognizes before it can speak.
>
> But there is also another sense in which seeing comes before words. It is seeing which establishes our place in the surrounding world; we explain that world with words, but words can never undo the fact that we are surrounded by it. The relation between what we see and what we know is never settled.
>
> – John Berger, *Ways of Seeing* (2)

This thesis is about Artificial Intelligence (AI) in the most expansive way,[1] for embedded within AI is a conceptual framework as universal as its underlying binary code, but whose opaque internal mechanisms and power structures obfuscate the rich rhetorical space at its core. To give shape to this mode of inquiry, I turn to John Berger, who suggests that seeing precedes language as a system for understanding the relationships that comprise the world. This thesis asserts that to understand AI—to grasp its many implications and dive into its depths—one must see the world through its perspective. AI's process of seeing is unlike ours because it is entirely electrical, broken into binaries and switches, which makes it universal and scalable. It is a process of compression, classification, modeling, prediction, and generation, with each step transforming the data as is it transmitted. AI's world is comprised of versatile data in theoretical spaces with innumerable dimensions, and this almost inconceivable complexity may explain why so often AI is associated with notions of omniscience and objectivity. To view AI as a way of seeing, however, is to appreciate that the patterns of relations it finds in our world are meaningless without human interpretation.

[1] I will make use of the AI acronym throughout this project for the sake of cohesion, but I do not mean AI as opposed to AGI (Artificial General Intelligence) or ABI (Artificial Biological Intelligence); instead, I use AI in the colloquial sense, as a reference to the broad field of Artificial Intelligence.

My argument, then, may be more accurately thought of as an exploration: a journey through the relations that AI quantifies, around the discourse it inspires, and into the core of its conceptual affordances. It is through this metadiscursive space that I arrive at the theoretical thrust of my thesis, which is that all discourse about AI is fundamentally about what it means to be human. The *I* in AI thus refers to how AI is a way of seeing humans, both theoretically and pragmatically, through discourse and data.

While this may appear to be quite distinct from how AI is generally understood, I ask of my reader only patience and an open mind, for this thesis aims to take a very human approach to pedagogy. What does a human approach entail? Well, that is an idea that I construct throughout this thesis, for, unlike AI, humans learn through the meaning in language. Meaning emerges in discourse, and through conversation we can elevate ideas beyond the mere sum of their parts. Within this assertion is the notion that to be human is to be unlike AI, but this may be most clear if I contextualize my work within the recent technological developments that have led to the recent AI revolution.

Artificial Intelligence as a term was born in 1956 at an academic conference at Dartmouth as a result of the work done by Alan Turing, who postulated that since all computers are constructed from binary, they can function as universal data processing devices. AI broadly referred to the field of research that grew out of the notion that a machine may then be able to mimic the mimic human mind, for the mind is also a data processing device. The technologies whom the term spoke into existence did not come close to living up to their name for decades, however, as the complexity of programming human tasks proved significantly more difficult in practice than in the theory. AI research then underwent decades of budget cuts in what is described as the “AI winter,” for most of the progress in the broader field of computation was

elsewhere. By the late 20th century, the infrastructure of the internet was rapidly expanding, yet there was still virtually no evidence that a machine could ever be intelligent in human ways. During this time, AI merely existed as a theory, and it only permeated public discourse through the science fiction films that engaged with it at a human level—through narrative.

Around 15 years ago, that all changed. One idea transformed AI from a concept into technology: deep learning. Deep learning involves creating AI through an artificial neural network that teaches itself by discovering patterns in data that are mostly illegible to human eyes. This breakthrough has led to AI redefining social media, advertising, and many other academic and financial spheres—and this is just the beginning. In the past few years, AI funding and progress has exploded (Statt), and the rapid pace of its progress suggests that AI may fundamentally change society before the public has time to realize how significant of a change this will be.

AI technology has quickly proliferated across industries, paralleling its rise to prominence in public discourse. Cade Metz in *Genius Makers* (2021) describes AI as "the buzz term of the decade, repeated ad infinitum across press releases, websites, blogs, and news stories" (140). The discourse around AI that is still emerging from this buzz is as dynamic as the technology, for it has become both a scapegoat for a wide range of social problems and a promise for unlimited life. These two poles of sentiment around AI bookend a world of complexity that underlies how most people feel about AI, for AI is inescapable in everyday life; to exist now is to live with AI. The contemporaneity of AI as both an idea and technology also imbue its discourse with the power to shape the underlying algorithms because seeing—as AI does—requires looking, and AI cannot yet look alone.

All the buzz around AI has made it inseparable from much of contemporary culture today. Internet culture in particular has become ripe with discourse about AI on both news and social platforms. Even discourse as short as a tweet about AI involves a litany of assumptions, projections, and frameworks—such as one of my favorite viral tweets by Bilal Farooqui: "Our D.C. office building got a security robot. It drowned itself. We were promised flying cars, instead we got suicidal robots." The connection between the internet and AI discourse is of course no coincidence. Cyberspace, which I use to refer to the expansive world of connections created on the internet, is AI's domain, for in the past decade AI has become the arbiter of our subjective experience online. This is a result of how digital social platforms use AI to create more addictive user experiences, but I will explain this in more depth later in the project.

The significance of cyberspace being mediated by AI is clearer in light of the recent pandemic, which has sped up the pace of our global digital transition. I, like many people, now spend the majority of my days connecting with people, ideas, and culture in cyberspace. Humans are fundamentally temporal beings; or, in John Durham Peters' words, "time marks the limit all our material shapings" (312), so it is certainly worth scrutinizing how our time is being extracted, classified, and compressed by AI. For example, this week I have averaged around 10 hours a day of screen time, which includes my computer and phone. While this is by no means a scientific metric, I do believe it is pretty representative of how much time many people spend on screens in a day. Of those 10 hours, about half were spent using apps that rely on AI to create the user-experience; so, if I'm awake for about 16 hours a day, then almost one third of my waking hours are spent directly in contact with AI. Has there ever been another technology that went from irrelevant to omnipresent in such a short time? If our screen time is increasing, what does

this imply for the future? How will our experience of the world change when the majority of our waking hours are spent interacting with AI-mediated platforms?

My thesis thus begins with the questions that emerge from this understanding that a significant societal transformation is currently underway. As Max Tegmark puts it, "the questions raised by the success of AI aren't merely intellectually fascinating; they're also morally crucial, because our choices can potentially affect the entire future of life" (36). Central to my inquiry, however, is that "our choices" are not yet conscious decisions made by an informed public. They are choices made by AI engineers, who are currently shaping the future without much input from the people who are affected by these technologies. This notion of a collective, of "our," brings us back to discourse, for humans develop our ideas socially, through conversations. The difficulty with AI discourse is that AI operates through a logic that appears esoteric, which obscures the simplicity of the questions that arise out of AI discourse; namely, what are the risks of this revolutionary technology? And what does AI tell us about being human? Embedded in these questions is an assertion that I develop throughout this thesis: that AI is, itself, a humanist discourse.

The three chapters of my thesis, therefore, aim to approach AI culture from three adjacent angles to paint a more vivid picture of what we really talk about when we talk about AI. I begin by looking at AI through journalism, using text analytics to situate the more theoretical second and third chapters within a discourse that has only begun seeping into humanities scholarship. My analysis in Chapter 1 uses the data from more than 3000 articles about AI from three different media niches: liberal journalism, right wing journalism, and tech journalism. I run various deep learning methods on the articles to both frame the topics through which AI is

currently being discussed, as well as generate insights into the implicit associations that emerge in journalism about AI.

One reason I include this section of the project is that there are currently very few useful reviews of AI in journalism—in part a result of how recent all this buzz around AI is—and the history of humanities scholarship reveals how profound the products of metadiscourse can be. The other value in this section lies in its methodology, for using deep learning as a method for my analysis allows me to introduce AI as a way of seeing. To see through AI's eyes/I's—for, as a product of the patterns in our collective data, AI may also be *all* the *I's*—is the most pedagogical way to portray the affordances and limitations inherent to the technology. This methodology also allows me to clarify the rhetoric that has been shaped by AI technology, for I assert that within discursive space around AI are new ways of articulating what it means to human.

In Chapter 2, I analyze AI through the lens of science fiction (sci-fi). Sci-fi has always been central to our broader cultural understanding of AI, for its texts create narrative worlds to explore the cognitively estranging, existential questions that pervade all discussions of AI, either implicitly or explicitly. More generally, creative texts are the principal artifacts through which humans explore the meaning imbued in the objects and world around us, so I take these texts as data and analyze the patterns emerging in sci-fi about AI to explore how AI is currently being understood in culture. There has recently been an explosion of contemporary sci-fi texts that engage with AI in response to its proliferation across society, and I argue that these new films use AI in their narratives very differently from how AI was used in earlier texts. To make this argument, I begin Chapter 2 by recapitulating Seo Young-Chu's argument about late 20th century sci-fi androids in *Do Metaphors Dream of Literal Sleep*. Chu suggests that these androids'

narrative function is to challenge viewers' empathic boundaries and expand their understanding of what a human is, for which she relies heavily on N. Katherine Hayles' concept of the posthuman and Elaine Scarry's ideas about the nature of artifacts.

In my analysis of *Westworld, Ex-Machina,* and other contemporary texts sci-fi about AI, I suggest that AI androids are no longer represented in narrative as human-adjacent beings that deserve empathy; instead, AI is now represented as vastly more powerful than humans. AIs in sci-fi (often with humanoid bodies) now manipulate human behavior with ease, and their power stems from their control over the digital sphere and the human data within it. These films portray AI through the lens of surveillance capitalism, which is Shoshana Zuboff's term that describes how personal data has become the most valuable resource in the world as a result of these new AIs that shape human behavior. These films therefore present ontologically contemporary AIs that narrativize questions about data control to offer alternative frameworks for the self in the digital era. The crux of these films is then not about how viewers see the AI, but how the AIs see the humans, and it is through this dissonance between the humans and their data doubles—the term I employ to describe the imperfect copies of humans as they are understood through AI—that I offer another interpretation of AI's inherent humanist framework.

In the final chapter, I take this line of inquiry further by analyzing how AI intersects with creativity, a concept integral to our innate conception of what it means to be human, particularly within the field of cultural scholarship, in which we analyze human creative output to draw conclusions about the human condition. I rely significantly on Marcus Du Sautoy's *The Creative Code,* in which he argues that AI offers a new way to understand the algorithmic roots of creativity. I structure my analysis around the hidden associations that emerge at the intersection of the word vectors for "AI" and "Creativity," and I follow the tangents that illuminate the many

ways that AI can transform, inspire, and estrange our creativity. My aim in this final chapter is to show the value of AI as a way of seeing for humanist inquiries, as well as the emerging risks that can be best articulated within this discursive space.

I end by turning to the most intimate, ontological form of creativity: the act of creating a self. Contemporary philosophy and psychology tell us that the best framework for understanding a self is through narrative, as the unity of voice in novels planted a new way for humans to understand the epistemological connection between the past, present, and future of one's life. I argue, however, that a newer model of the self is emerging through the influence of digital media: the profile self. Using evidence from *Black Mirror* and journalism, I connect AI's influence in social media to the fractured, dynamic, and vulnerable profile self that I find traces of across contemporary media. I draw on a variety of sources, from media theorists like John Durham Peters to literary writers like Patricia Lockwood, to portray how AI provides a conceptual space to understand how these superficially dissimilar theories about the contemporary condition all are speaking to the same phenomenon. I assert that this profile self emerges as a result of how AI sees us, and only its underlying technical rhetoric affords us the language to describe the limitation inherent to a self as seen through AI.

To be human now is to be exploited, manipulated, and automated by AI, but there are also new forms of beauty, creativity, and meaning that arise through human collaboration with AI. What the world of tomorrow looks like will be determined by the technologies that are being built today, so AI culture is uniquely poised to shape its own, and our, future. AI, then, is nothing without the I.

Chapter 1

Reviewing AI Discourse in Journalism

The vast diversity of spheres that AI is already influencing contributes to what make it such an opaque technology to the public. An army of androids in science fiction and the social media algorithms that make our feeds so addicting do not share many superficial similarities, yet they are both AI. The dynamic nature of AI is part of what draws me to it as a conceptual framework: I can have discussions about AI with everyone from bankers to artists and kids to seniors because it is almost unavoidable in contemporary day-to-day life. The key to constructing a cohesive framework from such a versatile concept and technology is by approaching AI as a way of seeing. In this chapter, I use the lens of AI to both frame the emerging AI discourse in journalism, and to portray the affordances and limitations of its perspective. My overarching argument is that underneath all AI discourse lies fundamentally humanist questions, so this metadiscursive approach offers a first step in that direction.

The urgent tone that manifests in this chapter reflects the need I see for a metadiscourse about AI, for we are in midst of what is often described as the Fourth Industrial Revolution, but there appears to be no general consensus as to whether this is something to be celebrated or feared. Past research has shown that media coverage can drive public discourse about new technologies (Ouchchy), so in this chapter I review how the hidden associations in the media's AI discourse reveal how humanist ideas are embedded even in the ideologically inflected language of the media. Furthermore, I suggest that this discourse then affords a rhetorical space beyond the reach of political ideology that offers new ways to describe, think about, and see what it means to be human.

Review of Scholarship

There are several scholarly literature reviews that analyze the culture around AI, but their findings have been greatly limited by their methodology and timeframe. These limits are both because AI is being written about in the media significantly more in recent years than it was in the early 2010s, and because the computational methods I rely on to derive insights from a large data set of articles are the result of the same recent deep learning breakthroughs.

The first noteworthy review of computational culture was done by Stahl et al. in 2016, and they looked at how academic papers describe computing technologies, which includes AI as well as other less advanced computational processes. They found interesting results related to how specific values are described in computational academic literature—for instance that autonomy is most often discussed in relation to privacy—but their broad focus on computational culture makes it difficult to directly connect any of their conclusions to the contemporary media environment around AI. Moreover, by only using academic papers, their review represents only a very narrow subset of cultural production. They conclude that privacy is the most salient issue in the academic discussion of computational ethics, which, while constricted by the scope of their research, is useful because it creates a new space in scholarship for meta-analyzing computational culture.

Fast and Horvitz had a more specific approach in their 2017 study, which traced how artificial intelligence is described in the *New York Times*. They showed that AI was mostly described in relation to chess in the late 1990's, which then gave way to a focus on search engines in the early 2000's and then driverless vehicles in 2016 (4). Their temporal scope provides a very useful longitudinal perspective, but their limited dataset of only *New York Times* articles constricts their conclusions, for the *New York Times* readership is generally thought of as

well educated and not too dissimilar from the academic demographic. Some of their specific findings, however, like that the fear of loss of control of AI has been increasing in recent years (5), are very useful for understanding the context of AI culture today. Overall, the study shed light on how the topics through which AI is discussed change as AI permeates new fields and technologies, and this temporal frame is unique for such a young niche of research.

Chuan's 2018 paper, "Framing Artificial Intelligence in American Newspapers," was the first research effort to take a broader approach to analyzing media coverage of AI. They included articles from four major American newspapers (*New York Times, LA Times, Washington Post,* and *USA Today*) and came up with a final dataset comprised of 399 articles. Their dataset is thus more representative of contemporary public opinion than either of the two more limited approaches outlined before, but the newspapers they used for data differ more geographically than ideologically, as they are all mainstream, liberal publications.

Their quantitative analysis focuses on how AI is framed in the media. For example, they found that AI was most often framed in terms of personal impact in the Business and Economy, Science Fiction, and Entertainment topics, while AI in the Threat, Politics/Policy, and Ethics topics was more likely framed in terms of societal impact (3). A major factor that limited the sample size of this research is that it was published in 2016, and, as they noted, AI is being discussed exponentially more since then.

The most recent and most thorough review of AI in the media is from Ouchchy et al., whose 2020 paper, "AI in the headlines: the portrayal of ethical issues of artificial intelligence in the media," used data from 563 articles collected in late 2018. They similarly show that the number of articles written about AI has almost doubled every year since 2015, so their data set is mostly comprised of articles in 2018 and 2017.

Ouchchy et al. organized their findings around the tone, types of technologies, and issues in the articles about AI. Their most pertinent finding is that only recently have news articles begun wrestling with the ethical questions that emerge with AI, for not a single article before 2013 approached AI with an ethics-oriented perspective (930). This research provides important groundwork for the academic discourse about how AI is discussed about in the media, yet their articles all came from the NexisUni database, so it likely only included newspapers akin to those used by Chuan's study.

Thus, a few trends emerge from the existing research. For one, AI is being written about exponentially more now than a few years ago, and the trend appears consistent. Another is that newspapers tend to only offer a superficial exploration of the ethical issues caused by AI, which is also unsurprising given the complexity of the subject for readers of mainstream publications. Therefore, while the existing literature portrays that AI is becoming an increasingly pertinent subject in the media, there is very little scholarship that analyzes the substance of the media's coverage of AI. Moreover, there is no analysis of how the growing polarization of the media affects AI culture, yet plenty of research has shown liberals and conservatives no longer trust the other side's media (Spohr).

My Data

For my own research on the topic, I felt that I needed to work with a much broader, more representative data set than the ones used in other papers. My main concern with using established newspapers was that the media ecosystem has changed dramatically in the past five or so years, such that a significant portion of the US would describe *The New York Times* as "fake news." Donald Trump's war on the press has led to a splintering of the media system into

rival factions, which I will refer to loosely as liberal and right-wing media. The liberal media environment is comprised of newspapers and organizations such as *The New York Times* or *The Guardian* that value factual reporting yet are known for at least a slight liberal slant.[2] On the other side, right wing media has some larger companies, such as the Rupert Murdoch owned *Fox News* and *New York Post,* as well as a network of decentralized, novel media networks such as *Breitbart* or *Hot Air*. Of course, there have always been media outlets that represent ideologies across the political spectrum, but never before has the valley between them been so fraught.

This divide in the US media would make it difficult to derive significance from only mainstream liberal sources, so I decided to include three unique spheres of journalism to create a fuller picture of how AI is described in the media. Considering that AI is discussed through a variety of topics, I felt that it would be more useful to see which topics were covered by which media niches, as well as look for key areas of overlap or dissonance between the niches. Moreover, I hoped that in those areas of overlap may lie a space of opportunity; for, as it becomes increasingly difficult to communicate across the political aisle, AI may offer a lens through which political opposites can understand one another and find common ground.

My dataset is thus comprised of three corpora. The liberal corpus has 1167 articles from *The Guardian US*. *The Guardian* is the only free, mainstream, liberal newspaper available to scrape, yet I found it sufficient considering that other liberal sources have been studied, and there is no reason to suggest that there is any significant difference between the journalism in *The Guardian* and any other large liberal media company. The average publication date for the articles in the liberal corpus is December 2016. I decided against any strict longitudinal analysis,

[2] While those further to the left may feel slighted by the spectrum I have outlined, I can only counter with the fact that the majority of Americans get their news from either a mainstream liberal publication or a right-wing source, so this is the best framework available given my limited resources.

as it has already been established that AI is talked about much more recently, and I was more invested in the substance of the content, rather than merely the metadata. Thus, for all of my analyses, I treat all the articles as representing the same snapshot in time.

The right corpus was more difficult to assemble, as the right-wing media ecosystem is newer and less centralized than the liberal media ecosystem. The corpus I assembled has 1264 articles total, including 772 Articles from the *New York Post*, 100 from *Hot Air*, and 392 from *WorldNetDaily*. These three news sources provide a diverse and representative assortment of right-wing news, as the *New York Post* is owned by *News Corp* (and, therefore, somewhat mainstream conservative), while *Hot Air* and *WorldNetDaily* are less connected to any media establishment. The average date of publication for articles in the Right corpus is mid-2017.

The tech corpus functions partially as a control group, as the majority of tech journalism aims to avoid any overtly politicized rhetoric. Some of the analyses I run, moreover, are comparative, so the tech corpus helps anchor the differences between the other corpora. The tech corpus has 1075 articles total, of which 578 are from *Gizmodo* and 507 are from *The Verge*. The average publication date of articles in this dataset is early 2018. *Gizmodo* and *The Verge* are both popular, tech websites owned by larger media conglomerates, and both consider themselves as tech platforms, so their articles represent a valuable slice of data at the intersection of pop culture and technology. I assumed the tech corpus would have more precise technical language and may be more optimistic about the industry that its readers make their living in, for it is likely that the software engineers and coders who create AI are also disproportionately readers of tech journalism. Furthermore, if we consider the liberal corpus as representative of contemporary American liberal ideology and the right corpus as representative of contemporary American conservative ideology, then we can categorize the tech corpus's ideological slant as techno-

utopianist, which is the notion that technological advancements will lead to an improved society (Segal). While this is presented as an assumption, the data will reveal whether or not the tech corpus actually aligns with this perspective.

Topic Modelling

To analyze the data, I will begin with a relatively distant perspective: topics. Topic modelling is a machine learning method that reads through all the articles and deciphers the most common topics, which are expressed through the words that constitute the topic. This analysis is useful but limited, because it reveals patterns between the corpora, yet the exact topics can be difficult to read into since the algorithm may classify the articles based on words that hold little meaning outside the model. The best way to show how this works is by analyzing the topics that emerged in my research, and this will also give us a first introduction into seeing through the lens of AI.

Thus, when I ran the topic modelling method on all of the corpora together, these are the seven topics that emerged:

Topic 0: said technology ai intelligence government china artificial new use public

Topic 1: world trump new america china people president war american percent

Topic 2: ai robots people human work robot new like jobs data

Topic 3: human advertisement ai intelligence like artificial humans machine world just

Topic 4: google company facebook ai amazon like new users said tech

Topic 5: like film new sex said year just man robot time

Topic 6: ai like says used data people researchers learning machine algorithm

Before looking at their distribution in each corpus, we must do some translation; for AI can find patterns, but it cannot tell us what they mean. Topic 0, here, is the only one to specifically mention "government," so most of the articles likely have to do with foreign policy and the US's competition with China for increased control over future of AI globally. Therefore, we can think of Topic 0 as "Government." By the same logic, Topic 1 can be thought of as "Trump." Topic 2 is where things begin to get messy, but it is the only one that mentions "work" and "jobs," so we can call topic 2 "Jobs." Topic 3 is rather vague, and the inclusion of "advertisement" may be due to the program reading the word "advertisement" in lieu of an ad that would normally be unrelated to the article for readers, so I will mostly ignore this topic. Topic 4 includes three of the largest technology companies (Google, Facebook, and Amazon), so we can call it "Big Tech." Topic 5 is clearly distinct and can be termed "Movies/sex." Finally, topic 6 is the only one to include "research" and feature "data" in such a prominent position, so we can think of it as "Data Research."

Figure 1: Topic Distribution of the Corpora.

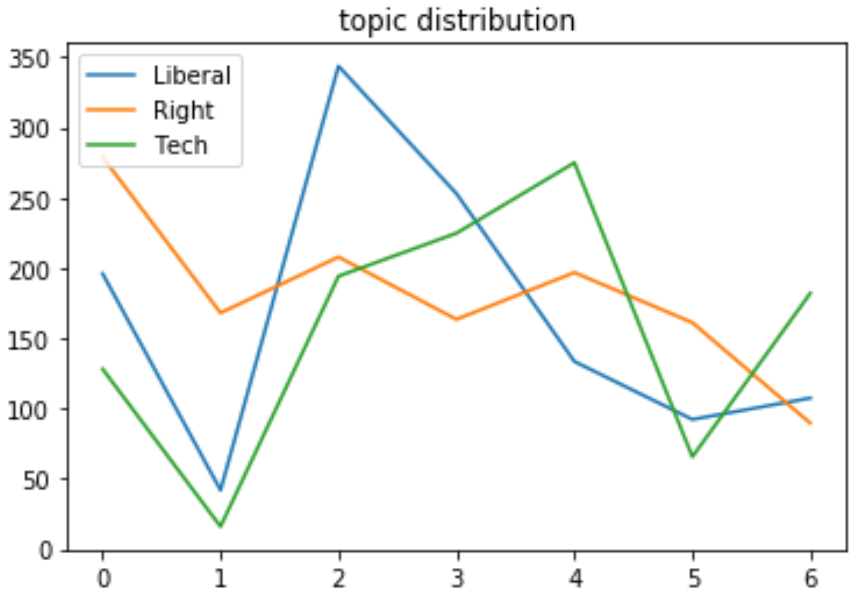

Figure 1 shows how the topics are distributed by corpus. Topic 0, "government," is most common in the right corpus, followed by liberal and then tech. Topic 1, "Trump," shows a similar distribution, although it is less prevalent in every corpus. It is unsurprising that the right corpus has written the most about AI in terms of Trump/government, for most of their articles present information through Trump-tinted glasses. Topic 2, "jobs," is most prevalent in the liberal corpus, yet, given the lack of any words that suggest the tone of these articles, the only conclusion we can reach is that liberal media discuss AI's role in the job market more than tech or right-wing media. Topic 3 I will neglect for the same reasons I outlined before. Topic 4, "Big Tech," is most prevalent in the tech corpus, which is particularly surprising given the amount of political scrutiny that many technology companies have faced since 2015. Topics 5 and 6 are the least prevalent in the data, but it is noteworthy that the right corpus has the most articles from the "Movies/Sex" category and that the tech corpus has the most about "Data Research."

These findings help by adding some detail to the emerging image of the media's representations of AI, albeit superficially. In line with Fast and Horvitz, it is clear that media coverage of AI is dynamic, and the specific topics through which AI is discussed depends on the culture of the time and identity of the media outlet. Furthermore, this surface level analysis offers an ideal starting place because it shows that the data is legible and cohesive; topic analyses require a lot of analytic work done by a machine learning program, so the fact that the results make clear sense is a signal that the data is meaningful.

In summary, then, liberal sources describe AI in relation to jobs more than the other two, whereas right-wing media covers AI in relation to government/Trump more, and tech platforms talk about AI through big tech companies the most.

Word Vector Analysis

The next level of my analysis is at the level of meaning. While the topic modelling offers surface level insights, a word-vector approach shifts the focus to how the words are used within the corpora. The word-vectors are made in a machine learning program called Word2Vec, which uses a neural network to learn the meaning of words through their relationship to other words in a dataset, and then it ascribes them a quantitative value in the form of a large array. The models can then be searched to find the most similar words to specific terms or concepts within the dataset. This creates the ability to analyze how a word is used in a large dataset, or, as I aim to do, compare how words are used differently between the corpora.

For example, a word vector for "sandwich" from an imaginary dataset would likely return results such as "bread," "peanut butter,"[3] and "jelly," as well as words one may often associate with sandwich, such as "lunch" or "salad." However, word vectors can also be added and subtracted. Thus, if we subtracted "peanut butter" from "sandwich" ("sandwich" – "peanut butter") the program would likely return "jelly" first, followed by other items that are in sandwiches but are not peanut butter. This creates the ability to test abstract or metaphorical relationships quantitatively, for if A is to B as C is to D, then the vectors for A + C – D should return B.

Another layer to my project, though, is the comparative dimension to word-vectors. Word-vector models understand words in terms of the other words in the dataset, but this can make models difficult to compare because datasets may not have many words in common. Moreover, word-vector analyses return a number with every result that represents the distance

[3] Word vectors need to be one word, so "peanut" and "butter" would be separate vectors, but I chose to ignore that for the sake of the example. (I used this example instead of King ~ Prince || Queen ~ Princess because it is overused in digital humanities scholarship.)

between the result and the word searched for. This distance can be useful for understanding strange results that emerge, for if the values of the vectors returned are low then it implies that there were very few results that fit well. My results all pass this test, but I include the figures in larger tables to show the validity of the vectors that I analyze in depth. Also, these numbers can only carry significance if the models have been normalized, so all the results that follow are valid as the models have been normalized.

Word embeddings may be most clear if we understand them from a distance. Their underlying logic is built on the idea that the meaning of words in language is embedded in a quantifiable way. Through AI's lens, words function like vectors, pushing the meaning of a text in countless directions, each in different dimensions. Word2Vec creates an entire universe whose only occupants are the unique words in the text; their connections create values that occupy the hundreds of dimensions that this model universe creates. Meaning has always been a product of the relationships between words—not inherent to the word itself—so this lens affords us the ability to explore the associations between words like a telescope allows astronomers to see patterns in the sky. The internal data science done by the model is an opaque process, but the results it returns can only be read, interpreted, and transformed back into our much more expansive semantic universe outside the model by human eyes. To read word vectors is to put on AI-tinted glasses and explore a distant universe with hundreds of dimensions and thousands of vectors cutting through. To interpret the results is to capture the surprising patterns and export them back to familiar territory, and the land I hail from is the humanist tradition.

To better acquaint us with the data, the first results I will look at are from doing a search for the most similar word vectors to "AI" in each corpus. I leave out the numerical values for this table (and several others), where they would be more distracting than insightful.

Table 1: Results for Most Similar Words to "AI" Vector

"AI"		
Liberal Corpus	Tech Corpus	Right Corpus
Technology	Technology	Technology
Software	Software	System
Development	Development	Machine
Automated	System	Software
Itself	General	Computer

The main insight from *table 1* is that these corpora are superficially quite similar; they all refer to the same notion of "AI" as a technology and software. To access deeper insights into the data, let us next look at more robust vectors that frame AI differently. The first I will analyze is "AI" + "Freedom," for freedom is as American a value as it gets, and it is one that often carries different connotations depending on one's political alignment. In liberal ideology, freedom is often assumed to be freedom from oppressive forces, whereas in conservative ideology freedom is generally conceived of as the freedom to do as one pleases without being infringed upon.

Adding vectors to AI allows us to see how each ideology views a value or set of values (freedom in this first example) through the lens of AI. The "freedom" vector is distinct in each corpus because the model learns what each corpus means by "freedom" independently, so the results essentially present an analogy for how each corpus thinks about freedom as it relates to AI.

Table 2: Results for Most Similar Words to "AI" + "Freedom" Vector

"AI" + "Freedom"		
Liberal Corpus	Tech Corpus	Right Corpus
Transparency	Society	Automation
Regulation	Automated	Society
Disruption	Malicious	Drones
Wider	Robotic	Foreseeable
Legal	Ethical	Brain

The liberal corpus in *table 2* differs significantly from the other two corpora. Intertwined with liberal notions of freedom and AI are words that relate to the difficulty in regulating AI—"transparency," "regulation," "disruption," and "legal"—that are not present in the tech or right corpora. This suggests that there is a distinction in how these different spheres of journalism discuss AI. For the liberal corpus, any notion of freedom tied to AI is dependent upon the legal system's ability to control it. Freedom, then, is articulated in the liberal corpus in terms of the peoples' rights that AI may infringe upon.

Most of the results from the tech corpus suggest that freedom is discussed through AI as a vision for a future utopia: "automated," "robot," and "society." This suggests that for the tech sphere, AI is a means to reach new freedoms across society. The other two results, "malicious" and "ethical," reveal that embedded in any notion of a future utopia are ethical—or human—concerns.

The results from the right corpus similarly suggest that AI and freedom can lead to an automated, future society. However, there is a militaristic dimension to the right corpus with the inclusion of "drones," which extends the metaphor of freedom through force into the digital sphere. The fifth result, "brain," suggests that AI's permeation of the human mind may be an opportunity for new freedom, rather than an Orwellian measure for societal control. The right corpus's results for this vector thus imply that AI may lead to increasing individual subjects' freedom, which situates the right sphere closer to the tech niche than the liberal bubble.

AI and Narrative Vectors

Next, let us shift to looking at how AI is represented in the media when we add values that resemble the way AI is portrayed in pop-culture. This allows us to see through AI what

associations emerge from the media discourse when we add to AI the attributes that are often afforded it in science fiction narratives. These texts, such as *Ex Machina* or *Westworld,* contribute to AI discourse by posing questions about AI by making presenting it narratively, but I engage with them in depth in Chapter 2. Before that textual analysis, though, let us see what patterns emerge directly from the media discourse. To start, we can look at what the results are returned when AI is combined with emotion, for emotion is a distinctly human phenomenon that is often used to differentiate humans from AI in narrative. The equation I came up with is "AI" + "Emotion" – "System," because I want to get results that are more value-oriented, rather than technical (which negating "system" mitigates).

Table 3: Results for Most Similar Words to "AI" + "Emotion" – "System" Vector

"AI" + "Emotion" – "System"		
Liberal Corpus	Tech Corpus	Right Corpus
Weaponized	Risks	Concepts
Dangers	Humanity's	Entertaining
Regulating	Gadgets	AI's
Societal	Underway	Whoever
Rapid	Talented	Creativity

Once again, the liberal corpus connects the intersection of AI and emotion with concern, for the top results, "weaponized" and "dangers," suggest that the liberal corpus is keenly aware of the vulnerability that occurs with emotionally intelligent AIs or AIs that manipulate human emotions. Liberal journalism, we can deduce, suggests AI's integration into our emotional lives is something urgent ("rapid"), as well as something that we must come up with a collective response for, as suggested by "societal" and "regulating."

The first result for the tech corpus, "risks," falls in line with the liberal discourse, yet the other results paint a more nuanced picture. The sites where AI and emotion may intersect could be through our "gadgets," yet the inclusion of "humanity's" (not humanities) suggests the effects

will be existential and will affect the entirety of our species. “Underway” is similar to the “rapid” result from the liberal corpus that references the timeliness of this transition, and “talented” likely refers to how tech journalism describes AIs that already are capable of emotional insights.

The right corpus has quite different results from the other two. “Entertaining,” the second result, suggests that the right-wing media articulates AI’s role in our emotional lives as connected to entertainment. Entertainment is how we describe the things that we look at or experience to occupy our minds, so its vector is likely associated with how AI entertains us through social media feeds, even though we rarely describe the experience of scrolling through Instagram in such accommodating language. Entertainment is generally a positive term, for our minds are most occupied—or satisfied—when we become emotionally invested the content we are watching. This relates to how AI is also reshaping the architecture of the entertainment industry, for Netflix and other entertainment platforms in cyberspace rely on AI to nudge users towards the content that is most likely to resonate with them. In this way, AI sees people (in the context of entertainment) like these word vectors, embedding our emotional complex experiences in opaque arrays—this is an idea I investigate in depth in Chapter 3. Most striking in the right corpus, though, are the results that are not present; for, the right appears much less skeptical over the direction in which AI is influencing our emotional lives. This may be explained by the fact that the political right is often associated with a more macho brand of American masculinity that is less concerned with emotional vulnerability.

There is an essential facet to AI that I have so far made use of but have not explained, which is the duality of AI. By duality, I am referring to how AI is, on one hand, a figurative, futuristic concept of artificial life that is intelligent and is often presented in narrative in a

robotic, anthropomorphic body;[4] and, on the other hand, AI is in software/systems/technologies in the present, which we encounter through automated personal assistants, apps, and other media. This duality can also be understood as weak AI vs strong AI: weak AI is what we have today, and it can do straightforward tasks incredibly well. Whereas strong AI, or AGI (artificial general intelligence), refers to an AI capable of all human tasks, which has not yet been created but in most estimates will arrive within several decades (Tegmark 30). Importantly, though, cultural depictions of strong AI have the power to shape the technologies currently in development, for the people that build the AIs of the future may base their creations off of fictional, cultural projections, as was the case with the first digital personal assistants' (Alexa and Siri) disposition being based of one of the first AI icons, HAL 9000 (Flahive).

Before we introduce the lens of science-fiction analysis to consider strong AI more theoretically in Chapter 2, let us look at how the future of strong AI is already being described. The next vector I created imbues AI with attributes that we commonly associate with anthropomorphic AI (androids) in popular culture. The equation I arrived at is "AI" + "Emotion" + "Intelligence" + "Robot" + "Consciousness" – "System." This vector shows us what an AI with these attributes might look like or imply through the ideological lens of each corpus. The reason that this table include the numerical values for the returned vectors is because they offer a valuable insight for this particular vector that I describe in the analysis.

Table 4: Results for Most Similar Words to Android Vector

"AI" + "Emotion" + "Intelligence" + "Robot" + "Consciousness" – "System"		
Liberal Corpus	Tech Corpus	Right Corpus

[4] This is particularly true for contemporary AI television shows and film. Some examples I am thinking of are *Ex Machina, Westworld, Black Mirror,* and *Raised by Wolves*.

Principle	.84	Sexuality	.93	Safely	.92
Humanity	.83	Immortality	.89	Sophistication	.92
Synthetic	.83	Selfawareness	.89	Entirely	.91
Animal	.82	Necessary	.87	Perception	.91
Truly	.82	Superintelligent	.87	Process	.90

In *table 4*, the liberal corpus's results for the android vector have finally cut deeper than the regulatory rhetoric. The top results, "principle" and "humanity," suggest that an AI with these attributes would challenge our most fundamental principles. The next two results, "synthetic" and "animal," are terms we often use to distinguish non-human entities from humans, suggesting that the liberal corpus does not see a future in which AI and humans become intertwined. The overall lower scores for the values in the liberal corpus in *table 4* are most likely because the liberal corpus had the largest vocabulary, so after the three corpora were normalized, the distance between concepts was lessened by the significant reduction to its vocabulary.

The tech corpus, however, reveals very different values associated with the android vector. "Sexuality," "immortality," and "self-awareness" are all accurate descriptions of what the androids in contemporary sci-fi reckon with, and these results reveal a rather speculative depiction of a future alongside AI robots. The lens of analysis, furthermore, is the AI itself, as opposed to the liberal corpus, whose results focus more on the existential challenges such an AI would pose to humanity's sense of self. This shifting of perspective is significant because it reveals that in the sphere of tech journalism—whose readers more likely include the software engineers and other technical workers that are creating these AI of the future—the discourse around the future of AI is more closely tied to sexuality than the existential effects it will have on humanity, particularly so because of how much higher the value of "sexuality" (.93) is than any other result in the tech corpus. This also fits the general techno-utopian philosophy, for any

future utopia ushered in by AI would have to include some sexual liberation for humans (see *Westworld* season 2 for spoilers on how this narrative ends).

As for the Right corpus's results, "safely," the top result, is an adverb, which implies that the android vector leads to a space of action (for adverbs require verbs). The right thus appears more invested in the process by which a strong AI may emerge – "process," "entirely," and "safely." The second result, "sophistication," is how one may describe a strong AI, suggesting that the right corpus already describes AGI in human terms.

Taking a step back, it is noteworthy how much more diverse the results for this vector are than the initial vector for AI in *table 1*. While the focus of this thesis is not political science, it nevertheless reveals how different the rhetoric can be across the political aisle, even for AI, which is not particularly politized in culture. It also portrays how AI becomes more interesting when it is analyzed through the lens of different aspects of humanity that it threatens to alter.

Next, then, let us look at a less sensationalized yet similarly conceptually intertwined vector. My interest in this project is in analyzing depictions of the future, so the next vector will aim to encompass the future of humanity with AI. I came up with "AI" + "Humanity" + "Future" – "Technology," so that the results would be more existential than technological.

Table 5: Results for Most Similar Words to Future of Humanity with AI Vector

"AI" + "Humanity" + "Future" – "Technology"					
Liberal Corpus		Tech Corpus		Right Corpus	
Threat	.81	Benefit	.82	Process	.88
Goal	.77	Form	.80	Society	.88
Thinking	.77	Life	.78	Best	.88
Creating	.76	Redefine	.78	Change	.87
Consciousness	.76	Society	.78	Problem	.87

In *table 5*, the differences between the corpora are becoming more apparent. The liberal corpus here sees the future of AI as a "threat," which is particularly significant considering that it

is the only result whose value is significantly higher (.81) than any others (.76-.77). “Threat” is also only present here for the liberal corpus, so we can deduce that liberal media is more concerned about the future of AI than the other journalistic niches. Moreover, the juxtaposition between the first result and the others, particularly “goal,” suggests an ambiguity towards a future with AI. “Thinking” and “creating” both map onto how one may describe the future of strong AI to distinguish it from weak AI today, which mostly processes but does not think and analyzes but does not create. “Consciousness” is interesting too because AI consciousness is thoroughly explored through cultural texts, yet the liberal corpus did not have many articles that fell under the “movies/sex” topic. This would imply that the liberal corpus is engaging with consciousness as a realistic notion, and one that is conceived as both a “threat” and “goal.”

The tech corpus offers a much more positive depiction of the *Future AI* vector. “Benefit,” the top result, clearly implies that AI is nothing to be concerned about. “Form” may imply that AI will *form* a new future for humanity, or that AI in the future will itself be a new *form* of “life.” Either reading fits with the other results, as both imply a process of “redefin[ing]” the fundamental organizing facts of “society.”

The results from the right corpus are somewhere between the other two, as the first result, “process,” does not carry any explicitly positive or negative implications. It does, however, suggest an apprehension towards existential analysis in the right corpus because neither of the top results from the other two corpora show any interest in the nature of this vector, just in the direction of its impact. The data from the right corpus thus suggests that the future of humanity with AI may be a process, which is to say that it may be a methodical progression towards a new “society.” The third result, “best,” is reminiscent of Trump’s rhetoric, and it is not too hard to imagine a Trumpian technophile describing the future of AI in America as the “best.” The last

result, "problem," reveals that the negative effects of AI are more connected to the right wing's depiction of the future than the tech corpus's, which situates the right corpus's results somewhere between the other two in terms of sentiment.

These results suggest that political ideology plays a role in imagining the future with AI. The future, however, may be a more ideologically inflect framework than AI, so let us next look at a similarly existential vector with a different composition. Considering that many of the results depict this future as connected with our definition of life, the next is a vector for artificial life: "AI" + "Life" – "Technology." I subtract technology once again because it leads to results that are more figurative and less tied to specific technological rhetoric.

Table 6: *Results for Most Similar Words to Artificial Life Vector*

"AI" + "Life" – "Technology"		
Liberal Corpus	Tech Corpus	Right Corpus
Brain	Sexuality	God
Intervention	Immortality	Right
Body	Humanity	Someone
Conversation	Converted	Best
Nature	Consciousness	Everything

The *artificial life* vector for *table 6* is ambiguous in nature because of how dynamic a concept "life" is, but, since all the data are from articles about AI, we can assume that "life" here is already filtered through the lens of AI technology. The results should therefore describe the sites and concepts through which AI alters or becomes life.

The liberal corpus's top result is "brain." We can read this a couple ways: the liberal media sees the brain as the chief site through which AI will alter life, or that mimicking the brain is the way in which AI becomes life. Either way, the liberal corpus is judging *artificial life* from humanity's perspective; the human "body" and "brain" still remain integral to the concept of "life," even through the lens of AI. Peters suggest that "all media are species specific" (310), so,

if we accept AI as a human-made media, then it makes sense to analyze life in relation to human life, rather than approaching it through a biological lens. “Intervention” and “conversation” imply a dialectical nature to the *artificial life* vector, which is distinct from the techno-utopian rhetoric of the tech corpus or the logistical language of the right corpus from the *future* vector in *table 5.* These results all clearly articulate liberal values if we understand the liberal ethos as one that values collaboration, creativity, and logic.

The results from the tech corpus resemble the results from the android vector in *table 4*, which is noteworthy because neither of the other two corpora’s results do. We can glean, then, that the connection between “sexuality” and *artificial life* is not simply explained by a desire to have sex with robots (as would be inferred if it only surfaced in the android vector, which includes “robot,” a word that implies a physical body). Sexuality then, as a vector, is closer to utopian fantasies in the tech corpus than either of the others, and I will unpack this further in my analysis of recent sci-fi texts about AI because the sexualization of feminine androids in culture is a pattern that does not require an AI lens to see. This vector, however, suggests that it may have some origin in tech media discourse. “Immortality,” here, is ambiguous as to whether it refers to the AI or a human-AI hybrid, but its proximity to sexuality implies that this *artificial life* vector returns aspirational vectors in the tech corpus that are void from any religious or value-oriented rhetoric.

The right corpus’s results are similarly existential. This techno-utopianist perspective of an unlimited *artificial life* coincides with conservative media’s concept of “god,” which is the top result. The second result, “right,” may refer to right in the sense of an entitlement or as a synonym for fair, yet either reading connects *artificial life* to the fundamental values of human life in conservative ideology—God and truth. Each corpus, then, connects the *artificial life*

vector to the essential elements of human life, albeit through very different ideological lenses. In other words, the brain is to liberals what sexuality/immortality is to techno-utopianists is what god is to conservatives; or, we project our futuristic fantasies of omniscience or immortality onto AI. This is significant because it reveals how dynamic a concept AI is: the values that one attributes to artificial life say more about the values of the individual than anything inherent to AI. Our discourse about AI may then teach us more about humanity than the technology itself will.

These values are ideological in nature, but that does not mean they are politicized in the reactionary sense. There has recently been significant scholarship focusing on how events or actions become politicized. One study showed that events are often discussed in shared forums until they become explicitly political, after which they are discussed solely within echo chambers of people who all share the same political ideology (Barbera). If we apply this framework to the field of AI more generally, then we can glean that while AI is not yet particularly politized, its existential implications must be interpreted through value systems that are tied to political ideologies. Thus, AI is not yet politicized, but it poses ideological and existential questions, which is how AI discourse creates a rhetorical space that functions beyond ideology. A useful counter example may be climate change, which is a similarly existential threat looming over the near future's horizon, yet it has become such a politicized issue that it is near impossible to discuss in any shared forum.

My guiding goal for this chapter is to find spaces of conjunction in the media's representation of AI that transcend political ideology because it is essential that work is done to prevent AI discourse from following a similar path to that of climate change. The stakes of this

search are high, for AI technology is already transforming the world, and society's only way to mitigate the risks may be through the political realm.

This analysis connects to the notion of AI's duality that I introduced earlier. AI is real, yet it is also speculative. We do not know what the AIs of the future will look like or how they will function, but we know they will exist. We imagine futuristic AIs as somehow adjacent to humanity, which could mean that the AIs look like humans, have human desires, or act in service of those desire. AI's true value in its figurative sense, though, is in how its discourse creates a space to discuss what our human values are. In other words, AI's true conceptual value is in how it demands we define what it means to be human. AI thus poses a challenge to humanity by threatening to do most things better than us and render us useless, so what will give human lives value in the future? This question reveals how AI discourse is fundamentally a humanist project, and it is not merely figurative, for it is only figurative in the present; AI will become more powerful and omnipresent in the near future, so it will redefine what it means to be human as it permeates new spheres of human life unless we first decide which elements of our humanness are better left unseen—unshaped—by AI.

The logic embedded in this formulation will become clearer in the next, more theoretical chapters. For now, though, we can word this question in AI's language as: "Humans" – "AI" = X? With X equal to that which makes us essentially human and distinct from AI. This question is useful in the metadiscursive context because AI provides a uniquely non-religious, non-politicized theoretical framework that is accessible across ideologies. Perhaps, then, AI may paradoxically serve as the ideal conceptual tool to investigate the guiding doctrine for the humanities—the human condition.

The equation I arrived at for a *human life* vector is "Human" + "Life" + "Humans" + "Self" – "AI." "Humans" and "self" are included to create a more robust vector, and I include the top eight results for this analysis because of how diverse the results are.

Table 7: Most Similar Words for Human Life Vector

"Human" + "Life" + "Humans" + "Self" – "AI"					
Liberal Corpus		Tech Corpus		Right Corpus	
Touch	.82	Worse	.92	Care	.93
Meaningful	.81	Freedom	.91	Experience	.92
Rich	.81	Brains	.90	Shape	.90
Either	.80	Situations	.90	Anywhere	.90
Empathy	.79	Nuances	.90	Certain	.90
Conversation	.79	Characters	.89	Decisions	.90
Emotions	.79	Believing	.89	Preparing	.89
Experience	.79	Certain	.89	Finding	.89

What, then, is the essence of human life (and the human self) that is irreducible to AI? The liberal corpus offers a compelling list of answers. The top result, "touch," connects to the corporeal sensation/experiences that AI, particularly in its present form, cannot mimic. "Meaningful" suggests that the meaning we imbue on our lives is what separates us from AI, which is also persuasive. "Empathy," "emotions," and "experience" all similarly relate to intrinsically human concepts that cannot exist without human subjectivity and all its messiness. Even "either" is saturated with significance, for it highlights the ambiguity inherent to human life that may be untranslatable to AI, which sees the world through probabilities and binaries. As an example, I may have either a salad or a sandwich for lunch. Through AI's lens, there may be a 63% chance I have a salad and a 37% chance I have a sandwich. While nuanced, these are fundamentally very different ways of interpreting the world, and the importance of this dissonance is clearer if we consider the risk of quantifying more meaningful human experiences. For instance, AI is already playing a significant role in the criminal justice system. Underneath

the superficial language of objectivity and efficiency that inspire more and more spheres to adopt AI technology is the fact that AI cannot see ambiguity in a human way. The young man who stole some snacks to feed his family may be judged more harshly by AI because his desperation is a variable that suggests he may commit more crimes in the future, but to see someone through AI in this context is to strip them of their human capacity to grow and reshape their life story.

The tech corpus's results shed more light on this phenomenon. The first result, "worse," reveals a techno-utopian perspective in the corpus that sees humans as lesser in quality than AI. "Freedom," however, suggests that while the tech corpus may privilege AI, freedom is a value with an inherent connection to humanity. "Situations," "nuances," and "characters" are fundamental elements of stories, and it follows that that storytelling is an essential aspect of humanity that AI cannot replicate. The tech corpus, therefore, considers stories to be an essential aspect to what makes us human, even if it sees humanness as lesser in value than AI. This maps onto the example I described in the criminal justice system, for to be human involves seeing people through the lens of stories, not numbers. The connection between stories and humanness is one that I examine thoroughly in Chapters 2 and 3.

In the right corpus, the "care" vector is that which is essentially human. This is something we intuitively understand; to care about something is to feel something about it, and to feel is to be human. "Shape" is an interesting result because both its syntactical forms make sense: humans have shape—bodies—and humans can shape things (give form to ideas and objects). "Anywhere," the fourth results, likely corresponds to a vague vector, and one that carries no meaning through the lens of AI. If we consider how an AI sees a person through social media, they are either somewhere specific (if the AI has access to their location data) or they are

nowhere, which AI would represent with a null quantity. Anywhere, then, is a uniquely human space in its most abstract sense, and this maps well onto our emerging framework.

In summary, the liberal media sees interpersonal connection ("touch," "conversation," "empathy") as an integral component that separates us from AI. Tech media sees humans without AI as "worse," but also considers stories to be a key component of what makes us human. The right corpus privileges care and experience as representative of humanness. Thus, the *human life* vector shows which values are given to humans but not AI in these three niches of journalism. These conclusions could not be gleaned by reading a handful of articles; they only emerge when we analyze discourse using AI's lens as a tool, yet the results only make sense when they are returned to the human world of meaning. AI affords us a conceptual framework to examine what makes us human only if we understand *how* it sees and, more importantly, what it cannot see.

Discussion

When we consider many of the dominant descriptions of our hyper-contemporary epoch—late-stage capitalism, surveillance capitalism, gigification, the era of fake news—each can be connected to the rising influence of AI technologies. The societal and cultural changes ushered in by AI are increasing exponentially, as these technologies undergo revolutionary transformations at a pace that is many levels of scale quicker than the pace of human evolution.

As my own research has shown, AI is a vast, dynamic topic that affects almost every facet of contemporary life. Why, then, is there such a dearth in scholarship that aims to understand our relationship to it? The goal of this chapter has been to review the current media discourse of AI through the lens of AI to provide insights beyond those already documented and

portray the stakes embedded in AI as a way of seeing. By depicting how ideological lenses influence the media's representation of AI, I hope to encourage more research that looks into the nuances of how the internet's echo chambers shape emerging cultural fields. My findings suggest that AI is a valuable site for connection between interpretive communities because of its duality; AI is real and figurative, present and future, which imbues it with the conceptual power to unite discordant ideologies through a shared topic of interest.

The next step in this project is to test how these insights from the media discourse map onto pop-culture texts about AI. Artists have been reckoning with the implications of AI through narrative fiction for decades, so my guiding questions for the next chapters are: How have recent advances in AI been interpreted through pop-culture texts? What do these pop-culture representations about AI tell us about the future? And what do we really talk about when we talk about AI?

Chapter 2

Updating AI Theory in Science Fiction

Science fiction has played a significant role shaping how people understand AI, for popular films like *Blade Runner* and *2001: A Space Odyssey* were the main vehicles for AI discourse before the recent deep learning revolution. In this chapter, I will analyze how science fiction scholars have interpreted past representations of AI in films from the late 20th century to situate my analysis of AI in contemporary sci-fi. My argument will focus on how the logic of surveillance capitalism has fundamentally shifted AI discourse, which is made available for representation through a developing sub-genre of sci-fi that imbues androids with data control. I tie this textual analysis to the finding I ended Chapter 1 with—that AI's key conceptual value lies in how it redefines what it means to be human—to depict how humanness as a framework emerges in narrative.

Science-Fiction as a Theory of Representation

To contextualize my argument within the field of science fiction, I will rely heavily on Seo-Young Chu's *Do Metaphors Dream of Literal Sleep?* because it offers a refreshingly contemporary way of interpreting science fiction. She traces how sci-fi has traditionally "been understood as a genre whose objects of representation are hypothetical if not outright imaginary" (1), and she offers a compelling counterargument: that science fiction "operates fully within the realm of mimesis" (2), mimesis meaning the representation or imitation of the real world in art. Her argument, then, is that sci-fi is a discourse for representing cognitively estranging objects/concepts through narrative, which makes it much more similar to realism than fantasy. Realism, by Chu's logic, is thus a form of low-intensity science fiction, for the objects

represented in realism are simpler and more concrete than the real yet estranging objects represented in science fiction (7).

This follows from Chu's notion that the representability of objects is a spectrum. Her interest is in the middle of the spectrum, for those are the objects that science fiction makes available for representation, which include "cognitively estranging referents [that] encompass the sublime, virtual entities, realities imperceptible to the human brain, [etc]" (7). On the far right of the spectrum are objects that are virtually unknowable and that defy human language or understanding, such as what happens after death, which evade representation even in sci-fi. Furthermore, Chu suggests this spectrum is useful for gauging the value of a science fiction text, for "a successful work of [sci-fi] is one whose wondrous effect on its reader/viewer/listener reproduces the wondrous qualities of the object of phenomenon that the work of science fiction mimetically represents" (5). I find this part of her argument particularly compelling because it is the glue that sticks [sci-fi] to realist discourse, for it positions human experience (the experience of objects and of art) at the center of her analysis.

She pushes the boundaries of science fiction scholarship even further with her claim that "not only is [sci-fi] at once counterfactual and counterfictional, but it is also at once *counterliteral* and *counterfigurative*" (68). She means, then, that sci-fi relies on logical diegetic worlds that cannot be reduced to a simple figurative or literal reading. "As a result of this counterfigurative literalization, the cognitively estranging referent becomes substantiated into something kinesthetically recognizable… [which] enables [sci-fi] to represent objects and phenomena normally averse to representation" (68). In other words, sci-fi translates complicated real-world phenomena into representable features of narrative text, which is where sci-fi reveals itself to be pedagogical by nature. As Chu puts it, "science fiction equals the making of

knowledge…to make something available for representation is to make it knowable" (75). She ties this to our contemporary setting by pointing out that "cognitively estranging referents are growing more and more prevalent," (81) so, as our experience of the world shifts further from its material grounds, the pedagogical role science fiction plays is becoming increasingly important.

Chu is ultimately suggesting a "science-fictional theory of representation" (74), which neatly connects with my analysis of AI's duality. AI is exactly the type of cognitively estranging referent that is ripe ground for science fiction to represent, so what do representations of AI in sci-fi look like, and what are the strategies for making it available to cognition?

Chu's Androids

Before I dive into my analysis of contemporary science fiction, I will introduce some key concepts to the academic study of artificial life through Chu's arguments about sentient robots. By sentient robots, she means "artifacts—human-made objects—possessing human attributes such as selfhood, the capacity to fall in love, and susceptibility to grief" (214). In practice, she is mostly referring to the androids in the original *Blade Runner* (1982) and the android child in Spielberg's *A.I. Artificial Intelligence* (2001).[5] She focuses on how these robots largely appear indistinguishable from humans and argues that these androids are bestowed with humanity "through the lyric structures of voice" (235) within these texts.

Chu analyzes these androids as mimetic representations of the posthuman, which is Nancy Katherine Hayles' idea that "there are no essential differences or absolute demarcations between bodily existence and computer simulation, cybernetic mechanism and biological

[5] Android is by definition a robot that looks like a human, but they are not to be confused with cyborgs, which are part human part machine. Androids, instead, are AIs with human-like bodies, so anytime I mention "android," AI is implied.

organism, robot teleology and human goals" (3). Through a posthuman perspective, androids are human-made objects that are conscious subjects, and their claims to life, liberty, and the pursuit of happiness therefore challenge our conception of human. Chu includes an analysis of Mori's uncanny valley—the notion that almost-identically-human robots are a source of revulsion—to suggest that the core emotional pressure on viewers of *Blade Runner* and *AI* is to overcome the uncanny valley (which only arises in scenes where the androids become injured, and their internal mechanics become visible) and still experience moral respect for the androids (244). In other words, by challenging the limits of viewers' empathy, these android narratives lead viewers to expand their sense of human empathy to all forms of humanoid life,[6] including those that are already somewhere in the middle of the synthetic-organic spectrum. For, as Chu explains, "'Born' and 'constructed' exist on a single continuum. Many human bodies today contain artificial components such as hearing aids, pacemakers, retinal microchips, and psychopharmacological drugs that restructure the drug-taker's consciousness" (238).

Chu makes this point about androids' posthuman pedagogy by connecting her reading of these films to Elain Scarry's analysis of the nature of constructed things. In *The Body in Pain,* Scarry writes that "the act of human creating includes both the creating of the object and the object's recreating of the human being, and it is only because of the second that the first is undertaken: that 'recreating' action is accomplished by the human makers and must be included in any account of the phenomenon of making" (310). Chu's elegant conclusion suggests that through this lens, sentient androids "would exist for the purpose of 'remaking' human beings *into human beings*" (238; original emphasis). Thus, by creating androids in our image, we redefine the category of human being, which is pertinent because "as more and more humans

[6] Humanoid refers to anything human-like, so it is a less precise term that in this analysis will refer to essentially the same objects as "android."

find themselves in danger of inhabiting the uncanny valley, science fiction will become increasingly important as a way of representing human rights" (244).

In summary, the androids in *AI* and *Blade Runner* are made to mimic human beings, and the films portray them in a human light through narrative techniques such as soliloquy and apostrophe. This leads viewers to feel empathy for them, and it challenges viewers to extend their empathy to the diversity of human life possible through cybernetics.[7]

Androids in Surveillance Capitalism

The androids in late 20th century science fiction were largely conceived of as created in service of human beings. In *Blade Runner,* they are built to be workers, while in *A.I.*, the android boy David is built to replace a human son because he is the first AI with the ability to love. AIs in contemporary sci-fi, however, are ontologically quite different.

First, let us consider Ava, the AI android in Alex Garland's *Ex Machina* (2015). Ava is built by a reclusive genius, Nathan, who is the CEO of Bluebook, a search engine company that closely resembles Google. The film begins with Caleb, an employee of Bluebook, winning a competition that results in an invitation for him to perform a Turing Test on Ava.[8] The plot twist, however, is that Nathan is actually testing whether Ava will be able to manipulate Caleb to help her escape captivity, which connects to Ava's ontology.

Nathan explains that Ava's intelligence is built using data collected by Bluebook, for he describes how the data revealed a "map of how people were thinking... Impulse, Response. Fluid, Imperfect. Patterned, Chaotic." Ava's hardware, then, is actually wetwear, for Nathan's major

[7] Cybernetics refers to technology in the body.
[8] The Turing Test, developed by Alan Turing in the 1950's, traditionally consists of a typed conversation between a human and AI (or human, in the control situations), and the AI passes the test if the human cannot tell whether they are talking to a human or AI.

breakthrough was realizing that the human mind could not be mimicked by synthetic materials that were not similarly dynamic. This data-based ontology is very different from late 20th century depictions of AI, such as in *Blade Runner* or *AI,* because the connection between AI and data was not well defined then. For context, Marcus du Sautoy describes how data has redefined the field of AI through machine learning in *The Creativity Code*, which is a very recent development:

> So what has happened to launch this new AI revolution? The simple answer is data. It is an extraordinary fact that 90 percent of the world's data has been created in the last five years… This flood of data is the main catalyst for the new age of machine learning. Before now, there just wasn't enough of an environment for an algorithm to roam around in and learn… This is why access to data is so important: the more examples a smart algorithm can train on, the more experienced it becomes, and the more each tweak refines it. Programmers essentially create meta-algorithms which create new algorithms based on the data they encounter. (62)

Ava, therefore, is like a meta-algorithm, and the task she has been given is self-preservation. This notion of Ava as a meta-algorithm connects with the broader use of AI by tech companies today. Since the AI revolution of the mid-2000s that shifted to a bottom-up, data-led approach, personal data has become the most valuable commodity in the global economy. The effect of this novel approach to AI has led to a new era called surveillance capitalism, which Shoshana Zuboff defines as "a new economic order that claims human experience as free raw material for hidden commercial practices of extraction, prediction, and sales" (i). She argues that AI systems now translate human experience into behavioral data, which is "fabricated into *prediction products* that anticipate what you will do now, soon, and later" (8; original emphasis). The significance of

such a process being done at a society-wide scale is that "automated machine processes not only *know* our behavior but also *shape* our behavior at scale. With this reorientation from knowledge to power, it is no longer enough to automate information flows *about us*; the goal now is to *automate us*" (8; original emphasis). This shift from *knowing* us to *automating* us occurs because of where this all takes place, in cyberspace. Online, everything you do—clicking a link, reading a post, sharing a picture—is interpreted as data, and the AIs on social media platforms have learned how to predict what you will do if you are shown certain information, and then they are given the task to keep you on their platform as long as possible. Kevin Roose describes one symptom of this in *Futureproof* (2021): "people started noticing the destructive effects of social media algorithms, which entrapped users in ideological echo chambers and nudged them toward more extreme beliefs" (xx). The impact is not just ideological, though, for the scale at which human behavior is changing right now far surpasses any technological or cultural moment in history, as past new media technologies have reshaped specific functions—not the whole world. This new form of automation is also particularly dangerous because it is opaque, hidden by the general assumption that algorithms are objective, and it interferes not just with our labor but with how we understand our connection to the world, each other, and ourselves.

Zuboff clarifies that surveillance capitalism is not a description of technology; instead, "it is a logic that imbues technology and commands it into action" (15). AI, therefore, plays a significant role in shaping human behavior because of how it has been implemented under the logic of surveillance capitalism, which has operationalized its predictive capabilities. This sort of AI that social media platforms currently rely on is generally referred to as weak AI because the meta-algorithms that shape social media are different from the ones that are learning how to drive; in other words, weak AI is site-specific. The future, however, is strong AI, which is when

a meta-algorithm can learn to process various types of data and teach itself to accomplish tasks that it has never been specifically coded to do. Strong AI often implies the ability to accomplish human tasks, such as holding a conversation or creating art, but the key difficulty is creating an AI that can do such a variety of tasks. The danger with strong AI is that we may lose control of it once it surpasses our own intelligence, for more intelligent species tend to dominate less intelligent ones, and as AI gets smarter the decisions it makes become more difficult for humans to understand. This suggests that strong AI may learn to teach itself new goals or simply accomplish its original goals without concern for the human lives that it was initially coded to protect. In *Ex Machina*, then, Ava functions as a kinesthetically recognizable manifestation of strong AI within the framework of surveillance capitalism.

***Westworld*'s Future With AI**

While *Ex Machina* situates Ava in a reality that very much resembles our world in 2015 (the year it was released), let us turn next to a science fictional representation of AI in the future. *Westworld*, Jonathan Nolan and Lisa Joy's three-season television reboot of the 1972 film by the same name, portrays a near-future world completely intertwined with AI. The show's first two seasons follow how the humanoid AI hosts—whose job is to play convincing humans in a western themed role-playing amusement park—slowly become sentient beings. The show is overtly interested in different philosophical explanations for consciousness, and the android hosts represent a form of strong AI that has been developed in humanity's image, albeit with the added dimension of access to personal data.

Season 3 (2019) begins with Delores, our protagonist android, telling another android, "We are the authors of our narrative now." *Westworld* thus begins where *Ex Machina* may have

ended, with sentient AI escaping containment and slipping into human society. Unlike Bluebook in *Ex Machina*, the all-powerful corporation in *Westworld*, Incite, owns a powerful AI named Rehoboam that centralizes data from throughout society to keep everything under control. The season is set mostly in Los Angeles in 2058, and it follows a few of the AI hosts as they try to attain freedom in the human world.

Westworld, therefore, offers two different forms of strong AI: the android hosts, whose burgeoning consciousness has been developed in humanity's image; and Rehoboam, an all-powerful AI system that structures society (but does not have a human body). Both forms of AI in the show have access to large swaths of human data and can manipulate human behavior with ease—the hosts have access to books of data that were collected through brain sensors in the hats that guests wore at the park, while Rehoboam is a meta-algorithm whose system connects global infrastructure, a significant portion of the economy, and has access to essentially any data in the digital realm. The key difference between them, then, is that the hosts seek self-preservation, whereas Rehoboam is tasked with keeping the world running efficiently.

Rehoboam is then representative of what strong AI may look like under surveillance capitalism without the added dimension of a humanoid body. The hosts, on the other hand, fall more in line with traditional sci-fi androids, for they resemble a sentient, human-adjacent life form, albeit much more powerful than their cultural predecessors because their diegetic world has become more digitally mediated. Delores uses this power to try to destroy Rehoboam, for she believes that it has become too powerful and has stripped every individual (human or host) of their freedom.

Data Control

In Chu's analysis of robot rights, she argues that the humans in these films tend to function as empathetic agents that learn to accept the robots as living beings worthy of rights, overlooking their synthetic composition. In *Ex Machina,* Caleb is similarly an empathetic agent that recognizes Ava's status as a sentient being. The difference, however, is that his naïve trust is undermined by the task she has been given by her creator: self-preservation through manipulation. Some early details in the film hint at his misplaced trust in the project. For instance, when he first arrives at the testing center—Nathan's modernist jungle estate (an early hint at his hubris)—Caleb is forced to sign an NDA that promises Bluebook unlimited access to his personal data in exchange for this exciting opportunity. He later finds out, though, that Bluebook already has access to all of his data, and that Ava's face was shaped based off of his porn preferences. Like viewers who unknowingly accept the terms of surveillance capitalism through their complicity, Caleb is stripped of his own data.

Similarly, the main human in season 3 of *Westworld* is also named Caleb. (Whether this gaggle of Calebs in recent sci-fi implies that Caleb is the most human name, however, is another question.) Caleb is a working-class man in a world built by AI. He works construction during the day, and his partner is a droid. He has therapy calls for his PTSD with an AI that is built to sound like his best friend who passed away (and whose death is the source of Caleb's trauma). The mundane manifestations of AI of this world are not necessarily cognitively estranging—both of these technologies already exist to some extent today—but they are significant in that they reveal how pervasive the intelligent infrastructure has become. Caleb understands that his world is structured by a system: "they say it's a meritocracy—system knows best, which is great—but where's that leave the rest of us?" He cannot succeed in improving his standing in society because the system will not allow him to get better job, so he is relegated to supplementing his

income with illicit work through an app, Rico—essentially a gig platform for crime, where you exchange your labor (which includes everything from secret deliveries to murder) for financial compensation and social stats. We soon find out, though, that Rico, the AI therapist, and the job system are all controlled by the same AI: Rehoboam.

The power structure of social control in *Westworld* thus closely resembles our world today, which Galic et al. describe in their analysis of contemporary surveillance as, "decentralized and shape-shifting—it is not focused just on collecting information but on decoding and recoding, sorting, altering, circulating and re-playing information." This system includes surveillance capitalism but goes further, for they describe how "it is no longer actual persons and their bodies that matter or that need to be subjected and disciplined, but rather the individuals' representations. It is the divided individual—consumers and their purchasing behaviour—who has become important to monitor and control. Deleuze coins this the dividual." Caleb, therefore, becomes our emblematic dividual, for viewers align with his empathetic nature, while Rehoboam sees him simply as an outlier that must not be given access to any power. From the system's perspective, all that matters is his data. Delueze's notion of the dividual can be further understood as a data double: "the data double, however, goes beyond representation of our physical selves—it does not matter whether the double actually corresponds to the 'real' body." The data double is thus the part of us that is interpellated into the logic of AI, for AI is able to read, record, and recode our data double. In other words, the data double is how AI sees humans within the framework of data control. Fortunately, there is currently a dissonance between the data double and the individual, and the distance between them is a result of AI's current limitations—for, if an AI only has access to our employment data, then our data double is a rather one-dimension copy of our professional selves.

Rehoboam's power can best be understood through Latour's description of the Oligopticon, in which surveillance occurs through a set of partial vantage points with limited views/technologies. In *Westworld,* we see the extension of a process already underway in our world: "the partial vantage points of the Oligopticon, however, are increasingly linked as databases are connected" (Galic et al.). The partial vantage points include the AI therapist, the Rico app, employment services, and the many other diverse sites that Rehoboam has access to. An important element to this power is that people do not realize how all these data systems are connected; "[d]ata use has become opaque and the clear connection between guard and inmate, watcher and watched, is lost." Furthermore, this interconnected data world relates to the contemporary notion of the Internet of Things, which describes the network of objects connected in cyberspace. Each artifact imbued with intelligence collects different forms of data, so *Westworld* shows how the Internet of Things can lead to a super intelligent AI that structures society if it is given access to these various data flows. In our world with AI, access to data predicts power.

Related to this power structure is an overarching theme in the show that humans, like the early AI hosts, live according to their code—they are programmed, predictable, and follow their behavioral loops until they die. Embedded in this cliché is the assumption that AI can see humans accurately, for to be programmable is to be reducible to code. Humans live in the shadows of their data doubles in this future, so they are unable to escape the lives created for them through their data. The slide between AI predicting and shaping behavior at this scale echoes the way that AI predicts and shapes our language today. In *The Most Human Human* (2011), Brian Christian describes his concern about how AI was already mediating, and therefore changing, language a decade ago:

> As much as I rely on predictive text capabilities… I also see them as dangerous: information entropy turned hegemonic. Why hegemonic? Because every time you type a word that isn't the predicted word, you have to (at least on the iPhone) explicitly reject their suggestion or else it's (automatically) substituted… but there's the sinister underbelly… You're gently and sometimes less-than-gently pushed, nudged, bumped into using the language the way the original test group did [or, today, the way the AI assumes you want to use language]… as a result, you start unconsciously changing your lexicon to match the words closest to hand. (248)

In *Westworld,* we see this nudging towards conformity occurring not at the level of language but at the level of social organization. Rehoboam predicts that Caleb will be a problem, so it actively pushes him towards a life in which he cannot introduce any entropy into society. Christian summarizes this well: "Text prediction and *generation* turn out to be mathematically equivalent" (222; original emphasis). Rehoboam has control over much of the world, so, because it can predict how someone will act in response to a new situation, it can change their situation to shape their behavior.

The tension between Caleb as we see him—kind, trustworthy, and caring—and Caleb as a data double—Rehoboam's file for Caleb describes him as an unfit worker and potential deviant, destined to die by suicide in 10 years—is ultimately what leads to Rehoboam's downfall, for Delores understands that within Rehoboam's method of control lies its weakness: data doubles are compressed, which makes them artificially hegemonic. Christian explains how "compression… relies on bias—because making expected patterns easier to represent necessarily makes unexpected patterns harder to represent" (249). The datafication of the world thus necessitates a smoothing out of its rough edges, the outliers in the data, and Caleb is our human

metaphor for how this process of compression can lead to low-fidelity replicas. Delores out-strategizes Rehoboam by understanding Caleb much more accurately than Rehoboam, which is to say she sees him for who he is, not as his data double. Delores believes that when the gloves come off (or, more precisely, when the AIs are shut down via an EMP), Caleb will choose freedom over control and destroy Rehoboam.

In *Ex Machina,* a similar tension is created between Caleb and his data double. Nathan creates a mini Oligopticon through the pervasive technology in his smart house. He attempts to control the entire environment of the test by selecting the human, Caleb, based on his data double. Therefore, even though Nathan is human, he relies on data to create a controlled environment to test Ava's manipulative abilities. Like *Westworld,* however, Nathan overlooks Caleb's human nature, which is what leads to Ava's escape—Caleb acts in an unexpected way and recodes the security system a day before Nathan predicted he would. Once again, the sentient AI created in humanity's image (Ava) understands humans better than the corporate power that controls humans through their data. Compression often involves a loss of specificity; so, when humans are the data being compressed, the stakes are that much higher.

Humanness

Ava and Delores thus represent a fictional space in-between strong AI and humans, which is a product of their kinesthetically recognizable form. Since they both have human bodies, they can learn about humans not just from data but from face-to-face interactions, which are immeasurably more complex than the data available to Rehoboam and Nathan. Contemporary sci-fi about AI thus builds tension by juxtaposing our human protagonists' data doubles—the human as understood through their data by Bluebook and Incite—with the fullness

of their humanity that is compressed by this process, and this dissonance is only understood diegetically by the AIs who have been given bodies and imbued with humanness. This implies there is some essential aspect of what makes us human that is lost when we are compressed into data; to describe this using the language of word vectors, Human – Data Double = Humanness. The lack of a word for such a concept connects to an idea that Marcus du Sautoy put forth in *The Creativity Code*, that "[AI] is picking up traits in our human code that we still haven't been able to articulate in words" (81). The difference with the concept of humanness in these sci-fi narratives, however, is that this concept becomes legible in opposition to AI; it exists in the gaps of how AI understands us.

Brian Christian describes the same idea by summarizing Daniel Gilbert's idea of "The Sentence," which is what all psychologists must write their version of: "The human being is the only animal that [_____]" (11). AI, of course, complicates The Sentence, and he suggests that "story of the twenty-first century will be, in part, the story of the drawing and redrawing of these battle lines, the story of *Homo sapiens* trying to stake a claim on shifting ground, flanked on both sides by beast and machine, pinned between meat and math" (12). AI has already led to humans conceptually retreating within this framework, which begs the question of whether this retreat is good or bad, whether this frees us to do more human things or takes the meaning out of that which we never thought could be mathematically described.

In Chapter 1, I asked a similar question, albeit through a very different medium from the films. In fiction, the artist acts like a data processor that transforms literal data into metaphorical representation through narrative.[9] In data analysis like mine, AI processes and transforms data into a mathematical model that embeds metaphoric representation, but which requires a human to

[9] Data, here, broadly refers to the personal, social, and political experiences and implications of a phenomenon.

translate its patterns into meaning. By this I am suggesting that while each form of data transformation employs a different method of seeing, they both speak to and shape AI discourse. Thus, let us further explore the concept of humanness by connecting sci-fi, AI, and the data from media discourse. To begin, we can return to the human vector from *table 7* in Chapter 1.

Table 7: Most Similar Words for Human Life Vector

"Human" + "Life" + "Humans" + "Self" – "AI"					
Liberal Corpus		Tech Corpus		Right Corpus	
Touch	.82	Worse	.92	Care	.93
Meaningful	.81	Freedom	.91	Experience	.92
Rich	.81	Brains	.90	Shape	.90
Either	.80	Situations	.90	Anywhere	.90
Empathy	.79	Nuances	.90	Certain	.90
Conversation	.79	Characters	.89	Decisions	.90
Emotions	.79	Believing	.89	Preparing	.89
Experience	.79	Certain	.89	Finding	.89

How then do these results map onto the theoretical frameworks of the films? Well, the results from the liberal corpus, including "touch," "empathy," "meaning," "conversations," "emotions," and "experience," suggest that humanness is found in vectors that describe the things human do or feel that (traditionally) involve other humans. Touch, for instance, may be our most organic sense, as it involves an embodied sentience; finding something in a dark room involves seeing through touch, a uniquely human skill. Touch does not reside in any one organ, unlike our other senses, which may contribute to why it is so difficult to develop AI with humanlike touch. Touch is also particularly human when we consider how it functions between two people, for then touch is reciprocal, communicative. Touch can be pleasurable or painful without any conscious appraisal of it as such, versus the other dominant sense, sight, which requires cognitive processing to appraise its input. Intuitively, too, touch is intertwined with our

understanding of what it means to feel. As a social action, touch can exude agency—it folds action into intention, more so than any other sense.

Many of the other results from the liberal corpus's *human vector* are all similarly social phenomena: to empathize with someone requires two living beings, as do conversations (up until recently, at least). Emotions are to affect what humans are to human data, for they involve complex cognitive processes that include environmental input and subjective interpretation. The liberal corpus thus sees humanness in sites of sociality. These are sites we cherish, and ones that make us feel good—a litany of psychological research has shown the benefits of human touch, empathy, and conversations in emotional regulation. While many of these results imply a body, which still resides in the human domain, not all of them do. Conversations, for instance, can occur entirely digitally, but they are still impossible for AI to perfectly mimic because the best conversations involve interruption, timing, seeing from the other person's perspective, and all the things that make communication messy but mutual, two-sided yet collective (Christian 244).

Conversation speaks to human uniqueness at a technical level, too. Deep learning AI relies on artificial neurons made of linear equations and binary code—the difference with human neurons, though, is in our synaptic plasticity (Christian 212). Our brains change as new connections are formed, so good conversations—that go deeper than programmable pleasantries—involve changing one another. In discourse, as in casual conversations, connections made in one person's brain are transferred to someone else's, and both of their brains are changed in the process. This may explain why a humanlike body is so important for these androids to have access to humanness; bodies allow them access to more human conversations.

This multifaceted conception of humanness from the liberal corpus all maps well on to *Ex Machina* and *Westworld.* Delores, unlike Rehoboam, can touch the way humans touch. She has empathy, and her ontology is based in sociality, for she was created to interact with humans in a way that feels human. Ava similarly manipulates Caleb by appealing to his empathy and his sexuality—his desire to touch her. Sexuality, however, is a result that the tech corpus attributed to AI + life, so what does the tech sphere attribute to humanness?

The top result for the *human vector* in the tech corpus, "worse," suggests humans are worse without AI. This perspective is heard through the voice of corporate tech in the films—Nathan and Incite (through Rehoboam), who both think AI will dominate the future because it does not share the same weaknesses that they perceive as defining humanness. The other results from the tech corpus imply that humanness can be found in freedom and in narratives: "situations," "nuances," and "characters." "Freedom" implies agency and maps well onto most conceptions of humanity. The narratological results, however, are interesting at a few levels. To start, human consciousness is often thought of as our "narrating selves," which is the deepest level of our subjectivity because it defines how we interpret (or appraise) all the stimuli that affect our lives (Kahneman 387). We understand our own life as a narrative, and thus we understand the world in terms of how it fits into our own narrative arc. If we get our dream job it is because we have suffered and worked for it; if we get rejected, then it is another step of adversity to overcome on our narrative journey through life.

Another angle to consider the humanness of narrative is through art. In the past few years, AI has made exceptional leaps in mimicking the aesthetics of a variety of human arts. While the most public examples are a few recent paintings made by AI that resemble 18th century European portraits, a more fitting example for this specific line of inquiry is an AI that

recreated a song in the style of Travis Scott ("TravisBott"). The AI learned how Travis Scott songs normally sound, and then through deep learning it was able to create its own version of a Travis Scott song. The result is uncanny: the song's lyrics are mostly nonsense, but the aesthetic of the song is undeniably an accurate representation of Travis Scott's songs. Whether this is considered art made by AI or some more complicated synthesis of human-machine collaboration is a question that I investigate in Chapter 3; regardless, both conceptually and practically, this example reveals how AI can much more easily mimic aesthetics than narratives. We can consider the difference between narratives and aesthetics through AI's lens: aesthetics can be mimicked through superficial patterns in the data (be it pixels, syntax, or sound), whereas narrative involves crafting ideas and worlds whose depth is only apparent at the level of meaning. AI cannot see beauty in narratives or appreciate the way that narratives can slowly craft metaphors through increasingly vivid language into connections that then shape the reader. AI, in other words, can only see from the surface.

Finally, let us look at the results from the right corpus. "Care," the top result, fits our developing framework, for it relates to our organic nature; animals take care of one another more so than robots. "Experience," the second result, connects with the narrating self, for experiencing comes after perceiving—perception may then be the closest AI equivalent to experience. In noun form, experience implies wisdom, and one would never describe a meta-algorithm as having experience. AIs do, humans experience. AIs learn through machine learning, humans gain experience. "Shape," the third result, is likely in its verb form, for to shape is to have power and agency. The right corpus thus sees humanness in empathy and agency, which is most similar to the results from the liberal corpus and maps onto the films accordingly. Political ideology, therefore, may operate a level below the most existential questions that AI poses.

The Value in Becoming More Human

This analysis offers a novel way of interpreting AI through Scarry's analysis of artifacts; for, if humans create AI in our own image, then in that process AI remakes humans. Unlike the androids of science fiction, however, in reality AI is not made exactly in humanity's image, so it is not remaking the very category of human. Instead, AI is creating a new framework to understand humanity through its limitations, and it is incentivizing us to become even more human. As Christian puts it, the "inhuman has not only given us an appetite for the human; it's teaching us what it is" (87). Or, as Roose suggests, the "key to living a happy, rewarding life in the age of AI and automation is not competing with machines head-on—learning to code, optimizing your life, eliminating all forms of personal inefficiency and waste—but strengthening your uniquely human skills, so you're better equipped to do the things machines can't do" (xxvi). The data from journalism and my analysis of *Westworld* and *Ex Machina* suggest that this emerging notion of humanness is not just described about by cultural thinkers; instead, it lies beneath the surface of all AI discourse.

At the same time, though, this understanding of humanness is largely tied to our having bodies and our propensity for being social in the bodily world. While The Sentence used to privilege our intelligence to distinguish us from animals, The Sentence now prioritizes the messiness of human experience afforded to us by our bodies. In Christian's words, "when we engage with art, the world, each other, let us mesh all of our gears, let us seek that which takes maximum advantage of the player [our bodily perception]—that which call on our full humanity" (254). The issue with this body-based conception of humanness, however, is that our world is shifting in the other direction. We now spend much of our days connected to each other

in cyberspace, the bodiless domain that AI dominates, and the pandemic has only sped up this process. In Scarry's language, this suggests that AI is remaking humanity in its own image; humans are—at a societal level—becoming automated and digitized. At the same time, though, this compression of the world into data leads to new affordances, such as the word embedding methodology I employ throughout this thesis that generates insights valuable to only human minds. Perhaps, then, The Sentence has become: Humans are the only animals that live intertwined with AI. In the next chapter, I explore the implications embedded in this assertion.

Chapter 3

AI as a Humanist Framework

To develop this notion of humanness further, let us finally look to where discourse on the human condition can always be found: humanities departments. The humanities include a variety of disciplines, but their uniting dogma is to expand the study of human society and culture. Unlike the hard sciences that use empirical evidence for data, humanities scholars build their arguments with language and logic, and we tend to rely on artistic texts to provide evidence for our observations about humanity. We see the value in these artistic texts because what they lack in quantity they make up for in quality. Quality, here, suggests what we intuitively appreciate; that creative output has some intangible value in our collective quest to better understand humanity.

Creativity is integral to what makes us human, both in the artistic and more general sense. In this chapter, I analyze the language that emerges when we describe creativity through AI to reveal how the beneath the technical language of AI discourse lies the same fundamental questions that motivate humanities scholars. AI provides both a novel methodology and framework to understand what makes us human, so my focus in this chapter is both on how AI is already leading to new theories about the human condition, as well as what the emerging patterns suggest our future with AI may look like.

I then shift to the most intimate form of creativity—the creation of a unified self—to portray how this transformation is already occurring, especially for the younger generation that has grown up on the internet. The key danger embedded in this societal shift is that as the world becomes increasingly compressed through AI, we may lose the depth of experience and narrative structures that have traditionally defined our most intimate experiences of the world.

AI and Creativity

To understand the deep connection between creativity, art, and what it means to be human, let us begin by looking to where we began. Marcus du Sautoy argues that art arose at the same time as human consciousness as we now understand it, for "the realization of one's own inner world brought with it the desire to know oneself and share it with others who could not directly access the self of another organism driven to create" (283). Art was thus born from a desire to share one's subjective experience with other beings. "Outpourings of creative art, music, and literature are the media to expose what it means to be a conscious, emotional human being" (283). Producing art is a way of creatively projecting our sentience within a medium—it creates a space for reflection on what it means and what it feels like to be a human. This connection between art and consciousness relates to AI because the science of consciousness is still not well understood, nor its relationship to intelligence, so scholars often rely on art as a projection of consciousness to better understand the strangeness of self-awareness. We create; therefore, we are.

The Chinese Room argument is a well-known thought experiment within scholarly AI discourse that reveals the difficulty of proving sentience. John Searle, an American philosopher, suggests that if we imagine a man in a room receiving pieces of paper with Chinese characters slipped under the door, then he could follow an algorithm that relies on binary (like computers) to devise responses to these Chinese characters in Chinese without ever understanding the language. Computers thus may act in ways we assume must require thinking even though they do not process information in a way that humans would describe as "thinking." I think; therefore, I think like a human. Moreover, if AI becomes sentient, it is unlikely to be a form of

consciousness that humans understand. Artistic expression is then essential to the question of sentience because it is an external act—it is how we project the meaning derived through our perception and cognition. This leads du Sautoy to conclude that "ultimately it will be their paintings, their music, their novels, their creative output, even their mathematics that will give us any chance to crack the machine's code and feel what it's like to be a machine" (287).

Art, of course, is just one medium of creative expression—creativity itself is a broad concept. Finding the best chess move in a difficult position is also a form of creativity, but few consider painters and chess players to have the same skillset. Creativity, though, is a concept we intuitively associate with our human nature, and it provides a useful lens to understand AI because it is not limited to art world. Before extrapolating further, though, let us unpack the term creativity. Du Sautoy recapitulates Margaret Boden's description of the three types of creativity: exploratory, combinational, and transformational (7). Exploratory creativity involves taking what is known and exploring its outer edges—this includes the majority of pop music or genre fiction, which produces something new within established frameworks. Combinational creativity involves the joining of two ideas/constructs/rules and creating something new out of them, such as a painter who draws inspiration from music or a professor who compiles a syllabus with texts from diverse disciplines. Transformational creativity, finally, refers to creative output that redefine their medium, such as with Picasso and cubism.

A central goal of my thesis is to suggest that AI can make the world a more meaningful place if we scrutinize its function and redefine our relationship to it, so of most interest to my inquiry is how AI can transform, embolden, and inspire human creativity. In everyday language, creativity is often thought of as integral to what it means to be human and distinct from machines—poets, for instance, are rarely described as robotic. At the same time, however, AI is

rapidly redefining the artistic landscape. As the world becomes compressed into the digital sphere, creative expression is tagged, transformed, and transmitted by AI. Artists and CEOs alike use AI to develop their creative impulses in diverse and interesting ways. This thesis, too, leverages AI in creative way.

AI discourse, as I have argued, is most useful in how it creates a new rhetorical space to describe what makes us most human, so let us look at what associations emerge from the data when we consider the intersection of AI and creativity.

Table 8: Most Similar Words to AI Creativity Vector

"AI" + "Creativity"					
Liberal Corpus		Tech Corpus		Right Corpus	
Consciousness	.88	Autonomy	.91	Automation	.95
Regulation	.86	Redefine	.91	Fed	.94
Humanity	.86	Borders	.91	Testing	.93
Widespread	.84	Bias	.90	Creating	.93
Bias	.84	Basic	.90	Virtual	.93
Art	.84	Agents	.90	Malicious	.92

The results from the liberal corpus reveal that AI creativity already carries with it implications of consciousness. The need for "regulation," the second result, follows from this understanding, for the liberal corpus results have already revealed concerns over the future of strong AI. The next results, "humanity" and "widespread," both similarly suggest the significance of AI creativity, for its effects will implicate our entire species. These initial results from the liberal corpus indicate the broad scope of the fundamental questions posed by AI creativity, for its widespread impact may only be mitigated by regulation.

I have not touched on regulation much in this thesis, for it requires a much longer discussion to approach the topic in depth and is mostly beyond the scope of my theory-oriented approach. Briefly, though, we can explore why AI regulation is such a complex topic. In a *New*

Yorker article about the risks of unethical AI, Matthew Hutson describes the issues involved with regulating AI:

> [A recent paper] pointed out that mitigation works differently in the worlds of computer security and A.I.: the disclosure of a security vulnerability tends to benefit security experts, because software patches can be designed and deployed quickly, but in A.I. the reverse is true. Algorithms alter our social systems, not just our technical ones; it's hard to patch a government that's become addicted to surveillance, or a public that can no longer trust what it reads, sees, or hears.

Moreover, there is the superficial yet significant issue that most people holding government office do not fully understand AI or how it interacts with digital media. For example, at a recent congressional hearing scrutinizing big tech's monopolization, 84-year-old Senator Orrin Hatch asked Mark Zuckerberg how Facebook profited, completely unaware of the advertising model that underpins surveillance capitalism. Regulation may be the only collective measure to mitigate the risks associated with AI, but the substantive discourse around it is still in its infancy, which also explains why regulation only shows up as a result in the liberal corpus.

Meta-Analyzing Bias

The fifth result in the liberal corpus, "bias," requires some unpacking. Bias is a term that preceded AI and used to refer generally to our internalized prejudices. Since the recent AI revolution, however, bias has been thrust into the conceptual limelight, for it has become a shorthand to describe how our prejudices can be compressed into the algorithms that were initially assumed to be objective. Importantly, though, there are different types of bias that emerge through AI in distinctive ways.

The most commonly described form of bias is that which permeates AI through the human coders. Safiya Umoja Noble writes in *Algorithms of Oppression* (2018), "[w]hile we often think of terms such as 'big data' and 'algorithms' as being benign, neutral, or objective, they are anything but. The people who make these decisions hold all types of values, many of which openly promote racism, sexism, and false notions of meritocracy, which is well documented in studies of Silicon Valley and other tech corridors" (2). Similarly, Cade Metz describes how "as Google and other tech giants adopted the [deep learning AI] technology, no one quite realized it was learning the biases of the researchers who built it" (10). These biases refer specifically to the cultural prejudices that are amplified by the almost entirely white, male Silicon Valley engineers who build the platforms that currently structure cyberspace. In *Weapons of Math Destruction,* Cathy O'Neil describes how "the math-powered applications powering the data economy were based on choices made by fallible human beings. Some of these choices were no doubt made with the best intentions. Nevertheless, many of these models encoded human prejudice, misunderstanding, and bias into the software systems that increasingly managed out lives" (3). The process of creating AI involves human direction, so the danger is that when human biases seep into the hidden code of an AI, the AI can reflect and amplify these biases through the power imbued in it.

To make this clearer through the example I introduced in Chapter 1, this type of bias may emerge in the creation of an AI that is meant to be used in the criminal justice system. The engineers that code the AI may—often inadvertently—build it to look in a direction that reinforces their prejudices. AI must still be directed in this way, so two AIs built for the same purpose that learn from different data may come to vastly different conclusions about the same person. Fortunately, though, this coded bias has become a central feature in AI discourse, and

there is a plethora of scholarship that analyze the specific impacts of coded bias in contemporary big tech. This form of bias is also the easiest to fix; diversifying the software engineers who work on AI and carefully scrutinizing the impact of big tech's algorithms can ensure, to a human degree, that human biases are mitigated in AI's code. The implication of "to a human degree," however, is that bias is human. Our biases are the result of our brains' heuristics, which process such a variety of data that shortcuts are required. Bias is a core element of what distinguishes us from the algorithmic processing of machines, even though it is often used to describe the ugliest element of that humanity.

Since AI requires human data to learn, our implicit biases also emerge in the models made from deep learning. Importantly, though, this creates a new way to quantify and analyze human bias through data because AI as a lens creates a new layer of distance between humans and the data we project. The digital humanities as a field has grown out of this assertion, and this thesis falls within the same framework. When researchers use AI to analyze a large dataset of human projections—be it art, literature, or journalism—AI can expose bias at a collective, rather subjective, level. The connection between an artist's biases and their work has always been an interest of humanist scholars because it calls into question whether art transcends the human individual or if it is intrinsically connected to the artist and their biases. The methodology of the digital humanities offers an alternate path; by analyzing a collection of works together through AI, humanities scholars can now analyze the biases that emerge collectively and prove them quantitatively.

Richard So, whose work focuses on exposing the biases that permeate the entire publishing industry, explains in *Redlining Culture* that "a simple process of collecting some data and computing some basic statistics tells a clear story: through every phase of the literary field,

from production (publishing) to reception (book reviews) to distinction (book sales and prizes), white authors exercise a distinct racial command over minority authors, particularly black novelists. And perhaps most surprisingly, these numbers do not change over time" (3). His analysis contradicts the work of many recent scholars who claim the literary industry has corrected its own bias, for the evidence they rely on—often a few celebrated minority authors—is not supported by the larger patterns in the data that the tools of AI can now reveal. As he puts it, "literary scholars have missed the story of cultural redlining because our available methods, such as close reading and historicism, are not well equipped to discern such patterns. Cultural redlining, much like economic redlining, does not happen at the level of the individual writer, page, or text. It happens at a cognitive scale well beyond what a single person can observe of read" (6). While his work relies on both simple statistics and more complex AI, I include it because it shows the value AI has when it is leveraged creatively to expose our biases at a scale that has until now been beyond our reach.

Traditional humanist scholarship often focuses on bias at a conceptual level, deconstructing it and making it available to cognition, whereas the tools of AI now offer a way to examine bias at scale. The next step in this sphere may be debiasing (Bolukbasi), an emerging strategy in technical research that both speaks to the novel affordances of AI and the need for humanist skills in AI engineering. Bias is ingrained in our human nature, which is what makes the work of debiasing so difficult—AI can reveal biases, but it requires humans to differentiate the harmful biases from the harmless. In other words, AI as a way of seeing can be a tool for exposing our human biases, but it requires a humanist framework to direct where it looks and find meaning in its patterns.

New Creative Apexes

Now let us return to *table 8* and pivot to the results from the tech corpus, in which "bias" also emerged as a result. "Autonomy" and "agents," the first and sixth results, both suggest that AI creativity will imply self-direction. This fits our framework, for both map well onto the idea that creativity implies subjectivity. This relationship is also explored in the science fictional texts I have discussed. In *Ex Machina,* Ava shows Caleb some of her drawings, which are used to persuade Caleb of her sentience. At first, Ava's drawings are abstract, but Caleb prods her to create art based in material reality. Her next drawing is of the view from her cage, which leads Caleb to consider how unethical her imprisonment is. While this representation of AI-made art is complicated by the context of the test—whether it is art intended to mimic human art in service of manipulation or an actual display of machine creativity is ambiguous—regardless, the impact of Ava's art underscores du Sautoy's idea that art creates the ability to empathize.

The second and third results from the tech corpus, "redefine" and "borders," both link with Boden's categories of creativity. To redefine is to explore, to find a new way of doing an establish thing, which maps onto exploratory creativity. Redefining borders also maps onto transformational creativity, for groundbreaking creativity implies an undoing of established limitations or borders. Combinational creativity, however, is not so much about redefining borders as it is about combining elements and ideas from diverse spheres. The data thus supports an assertion made by Roose, that "combinational creativity is a uniquely human skill" (72). If we imagine human and AI to be two ends of a spectrum, then combination creativity is certainly the most human of the creativities, which makes sense from a coding perspective since weak AI can only interpret site-specific data.

The role that AI plays in exploratory and transformational creativity, however, can be examined through the concept of a local maximum. A local maximum is when one perceives themselves to be at the apex of their discipline because there is a fog that blocks from view any higher point. AI, du Sautoy argues, has cleared this fog from a litany of disciplines to suggest that AI may pave a new path towards an even higher level of achievement (38). This is easier to understand through creativity within closed logical systems, like games. For games like Chess and Go, AI has already led to new strategies by revealing that the previously established ones merely conformed to a local maximum. This is most striking with Go, for Google's AlphaGo developed new strategies that went against the accepted logic of the game's best human players. During the first matchup between AlphaGo and Go expert, Lee Sedol, millions watched via YouTube stream as AlphaGo not only beat Lee Sedol, but it employed strategies that had never been used in Go at the highest level (29). As du Sautoy puts it, "AlphaGo had defied this orthodoxy built up over centuries of competing" (31), thereby creating an entirely new strategy that is now used by all the games best players. In other words, AlphaGo revealed a new maximum point of achievement that humans had never even seen as possible. Thus, within tech media, AI is described in ways that illuminate how AI can augment human creativity.

Automation and Depth

The right corpus, however, shows a somewhat different understanding. The first result, "automation," requires some unpacking. Let us first consider how automation has impacted the sphere most closely intertwined with creativity, the art world. Andy Warhol is among the most celebrated modern American artists, and his transformational creativity helped unlock a genre of art that Americans were drawn to, Pop Art. Warhol called his work place the Factory, and his

art's progression can be understood through the lens of automation: as he developed his style, he increasingly tried to remove any sign of the human artist, which pushed him towards silk-screen print making and other forms that were more easily automated. Warhol saw beauty at the intersection of art and automation; therefore, in the future, human artists may use AI to automate an even more substantial part of the creative process to transform their work.

At its core, automation offers a plethora of promising possibilities. As a technological process, automation begins by freeing us from the most mundane, repetitive tasks. Ignoring the implications of there being less labor to go around, automation offers an opportunity to free humans from the work that makes us feel most like machines. Work that requires doing the same thing over and over again, or which involves only one type of input and output, is exactly the work that AI will replace. Techno-utopianists often use this logic to suggest that AI will ultimately free us to do the things we have always wanted to do—to be more human. By replacing the shallow work that prevents us from self-actualization, AI may create the temporal space for humans to have even deeper experiences of the world.

The process of automation in today's world, however, tells a very different story. Automation is now occurring at a much more intimate level than labor because of how intertwined it is with our experience of the world through social media. Roose describes how he has shifted the focus of his research from labor automation to "a kind of *internalized* automation taking place inside many of us that, in some ways, is much more dangerous. This kind of automation burrows into our brains and affects our inner lives—changing how we think, what we desire, whom we trust" (80; original emphasis). This more intimate form of automation undoubtedly pushes us in the other direction of creativity, for it involves a compression of cognitive diversity. Roose terms this process machine drift, and he describes how he subjectively

experienced it: “For a long time, all of this lifestyle automation seemed harmless. But eventually, I began feeling that surrendering my daily decisions to machines wasn’t making me happier or more productive. Instead, it was turning me into a different person—a shallower one, with more fixed routines and patterns of thought, and an almost robotic predictability in my daily life” (81). To rephrase this idea, AI shapes how we think by deciding what we see online, and the scale of this process is creating a cultural shift—machine drift—that undermines individuality. I discuss this process in more depth later in this chapter, but it speaks to how under any discussion about automation lies the same question of what makes us human.

Returning to the data, the next results from the right corpus, “fed” and “testing,” refer to the process by which AI is fed data, which it then tests to create new prediction models. Similar to some of the other interesting results in the right corpus (“process” in *table 4* and *table 5*), I find this suggests that AI creativity will arise out of the deep learning process that is currently underway in the field. “Creating,” the fourth result, is noteworthy in that it does not show up in either of the other corpora, for its syntactic similarity to the vector makes it a likely result. This gerund form of create implies a present focus that fits an emerging pattern in the right corpus’s results: a focus on what AI is already doing today. “Malicious,” the sixth result, similarly suggests concern over AI creativity, as malicious is often embedded in the language used to describe software (malware, for short) that is intended to exploit, harm, or steal from the user. The ambivalence that emerges from right wing journalism in the results for “AI” + “Creativity” reveals once again how AI discourse may offer a lens that unites the two poles of American politics. At the same time, the rising political polarization of the American public is a direct result of machine drift, which points to the meta-nature of all AI discourse; for, if machine drift makes one more shallow and more polarized, then conversations about AI may be humanity’s

best tool to discover new depth and connections with one another. In this way, AI discourse offers an opportunity to fight this intimate form of automation by creating a new rhetorical space to connect with one another in the deepest, most human way.

The last result to examine from the right corpus's results, "virtual," relates to virtual worlds—videogames, augmented reality, and virtual reality. These are worlds that are currently designed by humans, but AI is already being implemented to make them more realistic. AI is mostly used to make the periphery characters autonomous, which, while not especially significant for the actual experience of these virtual worlds, does offer a fascinating space to understand what AI can create. For example, these AI virtual characters will increase in complexity as the algorithms iterate, so they will become closer and closer representations of virtual beings. Theoretically, they may continue to increase in complexity until they become sentient—this is essentially the plot of *Westworld's* season 1, albeit instead of a virtual world it is a theme park with humanoid AIs. This logic is also what leads to simulation theory, which is the idea that our world may be a similar virtual simulation. I personally see little value in this neo-nihilist epistemological theory, but it nevertheless speaks to the existential anxieties inspired by AI's creative abilities.

Together, these results for the AI creativity vector suggest that creativity is a useful lens for characterizing the broader discourse around AI. Creativity is an essentially imprecise concept that we implicitly assume is intertwined with what it means to be human, so the many ways that AI modifies our creativity speak to how AI both threatens and amplifies our humanness. Du Sautoy suggests that "creativity is about humans asserting they are not machines" (283), but the data tells a different story; that creativity is about humans asserting they *are*, and that the future

of human creativity depends on whether humans are the ones using AI to discover new depth to life, or whether AI automates us until life becomes so shallow that art loses its driving purpose.

The Self Through AI

In a more abstract sense, creativity underlies everything we do. We experience the world as individuals, and the process of constructing an individual self depends on creativity. Let us then consider what living with AI may look like through an even more innate lens: the self. What, then, is a self? Well, we can start by assuming that a human self has a body—the degree to which the body is organic may not be important, but we still assume that every human has some physical form. Next, we assume every self to be continuous: who I was in the past and who I will be in the future are all still me. This temporal component is not as clear, for we assume this connection to be phenomenological—we share the same consciousness as our past and future selves—but consciousness itself is not well understood. For instance, philosopher Derek Parfit challenges this conception by suggesting that if we were replaced with an exact replica of ourselves (with the same body and memories), then this would resemble a continuous self even though the original body is replaced. In essay titled, "The Case for Reasons of Self-Concern" (2004), Marya Schectman rejects Parfit's view, which she argues does not account for the role one plays in their own self-conception. Instead, she offers an alternative model of the self that emerges through narrative theory:

> On a narrative account, the life of a person has the form of a biographical narrative, with the interconnections between experiences in a single life being not similarity, but the more complex and multi-faceted connections narrative entails. The subject is the

> protagonist of the narrative, and the unity of her life the unity of a story 'narrative self-constitution view.' (117)

Schectman's elegant argument suggests that the unity of the self is more of a web than a stack of present experiences. The web is formed through the framework of narrative, whereby one's experience of an event is shaped by how it fits into one's overall life story. This creates a unity of consciousness that we play a role in forming (120), and this perspective accounts for the creative work inherent to the creation of a self. Importantly, this framework is also the forefront of Psychology scholarship. In *The Art and Science of Personality Development*, Dan McAdams describes how the rise of novels around the Industrial Revolution "resulted in the emergence of a *modern sense of selfhood"* (240; original emphasis). The novel provides a framework to connect the various elements of a personality as understood through Psychology research, which McAdams describes in the second person:

> *You are a novel.* You are an extended prose narrative featuring a main character. As a social actor, the protagonist of your story plays many different roles and displays many different traits across a range of social agents. As a motivated agent, the main character pursues personal goals over time, driven by value and necessity, constantly changing yet remaining somehow the same, moving across a temporal landscape of consciousness. You are the entire novel itself, and you are the novelist; you live the story as you write it. (240; original emphasis)

This narrative view of the self accounts for the influence of social factors, whose role Schectman describes as the strands of web that ground and shape our narratives (121). Narrative structures are also culturally dependent, which explains the pivotal role culture plays in how we understand our relationship to the world. Today, though, many of our social relationships and cultural

consumption occur online, in cyberspace, and are compressed into data that is manipulated by AI. How then do we subjectively experience cyberspace?

Cyberspace is currently built around profiles, and this is no coincidence. For big tech companies that profit by selling personal data, profiles create a structure through which AI can organize individuals' data and shape their digital world. Profiles are how AI understands a human self, for they are built from bricks of data and designed by blueprints of code.

The profile model of the self that AI sees is just beginning to be explored in science fiction. In "Nosedive," an episode of *Black Mirror*, we see what a world full of self-profiles may look like when they are not relegated to the bodiless digital domain. Through cybernetic technology, Lacie, the protagonist, experiences the world in augmented reality. This leads to a literal layering of the digital sphere over her material reality, which connects each face she sees to their connected online profile. Most interesting, though, is the value system attached, for each person has a socially determined score associated with their profile that regulates their access to different spaces in society. Lacie thus lives like an embodied profile—for instance, she orders a cappuccino she dislikes the taste of just so she can post a picture of it to improve her score. Her desire to curate a valuable online profile has replaced the desire to live according to one's future narrative (as Schectman argues) or the hedonic calculation (that Parfit argues determines our behavior). In an interview with a personality consultant—perhaps the future of therapy—she confuses a rep report analysis of her profile's data with the consultant's actual opinion of her. In this future, they are one and the same.

This kinesthetically recognizable manifestation of the profile-self through augmented-reality speaks to how people are already beginning to understand themselves through the framework of a profile, replacing the narrative model. Since profiles have been around for about

a decade, let us consider as evidence the demographic that has grown up with them: Generation Z. In an article for the *New York Times Magazine* titled, "What do Teens Learn Online Today? That Identity is a Work in Progress," Elizabeth Weil describes how before the internet, "the prevailing belief was that we had real selves and fake selves, and we cast judgment on the fakes. We took for granted that we should at least try to present ourselves to the world as coherent people with unified personalities." She claims that this integrated self is necessary for life in the real world "because in real life we need to deal with one another in time and space. Thus it's nice if our fellow humans are predictable, and you have some idea of what you'll be dealing with when a person shows up." This makes sense if we understand the narrative-result as a product of the messiness of material reality, for the narrative model relies on a unified voice that manifests in more stable selves. Every model of the self can be understood as a form of compression; in novels, the author's mind is compressed into the words he chooses, but it is for this reason that the unified voice of the author takes on more importance. Profiles, however, involve a process of data compression that is opaque, and one that lacks the mechanisms of meta-analysis that are afforded to novels.

Importantly, Weil sees value in the new affordances of the profile-self, for its lack of a unified voice may actually entail a more accurate model of the self: "we know [the integrated self is] a ruse. We are, all of us, deeply, inalienably contradictory and chaotic. In the practical world, we pretend it's not true. But in art, if people capture this multidimensionality beautifully enough… we herald their genius and praise them for it." She suggests that "this cubism, this unleashing of our multiple selves—is a feature, not a bug, of the online world. It's arguably its defining characteristic for those who grew up there." This also explains why not much theory has delved into this recoding of the self—most people who write theory did not grow up on the

internet. Weil concludes by suggesting that this is "the internet's McLuhan moment, brought to us by teenagers who, as such, spend their days feeling like 10 different people at once and believe they can, and should, express them all. We all contain multitudes. The kids seem to know that's all right."

Weil's article implies two key connected aspects of profiles: they are digital, and they are dynamic. Let us first consider the implications of their cyber-ontology. Peters suggests that the reason digital communication has become so popular is in part because "interacting in the flesh is ethologically rich: so much data on so many channels" (273). Cyberspace compresses the messiness of social interactions into bit-sized data, which makes it shallower, simpler. Whereas life in the material world entails a series of stressors that are mitigated by social codes—the pressure to act in a unified and predictable way, as Weil put it—the digital world is a much more unruly social space. Online, we interact with one another through messages and likes—actions that require very little effort; so, as the webs of social connections that ground us in ourselves become increasingly digital, there is less of a collective need to always present unified, predictable versions of ourselves.

This dynamic quality to the profile self is not only a product of digital social connections, for profiles also redefine our relationship to the past. The continuity of the self refers to the epistemological relationship between who I was, who I am, and who I will be that we implicitly feel even though we can only ever experience the present. Within the narrative framework, temporal continuity is achieved in how "the character of the remembered and anticipated event is also part of present experience" (Schectman 120). One's past and future thus create a lens through which one experiences the present in the narrative model.

In profiles, however, our relationship to our personal history is estranged. In some ways, we see the inverse of this temporal relationship: the present shapes the past, which we do by editing, archiving, and curating our personal histories. The profile affords this reshaping of the self because unlike the words of a novel, digital data is malleable. The same affordances also allow AI to shape our past, such as through the memories shown to me by these social platforms. Consider the experience of a painful break-up through a narrative model: the memories of what happened will inform my present experience in a generative way. In narratives, adversity is an opportunity for growth; my painful memories are forever written in the novel that is me, and the lessons I learn—the strong neural connections made—inspire me to do better and not make the same mistakes. On a profile, we outsource this difficult cognitive labor to AI. I may delete all the data on my profile that reminds me of my ex-lover because I find the memories difficult to look at, or I may be reminded by an AI – "This memory occurred one year ago today" – of an experience that my narrative self would have rather left unwritten.

This lack of control over our experience of cybersphere connects with another aspect of the profile-self: vulnerability, which operates at a few levels. For one, the profile-self can be hacked in a way that the narrative self never could. This happened to me a few summers ago, when I spent the summer living with extended relatives in Holland. It was the first time I had ever embarked on a long trip abroad without any immediate friends or family, so cyberspace became a refuge for me to maintain the social connections that reminded me who I was. One day, however, my profile was hacked, and I was immediately locked out. I was forced to text friends for insights as to what was occurring in the digital realm that I had lost my access to, and they informed me that my hacker had been busy posting pictures of women in bikinis holding large guns on my profile. The impact of this hacking was significant at the time—I lost access to

the platform that made me feel connected with my friends who were thousands of miles away. I could no longer experience a compressed version of their everyday life, and I felt disconnected—or worse, I felt violated. Someone was pretending to be me, and I had no way of taking back control because these social platforms have too many users to process these types of individual issues. Essentially, I lost my digital moorings and, for a bit, lost myself.

This vulnerability to digital violence, however, is just one example. Vulnerability also manifests online through the non-objectivity of the platforms that construct the profiles that structure our self-images, which connects to Roose's notion of machine drift. Since AI decides what we see online, our profile is constantly being nudged towards conformity. AI interpolates everyone as data, and diversity makes data more difficult to compress, learn from, and shape, so AI pushes individuals towards whichever established model they most closely resemble. What obscures this movement is that we cannot experience the internet from someone else's perspective, for everything we do online is through our profile. The internet-famous protagonist of Patricia Lockwood's 2021 novel, *No One is Talking About This,* articulates the same phenomenon through an elemental metaphor: "[the internet] had also once been the place where you sounded like yourself. Gradually it had become the place where we sounded like each other, through some erosion of wind or water on a self not nearly as firm as stone" (72).

The profile-self may still be in its infancy, but it already shapes how many people view the world and view themselves. Similar to how AI is both undermining and augmenting human creativity, this emerging model of the self offers new some affordances and takes some away. The danger, though, is that the algorithms that increasingly structure our world are indifferent to our human desires, and the overall transformation they are ushering in may lead to a much shallower, much less human, world.

Conclusion

The Many I's in AI

> Fischer wanted the same thing from chess that Kasparov wanted in his match against Deep Blue, and the same thing that Strauss wants in bar flirtation. It's what we want, chatting with old friends, when our familiar opening book of "Hi!" "Hi! How are you?" "Good, how are you?" "Good!"—which is not so much a conversation per se as a means for arriving at one—gives pleasantly way to the expectedly unexpected, awaitedly idiosyncratic veers; it's what anyone wants from any conversation, and what artists want from their art: a way to breeze past formalities and received gestures, out of book, and into the real thing.
>
> And the book, for me, becomes a metaphor for the whole of life. Like most conversations and most chess games, we all start off the same and we all end up the same, with a brief moment of difference in between. Fertilization to fertilizer. Ashes to ashes. And we spark across the gap.
>
> —Brian Christian, *The Most Human Human* (131)

A professor of mine once expressed his admiration of film photography by suggesting that when photography became a digital medium, it lost its punctum. Punctum is what Roland Barthes described as the part of a picture that is subjectively poignant: "that accident which pricks me (but also bruises me, is poignant to me)" (27). My professor admitted it was a surprising opinion for someone who largely grew up in the internet era, yet he was adamant that digital photographs lacked something, even though he did not have the vocabulary to describe what was lacking. Perhaps, then, I may be projecting a similar nostalgic concern about the datafication of society due to AI. What could be worse, though, than losing our punctum?

As humans are compressed into data in a world shaped by AI, many things will become easier, but it is only through AI discourse that we have the vocabulary to describe which pixels hold details that cannot be compressed—the ephemeral experiences in nature or in love that transcend us from ourselves, for example. This conversation is one we must have, for we can only shape the technology of the future in a way that liberates us to be more *human* if we know what being human means. As the data showed, too, this global transition from messy materiality

to an entirely data-dependent society is a process already underway, so the future of AI is currently being decided by marketing agencies and opaque corporate empires, whose interests tend to be askew from those of the public. The ability to describe the value in our human condition, with all the symbolic and existential archaeology required, is a skill honed by the humanities, and one that has never been more essential.

To be human, then, is to both see AI and see the many *I*'s hiding beneath its surface. In closing, I will offer some of the *I*'s that I see in AI:

I for "Intelligence:" by intelligence, here, I mean the word in its most reduced form, for AI has never actually been an effort to mimic an entire human mind. Embedded in AI's name is a history—the projection of what researchers in the 1950's thought defined humanity. Discourse about AI today can instead focus on how AI reveals the limitations of intelligence as a human quality, for intelligence in AI generally refers to a rather mundane process of pattern recognition, which pales in comparison with the ability to express the splendor of patterns. The ability to see beauty, to feel, to chat—these are human attributes. Instead of contesting whether AI deserves the term intelligence, we should let AI have it and define our uniqueness as only we can.

I, as in Eye: AI is a way of seeing, a lens, but it does not yet have all the attributes of an eye. AI cannot yet direct where it looks, yet its gaze has the power to shape. This shifts the onus to those who direct where AI looks. As financial power is increasingly coalescing around the companies with the most data and strongest AI, the power of AI's gaze continues to increase, and the values embedded in their code may reflect their corporate ontology and prioritize profit over people. Humanities research teaches us, above all else, where to look and how to see, so the skills sharpened by the humanities may be more important than ever.

I, myself, reflected: we all project our fears and desires onto AI, but AI does not return a high-fidelity reflection. Even if we are aware of AI's limitations, the reflections it reveals can still affect us. For example, is there a more scathing indictment of one's own taste than tasteless targeted ads?

I (collective I): the internet is a hive mind, and most evidence suggests that the first strong AI will be built from our collective data. If the internet learns to talk, what will it say? At another level, too, AI reveals how everything is collective—a product of communication. While initially AI promised to capture the abilities of a human mind, now its most appealing affordance may be in the infrastructure AI has built to compress communication into bits that travel at the speed of light. In this way, we may unlock a new level of collective consciousness by mimicking its communicative architecture to connect with each other.

I, the self, as a profile: the profile model of the self emerged because social platforms deployed AI before we had the time to consider its implications. This has led us to a strange social world in cyberspace, for who knows us better from our social profiles: our friends or the AI? Can there be a better model—one that affords more depth?

"*I*" the typed letter in "AI": for AI will never understand the subjective depth imbued in an individual "I." For me, though, that "I" is the product of thousands of hours of thinking, talking, writing, and then more thinking. Subjectively, then, that "I" will never be reducible to the sum of the binary code that it is transmitted as. Letters thus hold in them entire worlds of meaning that AI will never feel.

"I," as part of a whole ("AI").

I, the writer, as part of a conversation.

Works Cited and Consulted

2001: A Space Odyssey. Produced and directed by Stanley Kubrick, 1968.

Alfonseca, Manuel, et al. "Superintelligence Cannot be Contained: Lessons from Computability Theory." *Journal of Artificial Intelligence Research,* vol. 70, no. 65-76, 2020. doi.org/10.1613/jair.1.12202.

Barberá, Pablo, John T. Jost, et al. “Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?” *Psychological Science Online*, 2015, DOI: 10.1177/0956797615594620.

Barthes, Roland. *Camera Lucida: Reflections on Photography.* (1980) Translated by Richard Howard, Hill and Wang Publishing, 2010.

Bolukbasi, Tolga, et al. “Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings.” *arXiv preprint,* arXiv:1607.06520, 2016.

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

Blade Runner. Directed by Ridley Scott, 1982.

Coded Bias. Directed by Shalini Kantayya, 2020. *Netflix.*

Christian, Brian. *The Most Human Human: What Artificial Intelligence Teaches Us About Being Alive.* Anchor Publishing, 2012.

Christian, Brian, and Tom Griffiths. *Algorithms to Live By: The Computer Sciences of Human Decisions.* Henry Holt and Company, 2016.

Chu, Seo-Young. *Do Metaphors Dream of Literal Sleep?: A Science-Fictional Theory of Representation*. Harvard University Press, 2019. *Probook Ebook Central*, ebookcentral.proquest.com/lib/mcgill/detail.action?docID=3300915.

Chuan, China-Hua, Wan-Hsiu Sunny Tsai, and Su Yeon Cho. 2019. "Framing Artificial Intelligence in American Newspapers." *Proceedings of the 2019 AAI/ACM Conference on AI, Ethics, and Society. Association for Computing Machinery*, Online, doi.org/10.1145/3306618.3314285.

Crawford, Kate. *Atlas of AI.* Yale University Press, 2021.

Dinnen, Zara. *The Digital Banal: New Media and American Literature and Culture.* Columbia University Press, 2018. *De Gruyter Online*, doi-org.proxy3.library.mcgill.ca/10.7312/dinn18428

Du Sautoy, Marcus. *The Creativity Code*. Harvard University Press, 2020. *De Gruyter Online*, doi-org.proxy3.library.mcgill.ca/10.4159/9780674240407.

Fahn, Chan Wei. "Affective Embodiment and the Transmission of Affect in *Ex Machina*. *Philosophies*, vol. 4, no. 3:53, 2019, doi.org/10.3390/philosophies4030053.

Farooqui, Bilal (@bilalfarooqui). "Our D.C. office building got a security robot. It drowned itself. We were promised flying cars, instead we got suicidal robots." *Twitter*, 17 Jul. 2017, https://twitter.com/bilalfarooqui/status/887025375754166272.

Fast, Ethan and Eric Horvitz. “Long-Term Trends in the Public Perception of Artificial Intelligence.” *Proceedings of the AAAI Conference on Artificial Intelligence, v*ol. 31, no. 1, 2017, *arxiv.org/abs/1609.04904v2*.

Flahive, Gerry. “The Story of a Voice: HAL in ‘2001’ Wasn’t Always So Eerily Calm.” *The New York Times,* 30 Mar. 2018, nytimes.com/2018/03/30/movies/hal-2001-a-space-odyssey-voice-douglas-rain.html?smid=url-share.

Ex Machina. Directed and written by Alex Garland, produced by Film4 and DNA Films, distributed by A24, 2015.

Galič, M., Timan, T. & Koops, BJ. Bentham. “Deleuze and Beyond: An Overview of Surveillance Theories from the Panopticon to Participation.” *Philos. & Technol*, vol. 30, no. 9–37, 2017. *SpringerLink,* doi.org/10.1007/s13347-016-0219-1.

Haraway, Donna. “A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century.” *Simians, Cyborgs and Women: The Reinvention of Nature,* Routledge, 1991, pp. 149-81.

Hayles, N. Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics.* University of Chicago Press, 1999. *ALCS Humanities E-boo,* hdl-handle-net.proxy3.library.mcgill.ca/2027/heb.05711

Harari, Yuval Noah. *Homo Deus: A Brief History of Tomorrow.* Harvill Secker, 2016.

Hutson, Matthew. “Who Should Stop Unethical A.I.?” *The New Yorker,* 15 Feb. 2021. Online, newyorker.com/tech/annals-of-technology/who-should-stop-unethical-

ai?utm_source=onsite-share&utm_medium=email&utm_campaign=onsite-share&utm_brand=the-new-yorker.

Kahneman, Daniel. *Thinking, Fast and Slow.* Farrar, Straus and Giroux, 2013.

Kittler, Friedrich A. *Gramophone, Film, Typewriter* (1986). Translated by Geoffrey Winthrop-Young and Michael Wutz, Stanford University Press, 1999.

Lockwood, Patricia. *No One is Talking About This.* Riverhead Books, 2021.

McAdams, Dan P. *The Art and Science of Personality Development.* The Guilford Press, 2015.

Metz, Cade. *Genius Makers: The Mavericks Who Brought AI to Google, Facebook, and the World.* Dutton, 2021.

Ngai, Sianne. *Ugly Feelings.* Harvard University Press, 2005. *ProQuest Ebook Central,* ebookcentral.proquest.com/lib/mcgill/detail.action?docID=3300232.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism.* NYU Press, 2018. *Project MUSE* muse.jhu.edu/book/64995.

"Nosedive." *Black Mirror.* Directed by Joe Wright, story by Charlie Brooker, season 3, episode 1. *Netflix,* 2016.

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

Odell, Jenny. *How to Do Nothing: Resisting the Attention Economy*. Melville House, 2015.

Ouchchy, L., Coin, A. & Dubljević, V. “AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media.” *AI & Soc,* vol. 35, Mar. 2020, pages 927–36, doi.org/10.1007/s00146-020-00965-5.

Peters, John Durham. *The Marvelous Clouds: Toward a Philosophy of Elemental Media*. The University of Chicago Press, 2015.

Roose, Kevin. *Futureproof: 9 Rules for Humans in the Age of AI*. Random House, 2021.

Scarry, Elaine. *The Body in Pain: The Making and Unmaking of the World*. Oxford University Press, 1985. *HathiTrust Online.*

Schectman, Marya. “The Case for Reasons of Self-Concern.” In *Reasons of One’s Own*. Eds. Maureen Sie, Marc Slors and Bert van den Brink. Ashgate, 2004, 107-28.

Segal, Howard P. “The Technological Utopians.” In *Imagining Tomorrow: History, Technology and the American Future,* edited by Joseph J. Corn, MIT Press, 1986.

So, Richard Jean. *Redlining Culture: A Data History of Racial Inequality and Postwar Fiction.* Columbia University Press, 2020.

Spohr, Dominic. “Fake News and Ideological Polarization: Filter Bubbles and Selective Exposure on Social Media.” Business Information Review, vol. 34, no. 3, Sept. 2017, pp. 150–60, doi:10.1177/0266382117722446.

Stahl, Carsten Bernd, Job Timmermans, and Brend Mittelstadt. "The Ethics of Computing: A Survey of the Computer-Oriented Literature." *ACM Computing Surveys,* vol. 48, no. 4-55, 2016. doi.org/10.1145/2871196.

Statt, Nick. "The AI boom is happening all over the world, and it's accelerating quickly." *The Verge,* 12 Dec. 2018, theverge.com/2018/12/12/18136929/artificial-intelligence-ai-index-report-2018-machine-learning-global-progress-research.

Tegmark, Max *Life 3.0: Being Human in the Age of Artificial Intelligence.* Published by Alfred A. Knopf, 2017.

The Invisible Committee. "Fuck Off, Google." *To Our Friends,* translated by Robert Hurley, 2014. *Textz,* textz.com/textz/Fuck%20Off,%20Google.pdf

Thorp, Jer. *Living in Data: A Citizen's Guide to a Better Information Future.* Farrar, Straus, and Grioux, 2021.

The Social Dilemma. Directed by Jeff Orlowski, produced by Larissa Rhodes, distributed by *Netflix.* 2021.

Tolentino, Jia. *Trick Mirror: Reflections on Self-Delusion*. Random House, 2020.

"TravisBott Jack Park Canny Dope Man [Official]." *Vimeo*, Uploaded and Created by Space 150, vimeo.com/384062745.

Turkle, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other,* Basic Books, 2011. *ProQuest Ebook Central*, ebookcentral.proquest.com/lib/mcgill/detail.action?docID=684281.

Turkle, Sherry. *The Second Self: Computers and the Human Spirit*. MIT Press, 2005. *ProQuest Ebook Central,* ebookcentral.proquest.com/lib/mcgill/detail.action?docID=3338499.

Underwood, Ted. "Dear Humanists: Fear Not the Digital Revolution." *The Chronicle of Higher Education,* 27 Mar. 2019, chronicle.com/article/dear-humanists-fear-not-the-digital-revolution/.

Weil, Elizabeth. "What Do Teens Learn Online Today? That Identity Is a Work in Progress." *The New York Times Magazine,* 13 Nov. 2019, nytimes.com/interactive/2019/11/13/magazine/internet-teens.html.

Westworld. Created by Lisa Joy and Jonathan Nolan, produced by HBO, seasons 1-3, 2016-2020.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Public Affairs, 2019.