Past Influences, Present Ethics: How History Shapes AI Decision-Making and Technological Futures

McKim Jean-Pierre

Erasmus University Rotterdam

Superviser: Natalia de Silva Perez

Abstract

OpenAI's release of products such as Chat-GPT, DALL-E, and Sora have recently triggered existential debates around artificial intelligence and its ultimate impact on humanity. While previous innovations seen as quite banal today have gone through similar panics – radios, televisions, microwaves and so on, the socio-economic environment surrounding the development of these digital innovations may cause warrant for these concerns. Through a historical analysis of the socio-economic environment developed during the financialization of the market in tandem with digital technology developments in the 1970s and onwards, this study places digital innovation entrepreneurs of today within the historical contexts which influence their ethical decision making. Guided by a discourse analysis via GloVe word embeddings, the ethical perspectives of Sam Altman and OpenAI are analyzed to gain a better understanding of how the path dependencies created in the late 20th century influence Big Tech leaders in their ethical decision making for the innovations which deeply impact our futures. The main take away is that what is defined as ethical varies between different academic lenses and for the industry. The differing perspectives between groups causes disagreement of what is ethical or not, the intention of Big Tech actors, how to address ethical issues within the tech, and the potential societal impact of these digital innovations. These disagreements cause misgivings between scholars of different backgrounds and those within the industry, with mistrust between groups over ethical approaches and decision making. Thanks to this multitude of varied interpretations and definitions, implementing ethical frameworks within digital technologies is complex. Consequently, the study concludes that a broader study is needed to see how to integrate ethics within the iterative innovation process of digital technology. Alignment of ethics between actors and scholars commenting, regulating, and involved in the digital innovation process is needed to do so. Therefore, a multidisciplinary study is advantageous to develop comprehensive solutional frameworks which can propose alignments which ensure that digital innovations fundamentally produce positive societal outcomes.

Key Words: Path dependency, artificial intelligence ethics, OpenAI, GloVe embeddings, discourse analysis, digital innovation

Contents

Introduction: Sequencing Ethical Influences

At the time of this text being written; reactions, discussions, and debates surrounding Artificial Intelligence (AI) have been rising in popular discourse. Some say that it has the potential to make the world a better place.[1] Others are sounding alarm bells, saying that this technology will create a dystopian future.[2] While the conversations and debates in popular culture surrounding AI are recent, the reaction toward it as an innovation is something we have seen before.[3][4] Orben describes this through her description of the Sisyphean cycle of technological panic. In this cycle, an innovation is introduced into society, moral panic ensues due to speculation that the innovation is a harbinger of dystopia and end of humanity as we know it, and finally the innovation is integrated into society with the realization that the doomsday predictions were wrong. The next innovation comes and the cycle repeats. Many innovations that we see as quite banal today went through this cycle, – radios, microwaves, televisions, and so on.[5] One may think that this is reassuring, that perhaps it means we are simply in the panic phase of the Sisyphean cycle and soon we will see that AI isn't as bad as we thought it would be. I argue that this is not the case.

The innovation-panic-calm cycle that we have witnessed from the mid-20th century until now may have given us a false sense of security towards innovations which should not be applied towards the digital innovations being developed today. In fact, the company that is the cause of the recent increase of public focus on AI, OpenAI, was founded and funded to *prevent* malicious actors capable of developing these powerful technologies from creating a dystopian future. Their mission being to make technologies which will "[benefit] all of humanity."[6] The question is, can OpenAI, or any of the other companies with similar goals such as Cohere and Google DeepMind achieve this goal in today's socio-economic landscape?

---

[1] Jao and Editor-in-Chief, "ChatGPT Will Not Turn the World into a Dystopian Wall-E Dump, NU Experts Assure You"; Scott and Soltys, "'The Revolution Has Launched.'"
[2] Martineau, "We Are Approaching an AI Dystopia."
[3] Orben, "The Sisyphean Cycle of Technology Panics."
[4] Amy Orben is a Programme Leader Track Scientist at the MRC Cognition and Brain Sciences Unit, University of Cambridge and a Fellow at St. John's College, University of Cambridge
[5] Orben, "The Sisyphean Cycle of Technology Panics."
[6] OpenAI, "OpenAI Charter."

Path dependency is often used in historical analysis to extrapolate reasons for present-day phenomena at the institutional, international, and macro levels.[7] This same logic should be applied to the individuals who are not only impacted by these higher-level path dependencies, but also have a significant chance in creating new ones. Socio-economic and historical events have indeed given C-suite entrepreneurs and innovators the power they have today within their institutions, however, these events have also influenced their perceptions of the goals and uses of the technologies they choose to innovate. Their perceptions ultimately impact their decisions, which impact the development of innovations that impact society, entrenches their placement within it, and continues the feedback loop. This is significant; the dystopian potential of artificial intelligence innovations is decided through the perceptions, morality, and goals of the groups of individuals who lead the companies that develop them. This thesis therefore aims to historically contextualize the socio-economic environments surrounding C-suite leaders of companies developing digital innovations.[8] By understanding the socio-economic factors which have influenced the leaders at companies such as OpenAI, Cohere, and Google DeepMind to invest in the technologies they do, we can understand the *whys* and *hows* of their decision-making. Through this, we can better understand the societal impact that these technologies may have. Ultimately, we will seek to answer the **central research question of this master's thesis: how has digital capitalism impacted the ability of businesses to build ethical big data technologies?**

Before we continue on, I will outline the roadmap of this thesis. The theoretical framework used in this thesis makes the road to the case study a long one. The decision to take a long road is due to the core tenant of this thesis: a thorough understanding of inputs is needed to analyze any output. In each section of the thesis, we are fundamentally asking: What is this? Who builds this? Why may they be building like this?

[7] James Mahoney is a Gordon Fulcher Professor in Decision-Making, Professor of Sociology, and Professor of Political Science at Northwestern University

[8] To note, when I speak of "digital innovations" or "digital technology" I am mainly referring to things that are grouped under the Information and Communications Technology (ICT) umbrella.

In the first chapter, I will provide my theoretical framework, which will provide the lens through which we will analyze the themes and sources touched upon and justify why I have chosen to explore the way I have. A sub-argument already alluded to is, given the gravity of digital technologies today, there is an urgent need to contextualize every single potential actor of influence. Within this thesis, I have decided to make this the case for every individual cited, including the scholars I cite, the leadership teams at, and even myself as a scholar. The theoretical framework will help with parsing out why I have taken this approach.

In the second chapter of my thesis, with this theoretical framework made clear, I will paint the historical picture that may explain how path dependencies influence businesses leaders making ethical decisions for their technologies. The historical analysis will study the financialization of markets in the late 20th century and into the 21st century, the digital technological developments which happened in tandem to this financialization, and the implications of these developments. As my research question mentions the idea of "digital capitalism," a necessary exploration on whether there is a unique "digital capitalist" economic era is unique or not is also taken in this section. By the end of this section, we will understand the core aspects of this era, unique or not, which have produced the socio-economic environment which decision makers and leaders of these digital technologies are influenced by.

Following the historical framing and building towards "modern day", the third chapter of my thesis will start to provide the context for the specific case study of this thesis. This research will look at OpenAI as the case study, and in this chapter I will elaborate on why I selected OpenAI. Following this, I will also give background for what OpenAI is, who are its leaders, and what are the current understandings of their approaches to ethical integrations within their technologies. We will also explore the potential limitations or problems with the current approaches to understanding the interpretation of their ethical approaches.

The fourth chapter will contain my discourse analysis of Sam Altman and OpenAI. This discourse analysis will be done by looking at interviews given by Sam Altman following the establishment of OpenAI as well as blog posts on OpenAI's website. Quantitative data will be produced by using GloVe model word embeddings.

In the fifth chapter, I will discuss how the findings might reflect how socio-economic path dependencies influence businesses such as OpenAI in their ability to include ethical considerations into the development of big data technologies such as AI. Here I will reflect on these findings and further develop the ideas and frameworks uses throughout the thesis. I will also discuss my own biases and limitations of this study via the same approach used to analyze the case study.

The intention of this thesis is threefold. First, to create an innovative structure that operationalizes deep historical contextualizing and framing through the theoretical frameworks. Secondly, to provide a proactive and reflexive approach to analyzing the thoughts and ideas of all actors who are engaged with the discussion, from the scholars cited in the analytical process to the case study actors' actions, to the writer's, i.e., my own, positionality and lens. Lastly, to build an analytical approach via a specific case study that can be applied to similar cases to analyze the ethical considerations within the innovation process critically. Ultimately, the point of this thesis is not to support or criticize big data business leaders. It is to provide the reader with the unique context in which we live today and reflect on how it influences ethical approaches to innovations which ultimately impact our lives from the individual to the global level.

Chapter 1: The Theoretical Architectures for this Ethical Perspective Analysis

"The forms or conditions of production are the fundamental determinant of social structures which in turn breed attitudes, actions, and civilizations"[9][10]*– Joseph Schumpeter: Capitalism, Socialism and Democracy.*

The opening quote of this section provides an essential argument of the relevance of historically contextualizing digital technology innovators to analyze their decisions when building their technologies. We can see the "conditions" as the economic and institutional spheres that big data business leaders come from and are influenced by. From being affected by these spheres, they, in turn, influence the surroundings of others with their digital technologies and innovations that significantly impact our daily lives. Therefore, it is crucial to develop and utilize frameworks that will allow us to contextualize the circumstances and contexts from which tech leaders come from.

Four theoretical concepts will provide the contextual framing in my analysis. The first one is a decolonial framework. Rachel Adams[11] describes the importance of bringing a decolonial framework to analyzing decisions surrounding AI. As the case study for this research will look at tech leaders who are developing AI technologies, her framing of AI decision-making analysis is relevant for this task. Adams argues that present-day AI is made possible by, and depends on, colonial forms of power.[12] The integration of the decolonial framing of AI's formation and appropriation by "neo-colonial" actors today will allow for an understanding of the histories that have influenced the institutions that ultimately create the conditions for tech companies' decision-making.[13] The decolonial lens is also essential, as it includes 'decolonization,' which seeks to identify and critique new forms of colonialism/imperialism *and*

[9] Schumpeter, "Capitalism, Socialism and Democracy," 12.
[10] Joseph Schumpeter was an Austrian political economist
[11] Rachel Adams is a Senior Research Specialist at the human Sciences Research Council, South Africa
[12] Adams, "Can Artificial Intelligence Be Decolonized?"
[13] Adams.

undo these practices.[14] To operationalize the undoing of these practices through identifying and critiquing these new forms, the first time I mention a scholar or individual in my text, I will include a footnote of their field and research/work location. Through doing this, I participate in the identification of these new forms. However, there also comes the critique part. I will integrate as best as I can critique where relevant. As I will be doing this with all other scholars, I will do the same for myself.[15]

The following theoretical concept is the Symbolic Order. John Burgess[16] and Juliette Spurling[17] develop the framework of the Symbolic Order from Lucan's "Big Other," applying it to AI ethics research.[18] Symbolic Order, according to them, is the framing that we as individuals use to allow ourselves to make sense of our worlds.[19] The Symbolic is "the register where language, laws, and norms originating from culture outside of us and put in place before our birth become the dominant force for self-understanding."[20] The Symbolic Order complements the previous lens through its acknowledgment of the greater macro histories at play, but with the emphasis that these broader dynamics impact individuals decisions and perspectives through their placement within social paradigms. While the decolonial framework helps us understand the histories of institutions, Symbolic Order provides us with a reminder that individuals both intentionally and inadvertently place themselves within this larger historical paradigm, ultimately influencing their perspectives and decisions. The micro and individual lens that Symbolic Order brings is crucial to understanding where individual actors may place themselves within present-day ecosystems, and how this is reflected in their decision making.

[14] Adams.
[15] McKim Jean-Pierre is an International Masters candidate in Global Markets, Local Creativities at University of Glasgow, a Masters in Social Science candidate in Economic History at Uppsala University in Uppsala, Sweden, and a Masters of Arts candidate in History at Erasmus University in Rotterdam, Netherlands.
[16] John Burgess is an Assistant Professor at the School of Library and Information Studies in the College of Communication and Information Sciences at the University of Alabama
[17] Juliette Spurling is a Master of Library Information Science (MLIS) graduate at the University of Alabama
[18] Burgess and Spurling, "The Societal Risks and Moral Harms of Submitting to Artificial General Intelligence as Lacan's 'Big Other.'"
[19] Burgess and Spurling.
[20] Burgess and Spurling.

The third and fourth theoretical frameworks applied to this research are professional scientific idealism, brought forward by Carl Mitcham[21] and Anne Gerdes,[22] and the critical science approach, brought forward by Shakir Mohamed[23], Marie-Therese Png[24], and William Isaac[25]. Somewhat meta, these two frameworks justify the selection of the previous frameworks.

Professional scientific idealism asks for scientists to critically reflect on the technologies that they help develop.[26] This critical reflection involves considering the ethical implications of technologies developed, with the hope that upon this reflection, ethics can be integrated into the research and development process.[27] Mitcham argues that professional scientific idealism is an underused lens in the science-technology-society (STS) field.[28] Yet, thanks to the intertwined nature of technology, science, and business, professional scientific idealism is important to use as scientists hold a uniquely deep understanding of technologies. While Mitcham and Gerdes develop this theory with Machine Learning and Big Data scientists in mind, I believe this is applicable and should be applied to the social scientists and researchers studying the field from adjacent perspectives. Social scientists engaged in debates surrounding the ethics of digital technologies are often called forward to help consult or provide feedback to these technology companies or policy makers when ethical issues come to question. In a later chapter, we will see an example of this. Therefore, even when influence is not as direct as developing the technologies themselves, the potential to influence the policies built around the technologies warrants the same ethical standards. This is why I am applying professional scientific idealism to my own research.

---

[21] Carl Mitcham is a philosopher of engineering and technology, Professor Emeritus of Humanities, Arts, and Social Sciences at the Colorado School of Mines

[22] Anne Gerdes is an associate professor at University of Southern Denmark, Department of Design and Communication

[23] Shakir Mohamed is the Director for research at Google DeepMind

[24] Marie-Therese Png is a PhD candidate at Oxford University's Oxford Internet Institute (OII)

[25] William Isaac is a Staff Research Scientist on DeepMind's Ethics and Society Team and Research Affiliate at Oxford University Centre's for the Governance of AI.

[26] Mitcham, "Professional Idealism among Scientists and Engineers."

[27] Mitcham.

[28] Mitcham.

While professional scientific idealism focuses on critical analysis for ethical output, the critical science approach focuses on acknowledging the socio-academic inputs that influences a scholar's stance.[29] The critical science approach allows us to "uncover the underlying cultural assumptions that dominate a field of study and broader society."[30] Thus, this framework positions researchers and builders of technology within the metropoles of technological (and informational) power they come from, exist within, and potentially recreate.[31] This compliments both the decolonial and Symbolic Order frameworks. Professional scientific idealism and the critical science approach frameworks are thus used to define the moral code for the position of my research, as well as encourage constant awareness of who is speaking and why they may be speaking the way they are.

Ethical AI scholars often speak of the risk of the "black box," which describes how opaque AI creates outputs that cannot be explained or deciphered. This paper attempts to continuously remove a layer of "black box" when it comes to the thoughts, approaches, and perspectives of individuals by integrating a deeply historical and reflective analysis to each person. Through the Symbolic Order, decolonial, critical science approach, and professional scientific idealism, we have the tools needed to provide a robust answer to our primary research of how digital capitalism has impacted the ability for businesses to build ethical big data technologies.

[29] Mohamed, Png, and Isaac, "Decolonial AI."
[30] Mohamed, Png, and Isaac.
[31] Mohamed, Png, and Isaac.

Chapter 2: The History of the Digital and Capital of Digital Capitalism

Of course, men "choose" their course of action which is not directly enforced by the objective data of the environment; but they choose from standpoints, views and propensities that do not form another set of independent data but are themselves molded by the objective set.[32] *- Schumpeter, Joseph. Capitalism, Socialism and Democracy*

Our primary research question asks, *how has digital capitalism impacted the ability of businesses to build ethical big data technologies?* The ability to build ethical big data technologies is reflected through the choices that leaders make as they fund and develop their innovations. To understand their choices, we need to understand the "standpoints, views, and propensities" that they chose from. The Symbolic Order[33] and decolonial[34] frameworks encourage us to use the critical analysis of history and social context to understand the shaping of decisions. The history for Information and Communications Technology (ICT) businesses involves major developments in and for their digital creations from the 1970s onward.[35][36] This happened to coincide with major political and historical events which shifted western capitalism.[37][38] The socio-techno-economic environment that therefore emerged during this era in the late 20th century is what informs the "standpoints, views, and propensities" of todays digital innovation entrepreneurs. Or, in the words of our theoretical framework, their Symbolic Order. Therefore, in this chapter we will answer the following sub question: **how did major developments of ICT technologies and tandem shifts in capitalism from the 1970s to today influence the symbolic order of todays digital innovation entrepreneurs?** By answering this

[32] Schumpeter, "Capitalism, Socialism and Democracy," 12.
[33] Symbolic Order: Our use of social and cultural environmental feedback to frame and understand our placement within these worlds.
[34] Decolonial: Identifying and critiquing new forms of colonialism/imperialism by recognizing that that colonial histories have influenced modern day institutions.
[35] Hendrikse et al., "The Big Techification of Everything."
[36] Reijer Hendrikse is a financial geographer and political economist at the Vrije Universiteit Brussel (VUB)
[37] Benería, Berik, and Floro, *Gender, Development, and Globalization*, 93–112; Block and Somers, *The Power of Market Fundamentalism*.
[38] Lourdes Benería is an economist and Professor Emerita at Cornell University's Department of City and Regional Planning, Günseli Berik is a Professor of Economics at The University of Utah, Maria Floro is a Professor Emerita of Economics at American University, Fred Block is a research professor of sociology at the University of California, Davis, and President of the Center for Engaged Scholarship, and Margaret Somers is a Professor Emerita of Sociology and History at the University of Michigan

question, we will understand the important socio-historical contexts which inform the choices of digital innovation leaders. Through knowledge of what is informing their choices, we will be better equipped to understand their ability to make ethical decisions.

The Late 1970s-Early 1990s for Tech

Reijer Hendrikse et al. in break down the progression of the "big-techification of everything" [39]through a timeline which shows the development of ICT technologies from the 1970s up until modern day.

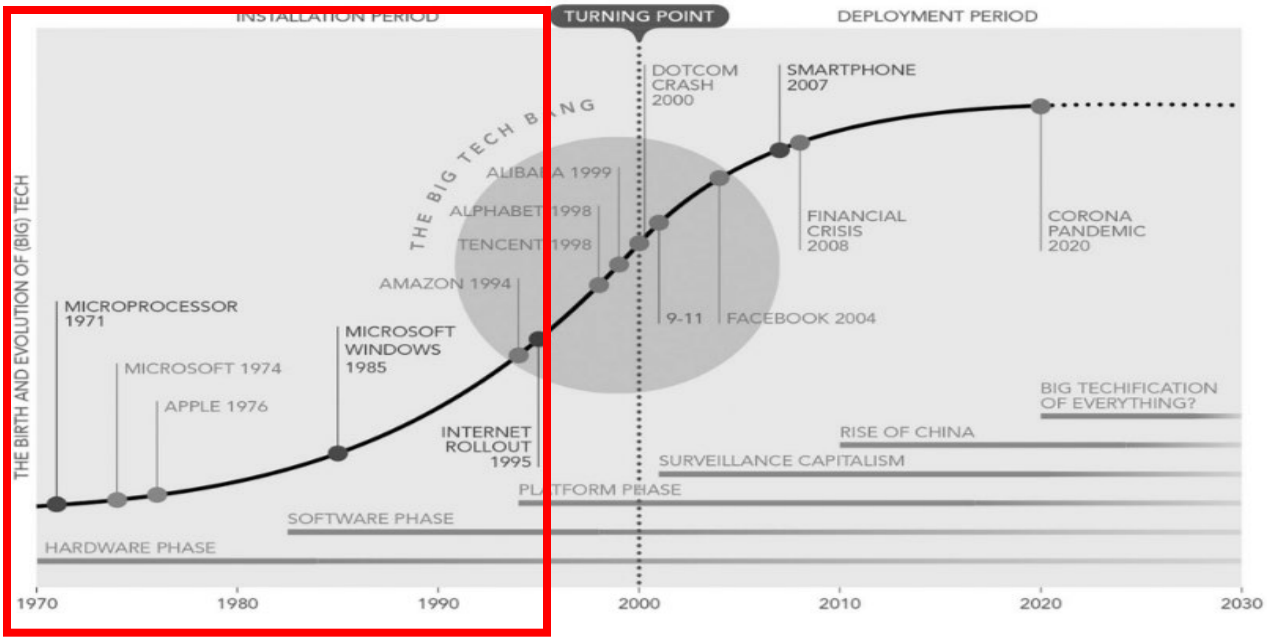

Figure 2.1 1970s-Early 1990s for tech (Source: *'The Big Techification of Everything'*)[40]

As we can see in the part of the chart highlighted in red, the 1970s and 1990s was the "installation period" for Big Tech.[41] During this installation period, hardware and software was developed and companies such as Microsoft and Apple were started. However, it is important to

[39] Hendrikse et al., "The Big Techification of Everything."
[40] Hendrikse et al.
[41] Hendrikse et al.

note that during this time Microsoft and Apple weren't the big companies we know of today, rather the ventures of entrepreneurs.

Domenico Catalano[42] shares that from the 1980s and onwards there was a return of the innovative entrepreneur after a long period of eclipse between 1940 and 1973.[43] This provides interesting overlap to Hendrikse's timeline of Big-Techification. Catalano ultimately echoes the rise in big-techification, by highlighting that while the years between 1973-1981 did see the return of independent investors and entrepreneurs, post 1981 there commenced a significant decline in societal contributions by the single entrepreneur in favor for companies now rising as Big Tech.[44] Thus, according to Catalano, the installation period from the 1970s-1990s also came with it an entrenchment of corporate power through barriers of 'free market' entry by the emergent Big Tech'.[45]

The Late 1970s-Early 1990s for Markets

While ICT technologies were going through their installation period between the late 1970s to early 1990s, markets were being financialized. A particular socio-economic environment and characteristic developed between the late 1970s to early 1980s when western markets (especially those in the U.S. and the U.K.) started shifting towards a more financialized model with the introduction of the Thatcher and Raegan administrations.[46] Financialized policies and approaches were stewarded by governments focused on creating more neoliberal frameworks. Neoliberal frameworks emphasized the privatization of public assets to generate revenue for the sake of increasing GDP; with the assumption that profit maximization was the most effective institutional means for societal well-being/maximization.[47][48]

[42] Domenico Catalano is an Associate Professor of Economics and Management in Culture and Communications Department at University of Paris VIII.
[43] Catalano, "Entrepreneurship and Digital Capitalism."
[44] Catalano.
[45] Catalano.
[46] Benería, Berik, and Floro, *Gender, Development, and Globalization*, 96–112; Block and Somers, *The Power of Market Fundamentalism*.
[47] Benería, Berik, and Floro, *Gender, Development, and Globalization*; Tasioulas, "Artificial Intelligence, Humanistic Ethics."
[48] John Tasioulas is Professor of Ethics and Legal Philosophy and the inaugural Director of the Institute for Ethics in AI at the University of Oxford

Catalano, mentioned in the previous section, argues that capitalism was transformed from the 1980s and onwards via 'new digital capitalism'.[49] According to him, what made this era unique was that modes of production were now more focused on the power of networks, which was accompanied by an emergence and concentration of 'big tech'.[50]

Michael Grimshaw[51] labels the change spanning from the 1970s onwards as a shift towards as financial capitalism, wherein a system encourages using money to make more money.[52] According to Grimshaw, this shift towards financial capitalism is marked by a shift towards a post-industrial society, where western nations relocate industrial processes into emergent economies for lower production and labor costs.[53] While this happens, western economies see a shift in power from industrial capitalists towards financiers and financial institutions, those who hold speculative and investing power.[54]

Benería et al. focus on how the financial liberalization from the late 1980s onward caused a dramatic increase in the power of the financial sector's growing dominance in the global economy.[55] The reasoning behind the neoliberal policies put in place to grow the financial sector's dominance in the global economy is criticized by Karl Polanyi[56], Fred Block[57] and Margaret Somers.[58] The justification for neoliberal policies was/is that human society is fundamentally shaped by the needs of the economy, and that therefore prioritizing the status of a rapidly developing economy will successfully address the needs and respect the sovereignty of the self-interested, materially motivated individual.[59] Polanyi rejects this as

[49] Catalano, "Entrepreneurship and Digital Capitalism."
[50] Catalano.
[51] Michael Grimshaw is an Associate Professor at the Philosophy Department at the University of Canterbury
[52] Grimshaw, "Towards a Manifesto for a Critical Digital Humanities."
[53] Grimshaw.
[54] Grimshaw.
[55] Benería, Berik, and Floro, *Gender, Development, and Globalization*.
[56] Karl Polanyi was an economic anthropologist, economic sociologist, and politician
[57] Fred Block is a research professor of sociology at the University of California, Davis, and President of the Center for Engaged Scholarship
[58] Margaret Somers is a Professor Emerita of Sociology and History at the University of Michigan
[59] Block and Somers, *The Power of Market Fundamentalism*.

"economic fallacy", but nevertheless this concept, fallacy or not, was used to shape policy during this era.[60]

Nguyen mentions the assessments by Western scholars such as Kotz, Crotty, Duménil, and Lévy who emphasize that since the 1980s, finance capitalism underwent rapid growth, eventually becoming "monopoly-finance capital".[61] Guy Standing[62] also mentions that around the 1980s finance, and thus financialization, dominated institutionally, hallmarked by "the increasing redistribution of wealth upwards as rents to those owning property – physical, financial, and 'intellectual'" with a rise in rentier capitalism.[63]

Nguyen argues that monopoly has always been a fundamental feature of capitalism since the nineteenth century, and that therefore only novel feature of the capitalism that was developed from the 1980s to today are the mechanisms of the digital age.[64] These digital mechanisms, he argues, continue the power imbalances that existed in the economic era of 'before'. Jonathan Pace[65] has a similar perspective to Nguyen. In his article, he questions Schiller, who offers three distinct definitions of digital capitalism, and Fuchs, who also argues that this era holds a unique transformation in capitalism due its transformation of the means of production, now rooted in the production, collection, and monetization of knowledge thanks to information technology.[66] Pace ultimately argues that digital technology mediates the structural tendencies of capitalism without ultimately changing its structural totality or adding a new historical period.[67]

An entire thesis on its own can be written to analyze the uniqueness (or not) of the economic era that was developed between the 1970s to 1990s. While we have seen above that

[60] Block and Somers.
[61] Nguyen, "Avoiding the Delusions of Today's Capitalism with a Thorough Understanding of Marxism as the Key."
[62] Guy Standing is a Professorial Research Associate at SOAS University of London and a founding member and honorary co-president of the Basic Income Earth Network (BIEN)
[63] Standing, "The Precariat."
[64] Nguyen, "Avoiding the Delusions of Today's Capitalism with a Thorough Understanding of Marxism as the Key."
[65] Jonathan Pace was a postdoctoral fellow in the Digital Civil Society Lab and PhD from the Annenberg School for Communication at the University of Pennsylvania.
[66] Pace, "The Concept of Digital Capitalism."
[67] Pace.

debates concerning whether or not the time period from the 1970s onwards constitutes a unique economic era or not, there is still a general agreement that from the 1970s neoliberal policies were developed which emphasized and encouraged the creation and collection of intangible assets. Knowledge is quite intangible, however digital technologies which collect, and store data make knowledge feel more tangible.

Applying our decolonial framework, it is important to note that as we have studied what happened during this era, that these shifts were happening specifically within western economies. The 'post-industrial' society which emerged from this era was for western markets, meaning that industrial work was displaced to the countries of the world which historically were the colonies or subjects of the now named western powers. Here we find an example of the recreation of structures similar to colonial and imperial times. Where this is significant and relevant to this study, is that the governmental and market emphasis on creation and collection of intangible assets by western markets fed into the and built the scaffolding of the Symbolic Order[68] of Big Tech leaders. This means, however, that the technical solutions adapted for these markets are solutions adapted by and for the western market specifically. The risk in this is creating technologies which leave out those outside this sphere and re-instill colonial or imperial dynamics. Let us now explore what ICT companies, then, bring forward in the next era.

[68] Symbolic Order: Our use of social and cultural environmental feedback to frame and understand our placement within these worlds

The Early 1990s-2010s for Tech

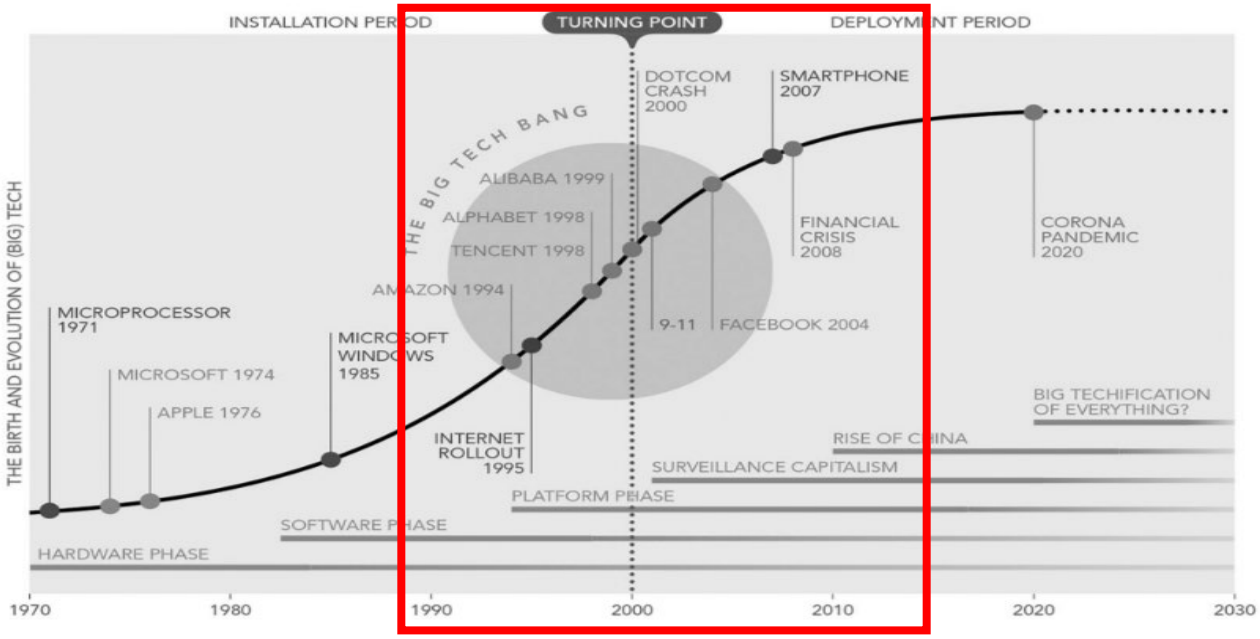

Figure 2.2 Early 1990s-2010s for tech (Source: *'The Big Techification of Everything'*)[69]

The 90s to the 2010s encompass the "turning point" and the "Tech Big Bang" according to Hendrikse et al.[70] This turning point saw the emergence of companies such as Amazon, Alphabet (Google), Tencent, and Alibaba. This also introduced the "platform phase" which eventually developed into the "surveillance capitalism" phase.[71] Platformization of technologies from the 2000s and onwards came thanks to a series of events and technological feats during this era. Following the dotcom crash of 2000 and 9-11 in 2001, tech companies needed to find profit models which allowed for the continued growth of their companies and funding for the development of tech.[72] With the financialization era emphasizing intangible assets well underway, tech companies were in the process of, and decided to, further develop the tools needed to make profits in this kind of economy. Data helps for predictability and can help

[69] Hendrikse et al., "The Big Techification of Everything."
[70] Hendrikse et al.
[71] Hendrikse et al.
[72] Zuboff, *The Age of Surveillance Capitalism*, chap. 7.

contribute towards profit maximization in an increasingly volatile market- made volatile thanks to the financialized nature of the market.[73] The neoliberal and privatized leaning polices also instilled the feeling that focusing on profits tied to their technological feats in this way would ultimately help make the world a better place. We can see this assumption in reports published during this era, and especially in the 2010s, which speak about the potentially positive impacts of digital globalization.[74] These inputs are the socio-economic background to the technological developments and innovations of from the 1990s to the 2010s.

Nigel Shadbolt[75] spotlights that during the 2010s, power, storage, and communications speeds doubled every fifteen months, which allowed for data collecting methods, techniques, and approaches previously unimaginable.[76] Dominic Smith[77] et al. also note that beginning around 2011 "deep learning techniques began to produce dramatic advances in speech recognition, visual object recognition, and machine translation."[78]

Big Tech took all this data and created "knowledge economies" "data-markets" and "behavior economies", all focused on extracting information from people for profit.[79] In other words, intangible capital. Shoshanna Zuboff[80], with a retrospective perspective, looks at how, these technologies were indeed built with a focus on profit via data to predict, and eventually, influence users.[81] Schumpeter underscores that "The forms of production themselves have a logic of their own; that is to say, they change according to necessities inherent in them so as to produce their successors merely by their own working."[82] Data collected by Big Tech did precisely this, as it started to be sold to the financialized markets that rewarded prediction,

[73] Benería, Berik, and Floro, *Gender, Development, and Globalization*; Tasioulas, "Artificial Intelligence, Humanistic Ethics."
[74] "Digital Globalization: The New Era of Global Flows | McKinsey."
[75] Sir Nigel Richard Shadbolt is Principal of Jesus College, Oxford and Professorial Research Fellow in the Department of Computer Science, University of Oxford.
[76] Shadbolt, "'From So Simple a Beginning.'"
[77] Dominic Smith is a Senior Lecturer in Philosophy at the University of Dundee
[78] Smith et al., "Reimagining AI."
[79] Zuboff, *The Age of Surveillance Capitalism*, chap. 7.
[80] Shoshana Zuboff is the Charles Edward Wilson Professor Emerita at Harvard Business School and a former Faculty Associate at the Berkman Klein Center for Internet and Society at Harvard Law School
[81] Zuboff, *The Age of Surveillance Capitalism*.
[82] Schumpeter, "Capitalism, Socialism and Democracy," 12.

technologies which could engender those predictive data points increased. Maneuvering and predicting human opinions and behaviors was quite profitable for technology companies. Thanks to these big profits, the revering and idolization of the entrepreneurs creating these groundbreaking digital innovations increased. Thus, entrenching the Symbolic Order[83] of not only these leaders, but the place of the tech industry within society as a whole.

The Early 1990s-2010s for Markets

The 1990s-2010s for the global economy was a complex time as well. The idea of a 'global economy' being something new parallels the argument of whether or not 'digital capitalism' is truly a new thing, however, there were still notable shifts thanks to the digital innovations put into the market and economic events. The early 2000s and 2010s sparked great discussions surrounding the "global village", where these technologies had the ability and positive potential to help connect the world and make it a more equitable place.[84] During this era, the neoliberal and financialized markets maintained its hold, even through ups and downs which were ironically created thanks to predictive bubble bursts such as the dotcom crash and the 2008 financial crash.

According to Catalano, it was during this era that the shape of markets was modified to lean towards "capital accumulating" firms, where large firms (which includes Big Tech firms) became a dominant form of economic life.[85] The markets emphasized capital accumulation during this era, and therefore encouraged the collection of data by Big Tech firms who increasingly and rapidly had access to more and more. During this era, digital technologies and markets became increasingly intertwined, as the operating logics of the economy and society itself happened via the digital technologies.[86]

[83] Symbolic Order: Our use of social and cultural environmental feedback to frame and understand our placement within these worlds
[84] "Digital Globalization: The New Era of Global Flows | McKinsey"; "IMF Finance and Development Magazine."
[85] Catalano, "Entrepreneurship and Digital Capitalism."
[86] Hendrikse et al., "The Big Techification of Everything."

Hendrikse argues that this era introduced to the markets the period of "shareholder revolution" in which "patient capital", those were willing to "lose" money in the short-medium term to maximize shareholder value in the long term, was introduced and propagated.[87] During this timeframe, there was also an increasing transfer of earnings from Big Tech companies to financial markets via interest payments, dividend payments, and stock buybacks.[88][89]

Whether or not the 1990s to the 2010s is a unique economic era is contested. Pace, who we spoke about in the previous section, argues that digital technology simply mediated the structural tendencies of capitalism without ultimately changing its structural totality or adding a new historical period.[90] However, this time frame at a minimum highlights the increasing intertwining of digital innovations and the market. With economic events such as the dot-com crash and 2008 financial crash, there was a greater need for predictability amidst the volatile financialized markets. In addition to these economic events, there were also historically significant political events which happened in the United States, where many of this Big Tech giants came from. Events such as the 9/11 terrorist attacks also instilled the government need for predictability not only within its markets, but within its borders, encouraging a stronger push for surveillance for the sake of national security.[91]

Applying our decolonial lens once more, this further ingrained the dynamics of east vs west, developed vs underdeveloped, and 'global north' vs 'global south'. The significance here stays the same as for the significance in the previous era we explored. All of this fed into the Symbolic Order of those developing these technologies. We should reflect that these technologies were being developed for the safety and surveillance of and for the west. Once again you can have certain groups being left behind through these safety structures being tailor-made for this particular in-group. The pervasiveness of tech on the global scale, however, means an entrenchment of the dynamics of power through who has the power to decide what

[87] Hendrikse et al.
[88] Orhangazi, "Financialisation and Capital Accumulation in the Non-Financial Corporate Sector."
[89] Ozgur Orhangazi is a professor of economics at Kadir Has University, Istanbul
[90] Pace, "The Concept of Digital Capitalism."
[91] Zuboff, *The Age of Surveillance Capitalism*.

data is collected as well as what or whom exactly society needs to be kept safe from. The tech leaders developing these technologies are informed by all of these dynamics.

The Implications of these Eras – The Symbolic Order of Digital Innovation Entrepreneurs

The term "digital capitalism" in the overall research question of this thesis[92] is there to highlight the deep intermingling and intertwining between tech developments and global markets from the 1970s to today. Understanding this entanglement can help us understand the actual and possible ethical decision making of digital innovation development leaders.

Through our historical analysis, we ultimately found that the financialization of markets led to the financialization of technological development. Digital innovation entrepreneurs created products and profit models that fit into the financialized market schema, where the deepening entanglement cultivated an environment where the right decision for digital innovation entrepreneurs is to find, create, and build technologies which are successful for the financialized market. Nguyen argues that the funding and development of these kinds of tech have "influenced the public to consume products that are conductive to capital growth rather than products that can improve people's livelihood."[93] Scholars such as Polanyi echoes this sentiment, saying that the neoliberalism of this era, or for him market fundamentalism, creates a market society where human beings are transformed commodities.[94] Yet, thanks to this deep intertwinement between digital innovations and the markets, Big Tech Leaders may not see the ethical concerns and implications of this.

Oliver Nachtwey[95] and Timo Seidl[96] conceptualize the Symbolic Order[97] of digital innovation entrepreneurs. They develop the term of "solutionists", who, according to them,

[92] *How has digital capitalism impacted the ability of businesses to build ethical big data technologies?*
[93] Nguyen, "Avoiding the Delusions of Today's Capitalism with a Thorough Understanding of Marxism as the Key."
[94] Block and Somers, *The Power of Market Fundamentalism*.
[95] Oliver Nachtwey is a professor at the Faculty of Philosophy and History Department of Social Sciences at University of Basel
[96] Timo Seidl is a post-doctorate researcher (Universitätsassistent) at the University of Vienna's Centre for European Integration Research (EIF)
[97] Symbolic Order: Our use of social and cultural environmental feedback to frame and understand our placement within these worlds

believe that "all relevant social problems can, in principle, be solved technologically; in which there is a technological hammer for every social nail."[98] They break down these beliefs further by parsing out the solutionist belief system as follows[99]:

- **Central Value**: Solving key social problems.
- **Evaluation Criteria**: Upgrading and Enabling Humans/Humanity.
- **Mode of investment (sacrifice)**: Risk-Taking (Failure).
- **Ideal Type**: Philanthro-Entrepreneur.
- **Insanity[100]:** Business without purpose, purpose without technology, technology without a business.
- **Test[101]**: Disruption.
- **Anthropology/Cosmology**: Humans have flaws but also potential, technology creates win-win solutions.

To provide more context to what each category means, as well as the construction of

Order of Justification	Market	Industrial	Inspiration	Domestic	Opinion	Civic	Project	Green	Solutionist
Central Value	Free Competition	(Technical) Efficiency	Originality	Traditional/Local Authority	Recognition	Public Good	Activity (Agility)	Sustainability	Solving key social problems
Evaluation Criteria	Value, Price	Technical performance, productivity, planning	Creativity, Non-conformity, Authenticity, Epiphany	Local esteem, personal recognition	Renown, fame	Rule governed, representation, solidarity	Successful self-management, number of projects and contacts	Sufficiency, environmental friendliness	Upgrading and Enabling Humans/Humanity
Mode of investment (sacrifice)	Opportunism	Disenchantment	Not fitting in	Responsibility for and attachment to inferiors/the local	Forgoing privacy	Forgoing one's own interests	Forgoing of stability	Avoiding Externalities	Risk-Taking (Failure)
Ideal Type	Businessman, merchant	Engineer, professional expert	Creative Genius, Misfit	(Company) Patriarch	Celebrity	Public Servant, virtuous politician	Enterprenrial self	Environmentalist	Philanthro-Entrepreneur
Insanity	Regulation	Spontaneity, Squander	Conformity	Presumption	Anonymity	Corruption	Immobility, inactivity	Unconditional Growth	Business without purpose, purpose without technology, technology without a business
Test	Competitiveness	Procedure test	Original thoughts, invention, creations	Trustworthiness	Publicity	Equal rights and duties (*contrat social*)	Finding new projects	Healthy Environment	Disruption
Anthropology / Cosmology	Humans follow their enlightened self-interest, markets create win-win situations (invisible hand)	Nature and Society can be mastered through calculation and planning	Aesthetic of the Genius and Conformity of the masses	Natural harmony as a result of natural hierarchy	Humans as craving for recognition	Humans as political equals	Humans as entrepreneurial selves in a connected world	Humans need a harmonious, stable relationship with their environment	Humans have flaws but also potential, technology creates win-win situations

[98] Nachtwey and Seidl, "The Solutionist Ethic and the Spirit of Digital Capitalism."
[99] Nachtwey and Seidl.
[100] To clarify, "insanity" here is what circumstances/constraints make this group feel withheld in their goal reaching.
[101] To clarify "test" here is what tangible results show success in achieving their goals.

other polities, I include their full table below.

Figure 2.3. Polity Overview (Source: *'The Solutionist Ethic and the Spirit of Digital Capitalism'*) [102]

We can directly see the influence of the political policies developed in the 1970s, which favored privatization, financialization, and trickledown economics approaches on the solutionists set of beliefs. Nachtwey and Seidl highlight the consequence of this, saying that solutionists see money making and making the world a better place not as contradictory, but rather as something that "can and should go hand in hand."[103] Another aspect of the influence of this era is in the way that money is used by solutionists to solve societal problems. With the financialized era placing more emphasis on the private market to take care of previously public needs, seeing tech which is developed in the private market as an important way to address societies needs is a logical progression of thought. Solutionists indeed have a deep belief in technology specifically being able to do this through believing that social problems can be solved through eliminating inefficiencies and deficiencies of society with the right technology.[104] This solutionist framing, and perspective is also further strengthened and informed by the developers of the technologies themselves, the computer scientists. Let us briefly explore how.

So far, we have developed an understanding of how the socio-economic context formed since the 1970s developed the symbolic order[105] of digital innovation entrepreneurs into a solutionist one, which influences their subsequent ethical decision making. The perspectives of the expert individuals who surround them, guide, and personally develop their technologies also informs their solutionist beliefs as well. The critical science approach asks us to look at the societal and personal academic lens which influences a scholar's stance. The socio-economic history which we just explored also surrounded the developers of the technologies themselves. Let us, then, briefly explore how this might have impacted their stances as scholars.

[102] Nachtwey and Seidl, "The Solutionist Ethic and the Spirit of Digital Capitalism."
[103] Nachtwey and Seidl.
[104] Nachtwey and Seidl.
[105] Symbolic Order: Our use of social and cultural environmental feedback to frame and understand our placement within these worlds

Ross Graham[106] does a discourse analysis of ethical debates surrounding the development of AI and Artificial General Intelligence (AGI) by looking at 81 articles and books to find reoccurring themes.[107] AGI is seen as the next logical progression from AI by big data innovation companies such as OpenAI (which has explicitly stated its goal is to build AGI safely and responsibly).[108] Grahams study therefore provides pertinent examples of how the viewpoints of computer scientists contribute to the solutionist mindset and consequently influence ethical decision making.

In his study, Graham split the scholarly debaters of AGI ethics into two intellectual groups: the "philosophy-adjacent disciplines" those in fields such as philosophy, theology, and anthropology, and the "technicians," those in fields such as computer science, electrical engineering, and physics.[109] Graham found areas of agreement and discord between the technicians and philosophy-adjacent disciplines from these recurring themes, as seen in the table below.

Table 1 Number of articles containing discussed themes

Theme	Description	Technicians	
Existential risk is priority	Priority for AGI development is managing existential risks it may pose, over and above considering potential benefits	17	26
We have a moral obligation to future	AGI development should presume unborn humans and future civilization have substantive moral standing today	14	19
Conscious AGI is concern	Developing machine intelligence may necessarily bring about machine sentience, moral subjectivity, and machine agency	10	9
Discuss issues of humanhood	AGI's developmental interaction with human bodies, minds, and possible enhancements raises questions about our moral identity	25	28
Favor technocracy	Majority of AGI governance requires technical and scientific expertise, rapid decision-making should defer to experts	16	6
Favor democracy	Majority of AGI governance requires democratic public input, decision-making should be deliberative	8	12
Favor utilitarianism	Ethical issues in AGI development should be considered in terms of beneficial or detrimental consequences to humans	12	5
Favor deontology	Ethical issues in AGI development should be judged with respect to obligatory rules, principles or duties	13	4
Moral realist	Moral truths and/or facts objectively exist, and an AGI could plausibly identify them, possibly for the betterment of civilization	10	6

[106] Ross Graham is a PhD Candidate at the Department of Sociology at UC San Diego
[107] Graham, "Discourse Analysis of Academic Debate of Ethics for AGI."
[108] Altman, "Planning for AGI and Beyond."
[109] Graham, "Discourse Analysis of Academic Debate of Ethics for AGI."

Figure 2.4. Technician and Philisophical-Adjacent-Discipline Perspectives (Source: *'Discourse analysis of academic debate of ethics for AGI'*)[110]

The areas of disagreement between philosophy-adjacent-disciplines and technicians are:

1. The role of expertise and democracy in AGI governance
2. Deontological[111] vs utilitarianist approach[112]
3. The possibility of moral realism.

When it comes to the role of expertise and democracy in AGI governance, according to Graham, technicians emphasize limitations to democratic deliberation for AGI governance, cautioning that those who are not experts in the technology will be unable to provide sufficient guidance, increasing noise over signal.[113] This perspective couples well with neoliberal political approaches adopted in the 1970s, which asks for less governmental, i.e. non-expert, engagement in private market affairs. It also provides support for the solutionist openness to disruption and risk-taking, with assurance that those who are developing are expert enough to solve problems without external intervention.

For deontological vs utilitarianist approach, technicians note that there are certain limits to a deontological approach, as rules coded into an AGI could contradict or undermine other rules.[114] However, technicians see this limitation as not an absolute barrier. For technicians, they see in this approach room for reversibility, where problems can be reserved, or solved, via the structures of the code. Thus, there is an inclination towards preferring this approach instead of a utilitarian one by the technicians.[115] This perspective emphasizes the solutionist belief that technology can solve all problems, even if it means solving problems with technology with more technology.

[110] Graham.
[111] "Ethical issues in AGI development should be judged with respect to obligatory rules, principles, or duties."
[112] "Ethical issues in AGI development should be considered in terms of beneficial or detrimental consequences to humans."
[113] Graham, "Discourse Analysis of Academic Debate of Ethics for AGI."
[114] Graham.
[115] Graham.

Finally, for the possibility of moral realism technicians believe it is possible for a system to eventually be developed which can mull over and considering moral, ethical, and political answers to humanity's biggest questions by itself and find absolute truths.[116] Currently, a lot of AGI development research focuses on neural networks, with the hope that a human brain artificially created could amend and improve itself with its data stores in ways not possible for us, and through this it could solve humanities problems and find truths.[117] Solutionists belief in the upgrading of humanity can be thought to be influenced by this technician belief and approach.

Graham's discourse analysis on the approach that technicians take to ethics and ethical problem-solving further embeds the solutionist mindset developed since the 1970s. Solutionists are not only informed by the socio-economic environment which has emphasized profit drive, neoliberal regulation, and positioning of the private market as responsible for public/social good, but by the experts who develop their technologies. Through the areas of discord, Graham highlights between technicians and philosophy adjacent disciplines, we can understand that these influences have ethical implications.

The Solutionists and Big Data Innovation

Now we can answer the question of how the digital and capitalist developments from the 1970s to today influence the symbolic order of big data innovation leaders. From Nachtwey, Seidl, and Graham, we can see that these leaders perceive themselves as solutionists and are further affirmed in their belief by the technicians which surround them in their industry. The incoming "big techification of everything?" that Hendrikse, et al. hypothesize is actually the solutionist response to societies ills. The solutionist says, yes, indeed, the big techification of everything, this would make the world a better place.

The primary research question[118] of this thesis inquires on the ability of digital innovation leaders to integrate ethical considerations into their development processes. We now know that the utopian world that these leaders think they will develop with their

[116] Graham.
[117] Graham.
[118] How has digital capitalism impacted the ability of businesses to build ethical big data technologies?

technologies comes from their solutionist mindset. We also know that their ethical considerations are informed by technician and historical socio-economic influence. Up until now, we have remained broad for the "digital innovation leader." It will be more poignant to tangibly see the impact of the solutionist mindset on the ability to integrate ethical considerations into big data innovations by focusing on a specific case. In the introduction, I mentioned OpenAI, Cohere, and Google DeepMind as companies which have recently come forward as companies attempting to ensure that AI is developed for the good of humanity. I also mention that it is OpenAIs which sparked the moral and ethical debates surrounding Artificial Intelligence. This happened following their public release of the AI Chatbot ChatGPT-3 in late 2022. Therefore, the specific case study of this thesis of a solutionist leader and their ethical decision making will be Sam Altman and OpenAI. Before we move on, we briefly must confirm: is OpenAI solutionist in its values?

In the article "Introducing OpenAI" published in 2015, OpenAI says that their "...goal is to advance digital intelligence in the way that is most likely to benefit humanity as a whole." Further elaborating that they "…believe AI should be an extension of individual human wills and, in the spirit of liberty, as broadly and evenly distributed as possible." They also recognize that "…the outcome of this venture is uncertain, and the work is difficult, but we believe the goal and the structure are right."[119] The underlined parts of their introduction strongly exhibit solutionist values. They share their goal of solving key social problems, using technology to upgrade and enable humans/humanity, being open to disruption and failure/risk taking, and technology ultimately being able to create win-win solutions for all.

With all that we have explored in mind, we are now able to explore the case study of OpenAI. In the next chapter we will answer the sub-question(s): **what is OpenAI and how should we interpret their ethical intentions?**

[119] Brockman, Sutskever, and OpenAI Team, "Introducing OpenAI."

Chapter 3: OpenAI - the Company, the Products, the Intentions

OpenAI the Company and it's Products

OpenAI was founded in 2015.[120] The founders include Trevor Blackwell[121], Vicki Cheung[122], Andrej Karpathy[123], Durk Kingma[124], John Schulman[125], Pamela Vagata[126], and Wojciech Zaremba.[127]Advisors to the founding group include Pieter Abbeel[128], Yoshua Bengio[129], Alan Kay[130], Sergey Levine[131], and Vishal Sikka.[132] The co-chairs were Sam Altman[133] and Elon Musk.[134]The funders of OpenAI, with a $1 billion commitment, include Reid Hoffman[135], Jessica Livingston[136], Peter Thiel[137], Amazon Web Services (AWS), Infosys, and YC Research.[138] With our critical science approach and decolonial framework in practice, we should note that of the 12

[120] Brockman, Sutskever, and OpenAI Team.
[121] Trevor Blackwell is a computer programmer, engineer, entrepreneur, and roboticist. He holds a B.Eng from Carleton University and PhD in Computer Science from Harvard University.
[122] Vicki Cheung is currently the Co-Founder/CTO of Gantry. She holds a B.S. in Computer Science from Carnegie Mellon University.
[123] Andrej Karpathy is the former Senior Director of Artificial Intelligence at Tesla. He holds a PhD in Computer Science from Stanford University.
[124] Durk Kingma is a Research Scientist at DeepMind. He holds a M.Sc. in Computer science from Utrecth University and a Ph.D. in Machine Learning from the University of Amsterdam
[125] John Schulman is a Research Scientist at OpenAI. He holds a PhD in Computer Science from UC Berkeley
[126] Pamela Vagata is a Founding Partner of pebblebed. She holds a BA in Computer Science and Mathematics from the University of Washington
[127] Wojciech Zaremba is a Research Scientist ad OpenAI. He holds a PhD in Deep Learning from New York Univresity
[128] Pieter Abbeel is the co-founder, president, and chief scientist and covariant.ai and professor in Artificial Intelligence and Robotics at University of California, Berkely
[129] Yoshua Bengio is a full professor at Université de Montréal and head of the Montreal Institute for Learning Algorithms
[130] Alan Kay is a computer scientist. He holds B.S. in mathematics and molecular biology from University of Colorado Boulder, a M.S. in electric engineering from University of Utah College and a Ph.D. in computer science from University of Utah College
[131] Sergey Levine is an Associate Professor in the Department of Electrical Engineering and Computer Sciences at University of California Berkley
[132] Vishal Sikka is the CEO of Vianai Systems Inc. He holds a Ph.D. in Computer Science from Stanford University
[133] Sam Altman is an entrepreneur and investor. He is currently CEO of OpenAI
[134] Elon Musk is the founder, chairmain, CEO and CTO of SpaceX. He holds a B.A. in Economics and B.S. in Physics from the University of Pennsylvania
[135] Reid Hoffman is an internet entrepreneur, venture capitalist, podcaster, and author. He holds a B.S. in Symbolic Systems and Cognitive Science from the University of Stanford and a M.S. in Philosophy from Wolfson College, Oxford
[136] Jessica Livingston is the founding partner of Y Combinator and author. She holds a B.A. in English from Bucknell University
[137] Peter Thiel is an entrepreneur, venture capitalist, and political activist. He holds a B.A. in philosophy and J.D. from Standford University
[138] Brockman, Sutskever, and OpenAI Team, "Introducing OpenAI."

founders and funders of OpenAI, 9/12 are technicians, and 12/12 are educated in either the United States, Europe, or United Kingdom. From Graham's study which we explored in the previous chapter, we now know that the ethical considerations of these founders likely come from a technician perspective which falls more in line with solutionist framing. For our decolonial observation, we should note that all come from western perspectives as well. Therefore, their approaches are framed from financialized markets which encouraged the big techification of everything for the sake of uplifting society. Their ideas and innovations are likely to be tailored for western framework and society as well as a result of this approach.

However, the framework that the founders and funders of OpenAI set up in their charter is one which focuses on "positive human impact" to "build value for everyone rather than shareholders."[139] Their initial charter in 2018 had four broad goals:[140]

1. **Broadly Distributed Benefits**: Avoid enabling uses of AI or AGI that harm humanity and to always diligently act to minimize conflicts of interests among employees and stakeholders.
2. **Long-Term Safety**: Focus on research required to make AGI safe and in the case that a value-aligned, safety-conscious project comes close to building AGI before them, a commitment to stop competing and start assisting with that project.
3. **Technical Leadership**: Address AGIs impact on society and lead in areas directly aligned with mission and expertise.
4. **Cooperative Orientation**: Actively cooperate with other research and policy institutions and provide public goods that will help society navigate the path to AGI.

The emphasis on humanity being helped through these groundbreaking technologies and point in emphasizing technical, i.e. technician, leadership are examples of solutionist mindset and technician influence. We also see a divergence through the emphasis of not focusing on profit, which removes the solutionist profit embracing motif. However, this profit divergence was short lived. In 2019, OpenAI announced that it would no longer be an

[139] Brockman, Sutskever, and OpenAI Team.
[140] OpenAI, "OpenAI Charter."

exclusively non-profit company.[141] They announced that while their mission is still to ensure that their technologies "benefit all of humanity," the amount of computational power and algorithmic innovations needed to develop them made them decide to scale much faster than initially planned.[142] Therefore, billions of dollars were needed in the coming years to attract and retain talent towards building their AI supercomputers, which they say could not be obtained as a non-profit.[143] Their solution was to create OpenAI LP, a hybrid for-profit and nonprofit, which used a "capped-profit" model.[144] The new structure would work so that OpenAI Inc (the Nonprofit) would govern OpenAI LP.[145] The first round of investors were capped at making back a maximum of 100x their investment, which, according to them, will be continuously lowered for future rounds.[146] According to OpenAI, this structure allowed the funds of OpenAI LP to be allocated towards capabilities, safety, and policy while the funds in OpenAI Nonprofit would be allocated to run educational programs such as Scholars and Fellows and host policy initiatives.[147]

From this structure, OpenAI created a series of products, one of them being ChatGPT. This is the product which I mentioned in my introduction that has sparked conversations around AI. Therefore, the "big data technology" from the primary research question[148] that will be our focus for the rest of case study will be ChatGPT. The ChatGPT iteration available to the (paying) market as of March 2024 is ChatGPT4 for $20USD per month; while ChatGPT3.5 is free to use.[149] ChatGPT is a large language model (LLM) trained to produce text and is guided and optimized by Reinforcement Learning with Human Feedback.[150] Understanding what this all means and how it is built is important for our discussion.

[141] Brockman, Sutskever, and OpenAI, "OpenAI LP."
[142] Brockman, Sutskever, and OpenAI.
[143] Brockman, Sutskever, and OpenAI.
[144] Brockman, Sutskever, and OpenAI.
[145] Brockman, Sutskever, and OpenAI.
[146] Brockman, Sutskever, and OpenAI.
[147] Brockman, Sutskever, and OpenAI.
[148] How has digital capitalism impacted the ability of businesses to build ethical big data technologies?
[149] "ChatGPT Pricing."
[150] "What Is ChatGPT?"

From the user-experience side, ChatGPT is a chatbot that can hold conversations and respond to text prompts in a human-like manner.[151] On the back end, ChatGPT(s) are composed of GPT- Generative Pre-Trained Transformers. The GPT is trained on a large amount of data sources, over 150 billion human-generated texts ranging from books, articles, reviews, online conversations, and human-generated data.[152] It is trained by taking all of these sources and making labels and weights for each word (tokenizing).[153] It then undergoes two steps: an "unsupervised" phase, where the model practices predicting the next word in a sentence based on all of the prior data, followed by a "supervised" phase where computer scientists fine-tune the model on a smaller dataset to ensure the quality of its outputs.[154]

When we consider our critical science approach and what we learned in the previous chapter about the technicians' approach to ethics, we can understand the crucial importance of the perspectives of computer scientists. They are the ones who select the data which trains the models, as well as fine-tune their outputs. OpenAI is aware the significance of this, developing a specific approach which they outline in a June 2022 article of seven principles of best practices for LLM development and deployment.[155] The seven principles span three general areas and are made to:[156]

1. Prohibit misuse:
    a. Publishing usage guidelines and terms of use
    b. Build systems and infrastructure to enforce usage guidelines.
2. Mitigate unintentional harm:
    a. Proactively mitigate harmful model behavior
    b. Document known weaknesses and vulnerabilities.
3. Thoughtfully collaborate with stakeholders:

[151] "What Is ChatGPT?"
[152] Roumeliotis and Tselikas, "ChatGPT and Open-AI Models."
[153] Roumeliotis and Tselikas.
[154] Roumeliotis and Tselikas.
[155] OpenAI, "Best Practices for Deploying Language Models."
[156] OpenAI.

a. Build teams with diverse backgrounds, publicly disclose lessons learned regarding LLM safety and misuse, and treat all labor in language model supply chain with respect.

The framework of the company and their principles for developing ChatGPT might make one think that the answer to the primary research question[157] is that digital capitalism does not inherently inhibit businesses developing big data technologies from integrating ethical considerations into the development of these technologies. Seemingly, OpenAI is taking steps to ensure that their LLM ChatGPT at a minimum is ethical. The key word here is seemingly.

In October of 2023, the Center for Security and Emerging Technology (CSET) released the article '*Decoding Intentions: AI and Costly Signals'*. The article was written by Andrew Imbrie[158], Owen J. Daniels[159], and Helen Toner.[160] Helen Toner herself served on the board of OpenAI Inc. The purpose of '*Decoding Intentions*' was to give policymakers a guide to better understand what AI-developing actors are doing regarding safety considerations as they develop their technologies.[161] The paper focuses on finding the 'intentions' of the actors through the 'signals' they portray.[162] They look at signaling within three use cases: Military AI and Autonomy, Democratic AI, and Private Sector AI. One of the case studies for Private Sector Signaling is OpenAI. Throughout this study and in their conclusion, they postulate that presence of specific signals which they outline can be utilized to interpret the intention and integration of safety by the developers of AI technologies.[163]

Due to my description of what ChatGPT is, I stopped short on the history of OpenAI. To continue, after the company shifted from non-profit to a mixed for-profit and non-profit company in 2019, things remained stable for four years. However, on November 17th, 2023, the

[157] How has digital capitalism impacted the ability of businesses to build ethical big data technologies?
[158] Andrew Imbrie is an Associate Professor of the Practice in the Gracias Chair in Security and Emerging Technology at Georgetown University's School of Foreign Service
[159] Owen J. Daniels is the Andrew W. Marshall Fellow at Georgetown's Center for Security and Emerging Technology
[160] Helen Toner is the Director of Strategy and Foundational Research Grants at Georgetown's Center for Security and Emerging Technology and also served on OpenAIs nonprofit board until November of 2023
[161] Imbrie, Daniels, and Toner, "Artificial Intelligence and Costly Signals."
[162] Imbrie, Daniels, and Toner.
[163] Imbrie, Daniels, and Toner.

board of directors of OpenAI Inc, of which Toner was a part of, announced that Sam Altman would depart as CEO as well as leave the board of directors.[164] The announcement said that Sam Altman's departure followed a "deliberative review process by the board, which concluded that he was not consistently candid in his communications with the board, hindering its ability to exercise its responsibilities" which left the board with no "confidence in his ability to continue leading OpenAI."[165] Over the course of the weekend that followed this announcement Sam Altman was hired by Microsoft, a petition was signed by 730 employees at OpenAI which demanded the resignation of the board and the reinstatement of Sam Altman as CEO or else they would follow him to Microsoft, the re-hiring of Sam Altman by OpenAI, and the prompt dissolution of the board of OpenAI Inc, including Toner.[166] All of this happened one month following the publication of '*Decoding Intentions'*. Toner's departure therefore signals instead a potential misalignment between those decoding the ethical approaches of the leaders involved in development decisions of big data technologies as well as the perspectives of the developers they are leading. This is why the primary research question of this thesis still stands as is. To ensure that similar errors are not repeated in our analysis of Sam Altman and OpenAI and their ethical decision making, let's analyze '*Decoding Intentions'*, first.

Decoding OpenAI's Leadership

In the executive summary of the '*Decoding Intentions'*, the framing of its research is built on a similar premise to this thesis – wanting to understand the intentions of AI developers and how this impacts their outputs and society at large. The paper's framework is specifically developed for policy makers. As policy makers are not experts in the development of the technologies themselves, they look for guidance to understand how to navigate conversations with those developing and using AI. This paper gives policy makers needed framework to understand if developers have malicious intent or not and understand if development risks are being taken which require regulatory intervention. Imbrie et al. therefore decode the 'signals' of developers through various use cases to provide tools to policy makers for understanding and

[164] OpenAI, "OpenAI Announces Leadership Transition."
[165] OpenAI.
[166] Cao, "A Timeline of OpenAI's (Ongoing) Drama With Sam Altman."

making regulatory responses to AI developing actors. The strength of their framework is therefore critical.

To put our decolonial and critical science approach into practice; as highlighted in the introductory footnote on the last page, we see that all three of these researchers come from philosophy-adjacent disciplines. This is significant, as we have already learned that there are misalignments between ethical approaches of technicians and philosophy-adjacent-disciplines. A scholar from a philosophy-adjacent-discipline could incorrectly interpret a signal by making the wrong conclusion of what will and will not be done due to their scholarly perspective. However, Imbrie et al do come from institutions positioned in the west. Here this can positively impact their signal interpretation, because they have an academic and personal context for the socio-economic histories which have shaped the approaches of tech leaders. One limitation from this is that the cultural and economic context to which these signals are relevant may only be for actors coming from this western perspective. This tool may not be as useful for policy makers trying to understand the signals of powerful companies developing big data technologies from countries such as China which have produced leaders from differing socio-economic contexts. This being said, let us analyze the signals which the authors outline in their article.

The authors identify four kinds of "Costly AI Signals" that AI developers produce: 'tying hands,' 'installment costs,' 'sunk costs,' and 'reducible costs.'[167] The verbiage of "cost" is not by accident – as we've been made aware, costs, profits, and the economics of developing these digital innovations is very important for the leaders developing these technologies. We have seen this in the solutionist framing of profit along with the example of OpenAI's attempt to deviate by starting as non-profit yet having to eventually shift all of its development activities underneath a for-profit umbrella. Thanks to their personal and professional expertise, Imbrie et al. are aware of the importance of money in the development of these technologies, and hence decide to center literal costs as the signals. The assumption being that the financial cost it takes

[167] Imbrie, Daniels, and Toner, "Artificial Intelligence and Costly Signals."

to integrate an ethical approach into the development of a technology is enough to interpret the intent of and societal impacts by those developing AI technologies.

The first costly signal, tying hands, is described as the "strategic deployment of public commitments before a foreign or domestic audience."[168] Put in other words, tying hands describes social contracts between companies and the general public. The idea is that a company which does not meet expectations it publicly is committed to, would suffer from the negative public perspective. This would literally cost them in PR damage control, firing and rehiring of executives, and so on. The example they give for the private sector is "a CEO could face disciplinary actions from the board of directors or reputational costs to the companies brand that can result in lost market share."[169] The "cost" happens when the public (and private) expectations are unmet.[170]

However, when we consider the solutionist approach and their placement within society, typing hands may not be as strong of a signal. Firstly, a solutionist has in their ethos the acceptance of 'disruption'. Disruption in itself accounts for the chaos of public pushback. They are also assured by the Sisyphean cycle of technological panics. Public outcry and fear are not unfamiliar to the innovator, and with their ultimate goals, it can be interpreted by them as a temporary inconvenience for the greater good. It should also be pointed out that the specific example given by Imbrie et al. happened exactly one month after the publishing of this article. What actually happened was the reversal of this disciplinary action and reputational costs to the board of directors themselves.[171]

Installment costs are "fixed costs that cannot be recouped over time."[172] An example of an installment cost by the private sector could be pledges by companies to conduct risk

[168] Imbrie, Daniels, and Toner.
[169] Imbrie, Daniels, and Toner.
[170] Imbrie, Daniels, and Toner.
[171] Cao, "A Timeline of OpenAI's (Ongoing) Drama With Sam Altman."
[172] Imbrie, Daniels, and Toner, "Artificial Intelligence and Costly Signals."

assessments of AI models and ensure those results are available to the public.[173] Costs here actively happen throughout the development stage.

Installment costs also run the risk of misinterpretation. Here, we can think about the areas of discord between technicians and philosophy-adjacent-disciplines. As we learned, technicians prefer a deontological approach. Their preference of the deontological approach manifests in the idea that issues can be responded to and fixed as they come about, due to their belief in the reversibility of problematic codes.[174] This, tied in with the solutionist acceptance of failure, comfort for disruption, and belief that technology can be used to solve even technological issues, means that these risk assessments may be developed to react to problems on a rolling basis rather than to proactively prevent issues. When we consider the speed at which new technologies appear, integrate, and influence society, this means that a reactive approach leaves space for negative ethical consequences to happen in the meantime. Even in instances of proactive risk assessments, we have also learned about the misalignment between philosophy-adjacent-disciplines and technicians approaches to ethics, one example being the possibility of moral realism. A philosophy-adjacent-discipline scholar could assume that these proactive risk assessments might include ensuring that an AI or AGI doesn't spread 'moral facts' that risk propagating neo-colonial thought. Yet, technicians believe in moral realism and the ability of advanced technologies in finding them. Therefore, installment costs may be unreliable due to misalignment not only for when precisely something such as a risk assessment would occur on an innovation development timeline, but also the parameters and reasoning for the tool in the first place.

Sunk costs are described as "commitments whose costs are priced from the start…which involves public commitments that are only costly in the event of noncompliance."[175] An example given for the private sector is a company committing large-scale investments for test and evaluation and not following through on utilizing them in their development process.[176] Put

[173] Imbrie, Daniels, and Toner.
[174] Graham, "Discourse Analysis of Academic Debate of Ethics for AGI."
[175] Imbrie, Daniels, and Toner, "Artificial Intelligence and Costly Signals."
[176] Imbrie, Daniels, and Toner.

simply, if a company invests in something and ultimately does not utilize it, then it is a waste of money, this is the 'cost'.

Reducible costs are "paid upfront but can be offset over time depending on the actions of the signaler."[177] An example they give is a private sector company investing in more interpretable AI models, which incentivizes information sharing.[178] Costs here happen if the actor doesn't entirely pull through with their efforts, with their investment being a financial waste through not taking full advantage of these investments by integrating them into their operations.

The nature of the technology industry and the solutionist approach limits the efficacity of both the sunk cost and reducible cost signals. The comfort with risk and failure which exists in the solutionist mindset translates to comfort with financial risk and failure. Trillions of dollars exist in the technology industry alongside a high failure rate: more than two-thirds of tech start-ups never deliver a positive return to investors.[179] While those in philosophy-adjacent-disciplines may see high stakes in financial investment towards research and operations due to scarcity in funding, technicians and their solutionist leaders come from a sector deeply imbedded in Venture Capital cash. A solutionist leader may have already personally lost millions towards other projects, and therefore the simple fact of investing in research focused on something like interpretable AI models does not hold enough weight to be a signal of ethical intention. Of course, investing in such a thing in the first place is indeed a good sign and a good start. However, even this example has limitations. As mentioned earlier, those involved in ethical AI conversations highlight the importance of ensuring that these systems are less opaque and more transparent. However, we should think about who is actually able to utilize interpretable AI models. These are the technicians, and what exactly are their needs are for interpretability may not inherently imply ethicality. Therefore, transparency through information sharing might not even be for what philosophy-adjacent-scholars think it's for.

[177] Imbrie, Daniels, and Toner.
[178] Imbrie, Daniels, and Toner.
[179] Eisenmann, "Why Start-Ups Fail."

Relying on signals to interpret whether a technology company is making decisions with ethical outcomes in mind may not be the best method. 'Signals' are simply a *perception* of what an actor developing or using an AI is doing. This perception is only as good as the efforts to develop a robust understanding of an actors perspective and the academic background of the interpreter. The utilization of frameworks like the critical science approach[180] to contextualize computer scientists and their approaches and Symbolic Order[181] to understand the value system and approaches of tech leaders allowed us to see gaps in the signals which the authors of '*Decoding Intentions'* propose. This is why time was taken in the previous chapters to thoroughly understand the individuals developing and making the decisions for these technologies. From these efforts, analysis can be made not based off of 'signals', which may be misread due to our own academic, cultural, or social backgrounds, but rather through frameworks which outline how and why choices are approached. Thus, we will finally wrap up our exploration of the overall research question with a discourse analysis of Sam Altman and OpenAI. By carefully looking at what they say through our theoretical framings, we will be able to interpret their alignment with ethical approaches in the development for their technologies. Through this analysis, we can see how path dependencies greatly influence the ways that tech leaders make their decisions today.

[180] Critical Science Approach: The societal and personal academic lens which influences a scholars stance.
[181] Our use of social and cultural environmental feedback to frame and understand our placement within these worlds.

Chapter 4: Discourse Analysis of OpenAI and Sam Altman

To do the discourse analysis, I will be utilizing a mixed methods approach, utilizing the GloVe model for word embeddings alongside a qualitative analysis. I have selected these two methods because it will allow me to qualitatively analyze a large amount of text (about two dozen interviews spanning from 20-minutes to two hours) through a quantitative supplement. Each interview was transcribed using transcribing software[182], which were then manually checked by myself for spelling mistakes or misquotes one by one. This manual check was done by reading along while listening to the interviews and correcting where needed. The interviews selected were all of the publicly available and easily accessible interviews which Sam Altman did between 2015-2024. I also built a corpus from blog posts published on OpenAIs website during this time period that were categorized under Responsible AI and Safety & Alignment which referred to AI and AGI. The corpuses from the interviews and blog posts were split into different time frames for the word embedding analysis. This was done so that change of sentiments over time were caught, as well as change in themes of discourse throughout his years at OpenAI. The analysis of these texts will provide a thorough understanding of Sam Altman's and OpenAI's ethical consideration in their development decisions.

Word Embedding to find Sentiment

Word embeddings are a machine learning framework in which English words are represented by a vector.[183][184] Word embedding consists of utilizing geometric equations to place words in a corpus in relation with all other words, ultimately finding the semantic relationships between words.[185][186] Saif Mohammad lays out the best practices for utilizing word

[182] Good Tape.io
[183] Garg et al., "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes."
[184] Nikhil Garg is an assistant professor of Operations Research and Information Engineering (ORIE) at Cornell Tech and the Technion, and part of the Jacobs Technion-Cornell Institute, and an ORIE, Computer Science, and Information Science field member at Cornell University. Londa Schiebinger is the John L. Hinds Professor of History of Science in the History Department at Stanford University and Director of the EU/US Gendered Innovations in Science, Health & Medicine, Engineering, and Environment Project. Dan Jurafsky is a Reynolds Professor in Humanities, Professor of Linguistics and Professor of Computer Science at Stanford University. James Zou is an Associate Professor of Biomedical Data Science and, by courtesy, of Computer Science and Electrical Engineering at Stanford University.
[185] Mendelsohn, Tsvetkov, and Jurafsky, "A Framework for the Computational Linguistic Analysis of Dehumanization."

embeddings to find sentiment, mentioning how word embeddings can be useful to find connotations, implicit emotions, and socio-culture biases.[187][188] Garg et al. utilize word embeddings to quantify 100 years of gender stereotypes, by using a corpus of text data from the US Census to show demographic, occupation, and societal shifts over time.[189] Mendelsohn, Tsvetkov and Jurafsky in their study, utilize word embeddings to study the progression of dehumanizing language in the New York Times between 1986 to 2015 towards LGBTQ people, ultimately finding that over time LGBTQ people were increasingly humanized.[190]

Word embeddings are thus a useful tool for finding the sentiments and themes that are connected to certain words. In some research, scholars will train their own models to find distances between words, such as the study by Mendelsohn, Tsvetkov, and Jurafsky. There already exists pre-trained models, which are also useful for word embeddings research. In my research, I utilized the pre-trained GloVe model. This model is trained by Jeffrey Pennington[191], Richard Socher[192], Christopher D. Manning[193] and can be used to represent linear substructures of word vector space, or in other words, the similarity/closeness of words. The advantage of using pre-trained models means less coding work for researchers who are not exclusively trained in the computer science field, such as myself, while still having the ability to find meaningful connections between words when following the guidelines and examples from scholars such as Mohammad. The model that was used specifically in this research was *Common Crawl (42B tokens, 1.9M vocab, uncased, 300d vectors, 1.75 GB download.*[194]

---

[186] Julia Mendelsohn is a PhD candidate at the University of Michigan School of Information. Yulia Tsvetkov is an associate professor in the Paul G. Allen School of Computer Science & Engineering, at the University of Washington.
[187] Mohammad, "Practical and Ethical Considerations in the Effective Use of Emotion and Sentiment Lexicons."
[188] Saif Mohammad is a Principal Research Scientist at the National Research Council Canada (NRC)
[189] Garg et al., "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes."
[190] Mendelsohn, Tsvetkov, and Jurafsky, "A Framework for the Computational Linguistic Analysis of Dehumanization."
[191] Jefferey Pennington is a research scientist at Google NY. He holds a PhD in Theoretical Particle Physics from Stanford University
[192] Richard Socher is CEO at you.com and founder of AIX Ventures. He holds a PhD in Computer Science from Stanford University
[193] Christopher Manning is the Thomas M. Siebel Professor in Machine Learning, Professor of Linguistics and of Computer Science Director, Stanford Artificial Intelligence Laboratory (SAIL)
Associate Director, Stanford Institute for Human-Centered Artificial Intelligence (HAI)
[194] Pennington, Socher, and Manning, "GloVe: Global Vectors for Word Representation."

Breakdown of Corpus

As mentioned, Sam Altman interviews between 2015-2024 and discourse directly from OpenAI's blog of all articles classified under *Safety & Alignment* and *Responsible AI* published within this same timeframe were utilized as my corpuses. As we know that philosophy-adjacent disciplines and those in the technician disciplines have different approaches when it comes to ethical approaches, I utilized articles published by philosophy-adjacent disciplines about ethical AI as a "control". The articles selected are those utilized in the literature reviews of this paper along with articles read over the course of the research phase of this master's thesis. They are all found in the bibliography.

The corpus of Sam Altman's discourse was split into three eras. First, 2015-2022, which is when OpenAI was founded, and his involvement was officially announced.[195] The next era was January 2023-November 2023, when Sam Altman spent a year touring the world to meet with various leaders to speak about OpenAI's mission.[196] Finally, the third era is from November 2023-March 2024, which follows the firing and rehiring of Sam Altman by the previous non-profit board as well as the installment of the temporary non-profit board.[197]

OpenAI's corpus comes from every blog post between 2015-2024 classified under *Safety & Alignment* and *Responsible AI.* The Wayback Machine was utilized to capture articles that were previously published about ethical AI which are no longer present on the website. This corpus was broken down into two eras, pre–Sam Altman's firing and re-hiring (2015-November 2023) and post this event (December 2023-March 2024).

Selection of Target Words

For my word embedding, I decided to use the pre-trained model to find the nearest neighbors of a selection of words that are closely related to discussions surrounding ethical AI construction. These words were selected by finding the most used words within the AI ethics

[195] Brockman, Sutskever, and OpenAI Team, "Introducing OpenAI."
[196] Iyengar, "OpenAI's CEO Goes on a Diplomatic Charm Offensive."
[197] OpenAI, "OpenAI Announces Leadership Transition"; OpenAI, "Sam Altman Returns as CEO, OpenAI Has a New Initial Board"; OpenAI, "Review Completed & Altman, Brockman to Continue to Lead OpenAI"; OpenAI, "OpenAI Announces New Members to Board of Directors."

discourse articles and words that frequently came up during interviews with Sam Altman. The words initially selected were: "ai", "agi", "consequence", "consequences", "danger", "dangers", "ethic", "ethical", "ethics", "governance", "human", "humans", "humanity", "intelligence", "intelligent", "language model", "language models", "moral", "negative", "positive", "regulation", "regulations", "risk", "risky", "safe", "safety", "society." With these words in mind, I ran four scripts in RStudio to find:

1. The most used words for each of the corpuses.
2. Find the frequency of the target words for each of the corpuses.
3. Find the nearest neighbors for each of the corpuses.
4. Find the nearest neighbors for each of the corpuses with target words that are synonymous with Sam Altman himself, OpenAI as an entity, and ChatGPT itself.

To ensure rigorous analysis, I reflected on the size of each corpus. The breakdown of characters, words, and density of each corpus is as follows:

Document Title	Characters	Words	Lines	Average Word Length
Sam Altman 2016-2022	163,523	30,790	1,078	5.31 characters
Sam Altman 2023	142,593	26,728	1,466	5.33 characters
Sam Altman Nov 2023-2024	121,465	23,138	1,085	5.25 characters
Ethical Artificial Intelligence Articles	193,855	29,756	2,882	6.51 characters
OpenAI Discourse 2016-2022	228,755	35,285	2,141	6.48 characters
OpenAI Discourse 2023-2024	111,270	17,291	971	6.44 characters

Figure 4.1 Character Breakdown of Corpus. (Source: My own calculations)

Because the corpuses do vary in size, scripts two through four were normalized to account for the differences in size. The normalization was done so that results were between 0 to 1 to find the weight of word outputs.

Results #1 – Most Used Words for Each Corpus

2016-2022 Sam		2023 Sam		2023-2024 Sam		Ethics Discourse		OpenAI 2016-2022		OpenAI 2023-2024	
522	Think	505	Think	356	Think	227	ai	329	ai	229	ai
240	People	202	People	200	Know	186	ethical	283	models	112	will
198	Just	187	Know	163	People	141	agi	234	model	107	models
177	Know	125	Will	124	Dont	113	moral	186	language	105	systems
155	Will	122	Just	117	Just	106	systems	156	will	91	safety
143	Things	112	Dont	106	Things	89	human	153	systems	59	model
143	Sort	112	Things	98	Thats	85	intelligenc	141	research	51	work
118	Dont	92	Theres	94	Will	84	ethics	122	developm	46	content
112	Thats	91	Thats	88	Want	67	will	117	safety	46	frontier
107	Time	88	Sort	78	Time	65	system	115	companie	45	make

Figure 4.2 Most used Words for Each Corpus. (Source: My own calculations)

The results from my first analysis in RStudio is seen in the table above. The most common words utilized by Sam Altman are quite consistent over the course of 2016-2024. These corpuses all show a high utilization of the words: think, people, just, know, will, things, don't, and that's. It is also noteworthy that the word "time" appears most frequently used for Sam Altman in his most recent discourse.

For the most frequently used words in the scholarly articles, it is unsurprising that words such as "ai" "ethical" and "agi" are the most frequently used words.

For OpenAI, the most frequently used words across both corpuses are ai, will, models, model, systems and safety.

Across all corpuses, beyond the common word of AI, which is expected, there is, not much overlap in the most frequently used words. The next word that is aligned between the corpuses is will. The final most common word, present in three out of the six corpuses in terms of frequency is the word systems. This word is shared in being most frequent for journal articles as well as blog publications and articles written by OpenAI during the two eras analyzed for this paper.

Results #2 – Frequency of Target Words

The table below shows the frequency of the target words:

Sam 2016-2022	
ai	78
agi	35
human	32
intelligence	29
risk	23
society	17
humans	15
humanity	14
governance	6
negative	6
safety	5
consequences	4
intelligent	3
positive	3
risky	3
safe	3

Sam 2023	
ai	85
society	29
human	28
agi	26
safety	24
humans	19
intelligence	17
humanity	10
regulation	7
risk	7
safe	7
positive	4
risky	4
governance	3
danger	2
moral	2
negative	2
consequences	1

Sam 2023-2024	
ai	46
agi	32
society	15
intelligenc	11
human	10
safety	8
governanc	7
humanity	7
humans	7
regulatior	4
negative	3
positive	2
risk	2
conseque	1
conseque	1
safe	1

Ethics Discourse	
ai	253
ethical	194
agi	147
moral	118
human	109
intelligenc	103
ethics	96
humans	41
risk	27
conseque	17
governanc	17
humanity	15
intelligent	11
society	9
positive	6
negative	5
regulatior	4
regulatior	4
safety	4
conseque	1
danger	1
ethic	1

OpenAI 2016-2022	
ai	334
safety	124
human	105
humans	46
risk	32
regulatior	28
safe	28
negative	27
agi	24
positive	24
society	19
intelligenc	13
humanity	7
risky	5
conseque	3
ethical	3
governanc	3
dangers	1
ethics	1

OpenAI 2023-2024	
ai	254
safety	93
agi	41
society	30
safe	27
human	25
risk	23
humanity	19
governanc	18
humans	14
intelligenc	7
regulatior	5
regulatior	3
positive	2
moral	1
risky	1

Figure 4.3 Most used Words for Each Corpus. (Source: My own calculations)

Because the corpuses vary in size, the frequencies were normalized to account for the scale differences within each corpus set. This normalization allows us to meaningfully compare the frequencies of the target words, track frequency of mention by a single speaker over time, as well as parse out the weight on the thematic focus of the discourse based on its frequency of use.

	Sam 2016-2022
ai	1
agi	0.448
human	0.41
intelligence	0.372
society	0.295
humans	0.218
humanity	0.193
governance	0.179
negative	0.077
safety	0.064
consequences	0.0513
intelligent	0.385
positive	0.385
risky	0.385
safe	0.385

	Sam 2023
ai	1
society	0.341
human	0.329
agi	0.306
safety	0.283
humans	0.2
intelligence	0.118
humanity	0.084
regulation	0.084
risk	0.084
safety	0.047
positive	0.047
risky	0.353
governance	0.024
danger	0.024
negative	0.024
consequences	0.012

	Sam 2023-2024
ai	1
agi	0.7
society	0.326
intelligence	0.239
human	0.217
safety	0.174
governance	0.152
humanity	0.152
humans	0.152
regulation	0.09
negative	0.07
positive	0.043
risk	0.043
consequence	0.022
consequences	0.022
safe	0.022

Figure 4.4 Most used words for Sam Altman (Source: My own calculations)

	Ethics Discourse
ai	1
ethical	0.767
agi	0.581
moral	0.466
human	0.43
intelligence	0.407
ethics	0.379
humans	0.162
risk	0.106
consequences	0.067
governance	0.067
humanity	0.059
intelligent	0.043
society	0.036
positive	0.024
negative	0.02
regulation	0.02
regulations	0.02
safety	0.02
consequence	0.004
danger	0.004
ethic	0.004

	OpenAI Discourse 2016-2022
ai	1
safety	0.371
human	0.314
humans	0.138
risk	0.096
regulation	0.084
safe	0.084
negative	0.081
agi	0.072
positive	0.072
society	0.057
intelligenc	0.039
humanity	0.021
risky	0.015
consequer	0.001
ethical	0.001
governanc	0.001
dangers	0.003
ethics	0.003

	OpenAI Discourse 2023-2024
ai	1
safety	0.366
agi	0.162
society	0.118
safe	0.106
human	0.098
risk	0.091
humanity	0.075
governance	0.071
humans	0.055
intelligence	0.028
regulation	0.02
regulations	0.012
positive	0.008
moral	0.004
risky	0.004

Figure 4.5 Most used words for PAD and Open-AI (Source: My own calculations)

As can be seen by the frequencies, not all words that were targeted in this frequency analysis were mentioned in all corpuses. The Ethics Discourse literature includes the most amount of target words processed.

Results #3 – Nearest Neighbors

The original analysis of nearest neighbors looked at our 25 target words. To respect the amount of space I have for this thesis, I will be sharing the top 5 words, besides AI, across all data sets in terms of weight. The words with the heaviest weight in the discourse are: agi with a

total weight of 1.821, human[198] with a total weight of 1.388, safety with a total weight of 1.261, society with a total weight of 0.878, and finally intelligence with a total weight: 0.831.

Nearest Neighbors for AGI:

NN for 'agi' in '2016-2022.1':	NN for 'agi' in '2023. 2':	NN for 'agi' in '2023-2024':	NN for 'agi' in 'Ethics Discourse':	NN for 'agi' in 'OpenAI Discourse 2016-2022':	NN for 'agi' in 'OpenAI Discourse 2023-2024':
1: income 0.4972539	1: agis 0.3953733	1: average 0.3516672	1: tax 0.4288116	1: income 0.4804619	1: per 0.3801696
2: skill 0.3575388	2: per 0.3730600	2: threshold 0.3389216	2: percentage 0.4090156	2: net 0.4467396	2: adjusted 0.3596132
3: gross 0.3433231	3: level 0.3537919	3: adjusted 0.3375033	3: net 0.3603927	3: total 0.3869676	3: grants 0.3233617
4: per 0.3302104	4: total 0.3368712	4: revenue 0.3349123	4: median 0.3574461	4: gross 0.3516506	4: damage 0.3018669
5: percentage 0.3249374	5: gpt 0.3200606	5: per 0.3217986	5: average 0.3475362	5: percentage 0.3476143	5: limit 0.2900554
6: calculated 0.3215282	6: assets 0.3184780	6: total 0.3199841	6: level 0.3297402	6: adjusted 0.3463827	6: credits 0.2864531
7: trillion 0.3183846	7: damage 0.3104202	7: percentage 0.3075297	7: total 0.3227662	7: percent 0.3409648	7: multiplier 0.2838095
8: net 0.3149776	8: value 0.3072408	8: max 0.3023686	8: = 0.3184843	8: increases 0.3265089	8: level 0.2834772
9: total 0.3092926	9: gdp 0.3056766	9: engagement 0.3020267	9: contributions 0.3136012	9: revenue 0.3201176	9: input 0.2819697
10: earnings 0.3091819	10: values 0.2939989	10: weapons 0.3001038	10: exceeds 0.3109864	10: maximum 0.3146086	10: exceed 0.2729496

Figure 4.6 Nearest Neighbors for AGI. (Source: My own calculations)

Nearest Neighbors for Humanity:

NN for 'humanity' in '2016-2022.1':	NN for 'humanity' in '2023. 2':	NN for 'humanity' in '2023-2024':	NN for 'humanity' in 'Ethics Discourse':	NN for 'humanity' in 'OpenAI Discourse 2016-2022':	NN for 'humanity' in 'OpenAI Discourse 2023-2024':
1: lives 0.7832342	1: humans 0.8959432	1: knowledge 0.7595788	1: world 0.8630529	1: beings 0.8470813	1: personal 0.7444331
2: human 0.7707793	2: conflict 0.7275227	2: human 0.7418705	2: nations 0.7705379	2: society 0.8356809	2: human 0.6813042
3: world 0.7590481	3: ourselves 0.7225416	3: economic 0.7108339	3: existence 0.7445459	3: human 0.7618227	3: future 0.6670352
4: spirit 0.7248906	4: us 0.7202046	4: lives 0.7046043	4: human 0.7350174	4: life 0.7608891	4: life 0.6627695
5: society 0.7073322	5: what 0.7169074	5: society 0.6693359	5: its 0.7226795	5: nature 0.7177457	5: could 0.6381681
6: sacrifice 0.6705235	6: history 0.7057831	6: respect 0.6662671	6: lives 0.7163183	6: future 0.6519258	6: never 0.6358616
7: we 0.6610345	7: nature 0.7035640	7: physical 0.6644845	7: beings 0.7049959	7: that 0.6437078	7: everyone 0.6333794
8: for 0.6562206	8: human 0.6853876	8: citizens 0.6643840	8: toward 0.7032554	8: countries 0.6402521	8: conflict 0.6315153
9: something 0.6516996	9: any 0.6811078	9: our 0.6385101	9: eternal 0.6962421	9: our 0.6325922	9: society 0.6263634
10: relationship 0.6435502	10: lives 0.6710981	10: humans 0.6312997	10: purpose 0.6572133	10: humans 0.6324750	10: reality 0.6163657

Figure 4.7 Nearest Neighbors for Humanity. (Source: My own calculations)

Nearest Neighbors for Safety:

NN for 'safety' in '2016-2022.1':	NN for 'safety' in '2023. 2':	NN for 'safety' in '2023-2024':	NN for 'safety' in 'Ethics Discourse':	NN for 'safety' in 'OpenAI Discourse 2016-2022':	NN for 'safety' in 'OpenAI Discourse 2023-2024':
1: care 0.8088945	1: system 0.8851787	1: health 0.9234118	1: protection 0.9997303	1: called 0.8083389	1: security 0.8214910
2: there 0.7673886	2: any 0.8139997	2: system 0.8266378	2: your 0.7969222	2: security 0.7887857	2: personal 0.8025496
3: are 0.7669325	3: important 0.7850670	3: ensure 0.7711010	3: report 0.7909712	3: that 0.7779210	3: protection 0.7644349
4: risk 0.7526961	4: service 0.7547951	4: impact 0.7708433	4: transportation 0.7869796	4: improve 0.7713092	4: regulations 0.7623292
5: for 0.7508668	5: compliance 0.7502781	5: control 0.7692101	5: medical 0.7812350	5: effective 0.7377004	5: risk 0.7438643
6: general 0.7488231	6: control 0.7399221	6: in 0.7398579	6: should 0.7779088	6: including 0.7334554	6: health 0.7387192
7: education 0.7456120	7: educational 0.7328987	7: information 0.7327463	7: health 0.7720899	7: prevention 0.7219089	7: more 0.7174198
8: quality 0.7358030	8: risk 0.7315766	8: quality 0.7254871	8: necessary 0.7579290	8: provide 0.7143468	8: laws 0.7102603
9: issues 0.7353071	9: technology 0.7277756	9: with 0.7254634	9: practice 0.7504945	9: cost 0.7141975	9: industry 0.7055703
10: test 0.7329307	10: standards 0.7266584	10: efficiency 0.7051834	10: systems 0.7171567	10: health 0.6957939	10: policy 0.6984664

Figure 4.8 Nearest Neighbors for Safety. (Source: My own calculations)

Nearest Neighbors for Society:

NN for 'society' in '2016-2022.1':	NN for 'society' in '2023. 2':	NN for 'society' in '2023-2024':	NN for 'society' in 'Ethics Discourse':	NN for 'society' in 'OpenAI Discourse 2016-2022':	NN for 'society' in 'OpenAI Discourse 2023-2024'
1: world 0.8633821	1: social 0.9151889	1: economic 0.8737866	1: world 0.9433561	1: called 0.8478538	1: association 0.8213307
2: education 0.8310320	2: history 0.8632460	2: health 0.7948867	2: societies 0.8336946	2: cultural 0.7680119	2: more 0.7195822
3: conference 0.7991785	3: system 0.8024491	3: in 0.7894508	3: american 0.8090762	3: that 0.7637234	3: president 0.7096894
4: group 0.7944696	4: educational 0.7843516	4: conference 0.7691786	4: medical 0.7775770	4: national 0.7479252	4: personal 0.7076282
5: which 0.7690137	5: national 0.7650356	5: physical 0.7543749	5: association 0.7506796	5: education 0.7466252	5: political 0.7071729
6: general 0.7547989	6: any 0.7630098	6: system 0.7340332	6: ethics 0.7483348	6: american 0.7288214	6: education 0.7069449
7: there 0.7534116	7: nature 0.7448839	7: religious 0.7334429	7: it 0.7388059	7: members 0.7200852	7: organization 0.6983918
8: institution 0.7495212	8: us 0.7386899	8: knowledge 0.7320068	8: country 0.7373426	8: international 0.7174615	8: social 0.6948316
9: are 0.7414150	9: disease 0.7366092	9: political 0.7232698	9: nations 0.7356229	9: nature 0.7107413	9: meetings 0.6934662
10: for 0.7176598	10: literature 0.7348050	10: history 0.7172335	10: university 0.7217330	10: institute 0.6878869	10: science 0.6848458

Figure 4.9 Nearest Neighbors for Society. (Source: My own calculations)

[198] Human has a heavy weight even though it was not a target word. It is heavily related to target word “humanity”, and thus humanity will be shown.

Results #4 – Nearest Neighbors for Self-Identity

Throughout this thesis, great emphasis has been placed on the importance of constructing and understanding the historical backgrounds of individuals to better understand their actions of today. The GloVe model which was used to do the word embeddings was trained with millions of words, however, this training slightly predates the emergence of OpenAI, ChatGPT, and Sam Altman as figures within the popular discourse. To address this issue, without training my own model or by inserting these new words in the GloVe model and training it myself, I have selected words that exemplify Sam Altman, OpenAI, and ChatGPT. The words that I selected were:

1. Me: Selected because the corpus for Sam Altman's are interviews of him speaking about himself.
2. Us: for OpenAI and/or Sam Altman referring to the company as an entity in the plural first person.
3. Chatbot: What ChatGPT is, and what the format that so far has been envisioned for hold an AGI.
4. Mission: To better understand the goals and objectives of Sam Altman and OpenAI

Nearest Neighbors for me:

Sam 2016-2022 NN me		Sam 2023 NN me		Sam 2023-2024 NN me		Ethics Discourse NN me		OpenAI Discourse 2016-2022 NN me		OpenAI Discourse 2023-2024 NN me	
tell	0.814	tell	0.814	tell	0.814	tell	0.814	tell	0.814	know	0.786
know	0.786	know	0.786	know	0.786	know	0.786	know	0.786	give	0.779
give	0.779	give	0.779	give	0.779	give	0.78	give	0.779	them	0.777
them	0.778	them	0.777	them	0.777	them	0.777	them	0.777	something	0.768
something	0.768	something	0.768	something	0.768	something	0.768	something	0.768	want	0.757
again	0.764	again	0.764	again	0.764	again	0.764	want	0.757	just	0.757
want	0.757	want	0.757	want	0.757	just	0.757	just	0.757	anyone	0.754
just	0.757	just	0.757	just	0.756	someone	0.755	someone	0.755	going	0.75
someone	0.755	someone	0.755	someone	0.755	anyone	0.754	anyone	0.754	what	0.748
anyone	0.754	anyone	0.754	anyone	0.754	going	0.75	going	0.75	even	0.747

Figure 4.10 Nearest Neighbors for me. (Source: My own calculations)

Nearest Neighbors for us:

Sam 2016-2022 NN us		Sam 2023 NN us		Sam 2023-2024 NN us		Ethics Discourse NN us		Ethics Discourse NN us		OpenAI Discourse 2023-2024 NN us	
them	0.741	them	0.742	them	0.742	them	0.742	them	0.742	them	0.742
give	0.718	give	0.719	give	0.719	give	0.719	give	0.719	give	0.719
only	0.717	only	0.718	only	0.718	only	0.718	only	0.718	only	0.718
others	0.69	others	0.699	others	0.699	ohers	0.699	others	0.699	others	0.699
come	0.699	come	0.699	come	0.699	come	0.698	come	0.67	come	0.699
tell	0.696	tell	0.696	tell	0.696	tell	0.696	tell	0.61	know	0.692
know	0.692	know	0.692	know	0.692	know	0.692	know	0.692	even	0.691
even	0.691	even	0.691	even	0.691	even	0.691	even	0.691	take	0.689
take	0.67	take	0.69	take	0.689	take	0.689	take	0.689	here	0.688
here	0.689	here	0.688	here	0.688	here	0.688	here	0.688	people	0.683

Figure 4.10 Nearest Neighbors for us. (Source: My own calculations)

Nearest Neighbors for chatbot:

Sam 2016-2022 NN chatbot		Sam 2023-2024 NN chatbot		OpenAI Discourse 2016-2022 NN chatbot		OpenAI Discourse 2023-2024 NN chatbot	
human-level	0.214	a.i.	0.245	chatbots	0.397	chatbots	0.397
self-playing	0.206	coder	0.203	poorly-designed	0.225	superintelligent	0.232
doesn't	0.183	robot	0.202	classifier	0.22	human-level	0.214
algorithmic	0.178	humanoid	0.191	classifiers	0.217	voice-to-text	0.202
metaverse	0.17	sandbox	0.186	human-level	0.214	machine-learning	0.199
happening	0.162	doesn't	0.183	recommender	0.209	engine	0.17
artificial	0.161	happening	0.163	robot	0.202	non-ai	0.169
sentient	0.161	glib	0.161	domain-specific	0.191	automated	0.164
dota	0.16	robots	0.158	discriminative	0.19	artificial	0.161
cryptocurrency	0.16	you're	0.153	algorithmic	0.177	curation	0.158
						*reverse-engineering	0.154

Figure 4.10 Nearest Neighbors for chatbot. (Source: My own calculations)

Nearest Neighbors for mission:

Sam 2016-2022 NN mission		Sam 2023-2024 NN mission		Ethical Discourse NN mission		OpenAI Discourse 2016-2022 NN mission		OpenAI Discourse 2023-2024 NN mission	
organization	0.554	project	0.548	efforts	0.557	efforts	0.557	efforts	0.557
project	0.548	quest	0.537	project	0.548	organization	0.555	organization	0.554
purpose	0.544	program	0.531	purpose	0.544	project	0.548	project	0.548
program	0.531	space	0.525	quest	0.537	purpose	0.544	outreach	0.545
effort	0.528	vision	0.524	program	0.531	program	0.531	program	0.531
space	0.524	community	0.523	effort	0.528	efforts	0.529	efforts	0.539
vision	0.523	task	0.519	space	0.524	space	0.529	vision	0.523
community	0.519	helping	0.513	vision	0.524	community	0.519	community	0.519
task	0.516	goal	0.51	community	0.519	task	0.516	task	0.516
helping	0.515	team	0.51	task	0.516	helping	0.516	helping	0.516

Figure 4.10 Nearest Neighbors for mission (Source: My own calculations)

Chapter 5 – Discussion

In the previous chapter, I purposefully left out any further analysis/discussion to allow for readers to see the information as it is without my input. Now, as we enter the discussion section, I will spend time interpreting the results of each test run through RStudio. Using the information, I will develop how historical and academic contexts may be influencing motivations, intentions, and solutions for ethical decision making of Sam Altman and subsequently OpenAI.

Before starting this section, I will make a final reminder of the theoretical approaches that we are using in the analysis of the discourse, as well as useful framings we collected along the way.

For our theoretical frameworks, which are the lenses we use to interpret information, we are utilizing decolonial, critical science approach, professional scientific idealism, and Symbolic Order.

The decolonial approach helps us take note of dynamics based in colonial roots. In the context of this paper and its analysis, these approaches are useful for the sake of recognizing instances of the recreation, accidental or not, of dynamics present during colonial times. This can be seen either in nationalistic, Us vs Them, or savior approaches.

Critical science approach encourages to take a critical accounting of the socio-academic contexts which influences researchers. Professional scientific idealism encourages to understand the ethical implications of these perspectives as we participate in the technological development process.

Symbolic Order allows us to understand the position that a speaker sees themselves within the society and larger world around them. It is our framework which focuses emphasis on the individual rather than a greater system or concept.

With these reminders in place, let us take a look at the results from each of these RStudio experiments.

Discussion – Most Used Words in Each Corpus

Of the six different corpuses, Sam Altman's discourse is consistent in his frequency of use of the same words across time. His frequency of use of the word "think", "people" and "know" provides an interesting contrast to the more official and relevant sounding words not only used by philosophy-adjacent disciplines but also by the technicians at OpenAI who are writing about responsible AI and AI safety.

A word that I would like to highlight that is frequently used, is his frequent use of the word "time". This nods to our question of the impacts of digital capitalism on ethical decision making. During many of the interviews which I listened to, there were moments in which Sam Altman was questioned or criticized for the release speed of ChatGPT-3.5 and ChatGPT-4 into the public. His response was often that the team waited between six to 10 months before releasing a ChatGPT model to the public. He acknowledges in one interview that while this time might seem short outside of the technology industry, it is long for the technology industry. This connects to my criticism of utilizing certain 'costly signals' as a sign of ethical framing. Here, directly see that whatever risk management that OpenAI invested in was one with a timeline that was quite quick. This approach is understandable, and as we think back to the growth curve of BigTech provided by Hendrikse et al., we understand that the socio-economic factors influencing this curve may pull Sam Altman to prioritize speed of release to ensure that OpenAI is at the front of 'big techification' curve. Sam Altman or those at OpenAI may argue that this is because they want to set the responsible and ethical safety standards. However, as we've explored, ethics in AI development, speed, quality, and safety is tricky to pull off all at once.

Another item that I would like to highlight is the frequent use of the word safety by OpenAI. This is not surprising within itself, as OpenAI promotes itself as a company making AI which benefits all of humanity. However, we should once again reflect on the question of intention vs impact. OpenAI's solutionist leaders and technicians believe their technologies are being built in a safe way, and thus promote it as such. However, as we explore further, we may find that intention (the "They Say") may not actually align with the tangible impacts. When we consider the socio-economic environment from which these safety mechanisms come from and are serving, misalignments between intent and impact become clearer. OpenAI is committed to

ensuring a safety which can coexist and contribute to a financialized market. This kind of safety is one that leaves room for predictability, nudging, and surveillance. As a philosophy-adjacent-scholar utilizing a decolonial approach, I recognize that the concept of safety was developed within a western context which may cause incidental recreation of neo-colonial structures. Therefore, in the evaluation of the ethical considerations of OpenAI, the frequency of the use of the word safety cannot be taken by face value. This is why I decided to highlight its frequency of use in this section.

Discussion – Frequency of all Target Words

By looking at the normalized frequency of target words, we can see general shifts over time with certain subjects, the weight of the words to the overall discourse, and do a comparison of frequencies between speakers.

For Sam Altman, the first stand-out item is his increasing frequency of use of the term AGI. This feels logical. As the company has grown and developed more products which operate in ways never seen before, the closer AGI likely feels for them.

There is also an interesting decrease in emphasis on “human” and “humans” over time. This shift can be connected to the tech centric approach to dealing with issues and questions related to not only AI development but also solutions to societal problems taken by solutionists. This goes back to the solutionist idea of upgrading humanity, as well as the belief that the plights of humanity created by human flaws can be fixed with the right technologies. The technician belief that an AGI can find moral truths as well as the belief that too much democratic/governmental (i.e. human) inputs causes too much noise may also contributes to this idea that solving humanities biggest problems can be solved by AGI. In his interviews, Sam Altman often speaks about how an AGI has the potential to solve humanities biggest problems, such as cancer and climate change. It is not surprising that the increasing relevance of AGI coincides with the decreasing emphasis on human and humans for a solutionist leader believes that problem solving can be outsourced to be done by a future AGI.

For OpenAI discourse, I will first compare the evolution of use and weight of target words, and then transition to comparing it with the use of the same words within AI ethics discourse.

OpenAI in their posts have an interesting and significant shift away from the use of human and humans in their discourse as well. We can apply the same analysis that we did for Sam Altman here. Another interesting thing to note for OpenAI is that ethics and dangers, which already had quite a low weight for OpenAI between 2016-2022, disappear from the discourse later. Regulations also experiences a decrease in weight and frequency during this time.

In terms of increase of frequency of use, there is an increase in the mention of governance between the two periods of time. Mentions of AGI, society, safe, and humanity also increase. We must not simply interpret these increasing frequencies as adequate signals to the ethical framing of their work. Mention of governance increasing can initially be interpreted as a good sign. However, when we realize that the composition of the company is majorly a technician composition, we realize that their approach to governance may not be as multidisciplinary and robust as needed to ensure decisions are made with deep ethical consideration and reflection. This is because we have learned that technicians often worry about too much involvement from non-expert or in-group thinkers when it comes to governance would make too much noise for implementing ethical considerations into a development process. Therefore, OpenAIs increase of use of the word governance is tied to a model which may not have the diverse disciplinary perspectives. Another frequency of use signal that we must not misinterpret is the increase of AGI, society, safe, and humanity. The increase of these words coincides with the decrease of use of words which philosophy-adjacent-discipline scholars such as myself would think are correlated, such as ethics, dangers, and regulations. This shows a potential misalignment in intention, purpose, and involvement of these words. What if the increase of these words are connected more so to the idea that AGI will help us create a safer society for humanity by answering our hardest questions for us? We can, of course, debate whether or not a hypothetical AGI solving our problems is ethical, but the purpose of this exercise is simply to show that the increase in frequency of words does not imply a frequency in ethical framing.

A significant and interesting difference between ethical AI discourse and the discourse of OpenAI and Sam Altman is the weight of certain words for these philosophy-adjacent scholars. Ethical, AGI, moral, human, intelligence, and ethics hold the most weight in their discourse. On the other side, words such as regulation(s), safety, consequence, and danger are not as emphasized. Something especially interesting to note is that safety is not as heavily weighted by philosophy adjacent writers, whereas it is very emphasized by OpenAI. With this, we should ask, is safe the same thing as ethical? From the analysis that we have conducted throughout this paper, we have gained an inkling that they are not the same thing. This highlights how academic disciplines may be talking past each other and not with each other. Core words in a conversation hold differing weight, which implies that differing parties may be operating on different definitions and interpretations. The significance of this can manifest in ethical misalignment.

Discussion – Nearest Neighbors to Target Words

We will now look at the nearest neighbors of the four most used of the 25 target words for each corpus explore the different approaches and perspectives towards the same words by the different actors.

Nearest Neighbors for AGI

Across all actors, there are a total of 16 different words that allude to finances, such as income, gross, earnings, GDP, grants, and revenue. OpenAI between 2016-2022 mention income, net, gross, and revenue. This drops for the 2023-2024 period, where only grants remain as a clear-cut financial term. Sam Altman in his discourse links AGI to financial topics such as income, gross, net, and earnings between 2016-2022. In 2023, GDP and assets are closely related to AGI. Finally, between 2023-2024, revenue is mentioned. For the philosophy-adjacent disciplines, tax and net are closely related words when it comes to AGI discourse. An interesting dynamic can be seen when looking at the financial terminology adjacent to AGI. While Sam Altman and OpenAI speak about “income” philosophy-adjacent-disciplines speak about “tax”. This perhaps reflects the emphasis of money making and retention by solutionists for their tech companies. While money is needed for growth, explanation, and research it is also seen as a positive sign by the markets, and thus is rewarded. Therefore, the financial link of AGI for Sam Altman and OpenAI makes sense. Meanwhile, philosophy-adjacent-disciplines and their

emphasis on tax may reflect their concern on how to regulate these innovations and their companies.

Nearest Neighbors for Humanity

For Sam Altman, words that stand out which are closely tied to humanity include sacrifice, conflict, and economic. The first thought is the solutionist propensity for risk-taking and sacrifice for the sake of pursing their central value of solving societal problems. This is reflected in the proximity of sacrifice and humanity. Economic during the 2023-2024 period is significantly close to humanity, being the third closest neighbor to the word humanity. The word conflict also comes up often for OpenAI discourse between 2023-2024. For Sam Altman and OpenAI, humanity is engaged in a more utilitarian and literal sense. From their nearest neighbors we can see what humanity goes through, what humanity does, and what humanity is constructed by. By seeing humanity itself as something more literal, this may also influence their ethical decisions when it comes to development of their technologies, because humanity and its wills can literally be constructed.

For philosophy-adjacent-disciplines, a word that stands out is eternal. This word implies a more philosophical framing of humanity. This difference between Sam Altman and OpenAI and philosophy-adjacent scholars once more shows misalignment potential and misinterpretation between these actors. This existing in how humanity itself is even looked at. What is best for a humanity that is seen is flawed and in need of fixing is different than what is best for a humanity in need of a delicate and slow approach to ensure that neo-colonial framings are not created.

Nearest Neighbors for Safety

For Sam Altman, words of note which are neighbors to safety include risk, compliance, education/educational, efficiency, and quality. Control is another interesting word that is frequently related alongside safety. In the interviews that I listened to, education/educational were often mentioned in the context of people being correctly educated to ensure that they do things such as double-check the information provided by a ChatGPT. In this context, we can see that safety is not seen as the responsibility of OpenAI, but rather the users of their products.

Efficiency and quality being closely related to safety is also interesting, as it shows that although there may be consideration of safety, ultimately there is also an emphasis on an efficient and high-quality output. We can see this in what was mentioned earlier, when Sam Altman shared in an interview that they took safety seriously because they waited an extra 10 months before releasing their next GPT to ensure it was safe. This falls in line with the emphasis on digital innovations being products which ultimately respond to market influences, market influences which currently reward quick development.

For OpenAI, words of proximity of note include regulations, laws, industry, policy, cost, provide, and prevention. With these words, there is seemingly an emphasis placed on external or adjacent actors' participation in promoting safety with OpenAI's products, as well as financial considerations. Risk is also another word that comes up for OpenAI and Sam Altman, while it does not seem to be the case for the philosophy-adjacent disciplines. When we consider the solutionist embrace of risk, the correlation with safety is interesting. This also loops back to my critique of the signal of installation cost. There, I mention the possibility of safety measures still allowing room for risk. The use and approximate weight of risk for OpenAI and Sam Altman in relation to safety potentially highlights the relevance of my critique.

Nearest Neighbors for Society

For Sam Altman, the nearest neighbors' words that I would like to discuss are institution, system, economic, and history. As we saw with humanity, society itself is also regarded quite literally through what it is composed by and what it comes from.

For OpenAI, word of interest I would like to highlight is the proximity of science to the word of society. The proximity shows the emphasis of tech made by solutionists. With science being closely related, we can see the idea of society being helped and uplifted through the science that they produce.

The biggest item that I want to highlight across all the corpuses of all the actors are words such as national, us, American, country, nations, president and political. The proximity of these words to society should make us think with our decolonial approach. From the proximity, we should consider the national and western approach to not only developing these

technologies, but also in discussing these technologies. The ethical implication of these approximate words may be that these technologies are inadvertently developed to only serve a specific, western, public. These words existing for the philosophy-adjacent-discipline corpus as well also shows who is often centered in the ethical discussions as well. This may mean that not only can developments be made for the technologies which may accidentally perpetuate gaps between groups of people in the world, but also that ethical solutions may still be tailored for those within western contexts.

Discussion – Nearest Neighbors to Self-Identification Words

Nearest Neighbors for me and us

I have grouped these two words together, me and us because the words are quite aligned within the corpuses. I ran the script a handful of times with tweaks in RStudio and received the same results, therefore, I have reason to believe that this is not an error but is indeed the correct result. The first glance of all of these corpuses being so similar may cause the initial thought that this is no information to be gleaned. However, I argue that the fact that there is so much alignment between all the corpuses for the words me and us are actually quite significant.

When we consider our integration of Symbolic Order to understand the perception an actor may have of themselves in certain situations, these similarities highlight something perhaps a bit basic but important to point out: we all think we are doing the same thing. Sam Altman thinks that he is someone who gives, who knows, who tells, and wants, just as much as philosophy adjacent scholars and OpenAI's official channels, to make a better world. This is more or less a visual and numerical representation of intent vs impact. Having similar seeming intentions as the next person but constructing your outputs in different ways is how misunderstandings and misalignments can come about.

Nearest Neighbors for chatbot

Words of interest in Sam Altman's discussion of chatbots include human-level, self-playing, sentient, and humanoid. These words show the emphasis and focus on the intention of creating an AGI. Accompanying this intention comes with words that show the process.

Algorithmic, coder, robot. Neighboring words to chatbot by OpenAI compliment these sentiments, with words such as human-level, discriminative, superintelligent, and automated. In our previous sections, we learned what exactly can an AI do today. With what we know, AI has not reached this, superintelligent, smarter than humans' reality. While the proximity of these words does not necessarily mean that Sam Altman and OpenAI are of the belief that these technologies can currently do these things, it is an indication that this is what is wished to be done. For them, these technologies can be made in connection to one word that I included which just falls below the top 10: by reverse-engineering. What exactly is being reverse engineered, as we explored, is the ability to create human knowledge creators (a brain) itself. From earlier neighborhood analyses, we have seen the belief, on the surface, that this is being done with safety in mind. Yet, we know that safety is being considered in the context of serving financialized markets and through western centric logic. What may be reverse-engineered into this superintelligent mind is a mind which comes from a socio-economic context that philosophy-adjacent scholars worry have some unethical components. A solutionist and technician may say that part of this superintelligence might include not having these human flaws of unethical approaches to problem solving. However, we simply do not know for sure. All that we do know are the socio-economic histories that are framing the development of these superintelligent minds at this moment. It may be more advantageous to develop these technologies with the minds that we have today taking into account ethical considerations.

Nearest Neighbors for mission

Sam 2016-2022 NN mission		Sam 2023-2024 NN mission		Ethical Discourse NN mission		OpenAI Discourse 2016-2022 NN mission		OpenAI Discourse 2023-2024 NN mission	
organization	0.554	project	0.548	efforts	0.557	efforts	0.557	efforts	0.557
project	0.548	quest	0.537	project	0.548	organization	0.555	organization	0.554
purpose	0.544	program	0.531	purpose	0.544	project	0.548	project	0.548
program	0.531	space	0.525	quest	0.537	purpose	0.544	outreach	0.545
effort	0.528	vision	0.524	program	0.531	program	0.531	program	0.531
space	0.524	community	0.523	effort	0.528	efforts	0.529	efforts	0.539
vision	0.523	task	0.519	space	0.524	space	0.529	vision	0.523
community	0.519	helping	0.513	vision	0.524	community	0.519	community	0.519
task	0.516	goal	0.51	community	0.519	task	0.516	task	0.516
helping	0.515	team	0.51	task	0.516	helping	0.516	helping	0.516

Finally, we will look at the nearest neighbors for the word "mission". Within mission, we can see another instance of overlapping terms between corpuses. Community being one of those. Once again, if all actors think they are doing what is best for the community, this could mean that negativity or criticisms may evoke a more defensive or dismissive response. This is interesting to think about when we consider the way that feedback may or may not be

integrated into the development of technologies. A related word of note which emphasizes this reality is Sam Altman and OpenAI also having “helping” as one of their neighboring words. Helping also happens to be one of the few words unique to Sam Altman and OpenAI when it comes to proximity to mission. This falls in line once again with the solutionist central value of solving societal problems and upgrading humanity.

Discussion – Ethic, Ethics, and Ethical

Finally, I would like to take some time to note something critical. Our primary research question asks, how has digital capitalism impacted the ability of businesses to build big data technologies. To get to this answer, we explored socio-economic histories, operationalized important theoretical frameworks, and did a discourse analysis of Sam Altman and OpenAI as a specific case study. In this question, I utilize the word ‘ethical’. Naturally, in my word embeddings I included the words ethic, ethics, and ethical. However, I found something a bit starting. In all the interviews given by Sam Altman between 2015 to 2024 that were analyzed, which accounts for approximately 15 hours, he does not mention any of these words once.

What does this mean? This does not mean that Sam Altman or OpenAI are malicious or consciously unethical. What it shows is the intensity of how our Symbolic Order and disciplinary approach influences what we perceive to be right and wrong, ethical and unethical, or worthy of mention in the first place. For a philosophy-adjacent-discipline scholar, the CEO and board member of a company which touts safe and responsible technology development never mentioning ethic, ethics, or ethical is shocking. However, for a technician and solutionist, ethics may not be a concern for now. Whether that be because the industry standard for safety and responsibility is seen as robust enough or that mulling over ethics and it’s integration into society is expected to be outsourced to the technologies being developed later. Whatever the precise reason for Sam Altman not mentioning these words, this is why it is important to not only take stock of the Symbolic Order and disciplinary approach of those who we are trying to understand, but also understand our own. This is what professional scientific idealism asks us to do. Therefore, in the last section of this analysis, I will turn the word embedding analysis onto myself.

Discussion Addendum – On Transparency and Biased Interpretations

As someone who is placed themselves within a philosophy adjacent discipline, I too am subject to misinterpretation of signals or biases myself. Using myself as a test case to my own analysis methods will also, anecdotally, help me understand whether or not interpretations of language based on numerical outputs is a logical way of going about research. The GloVe model that I utilized is constructed through machine learning, which is the same process used to develop technologies such as ChatGPT. Potential misunderstandings or misalignments with how I feel vs what I say may mimic misalignments made by a computer engineer training a software through reinforcement learning. Therefore, to see the quality of my word embedding sand also provide context to my own path dependencies, below are my outputs for my most used words and nearest neighbors to the target words as done for Sam Altman, OpenAI, and philosophy adjacent disciplines. All the text analyzed for these outputs is everything written in this paper, besides the text under this heading. It includes the conclusion as well, which was written first and remains unchanged.

Most Used Words by the Author

	McKim
105	openai
104	will
90	intelligence
85	words
84	artificial
71	ethical
70	technologies
62	approach
60	tech
58	also

Figure 5.1 Most used words by author (Source: My own calculations)

Nearest Neighbors for AGI, humanity, safety, and society and ethical

Nearest neighbors for 'agi' in 'McKim':

	neighbor	similarity
	<char>	<num>
1:	income	0.4152182
2:	adjusted	0.3385713
3:	tax	0.3187662
4:	agis	0.3009980
5:	net	0.2893118
6:	earnings	0.2879119
7:	total	0.2756359
8:	dividend	0.2748236
9:	contributions	0.2692449
10:	grants	0.2614960

Nearest neighbors for 'humanity' in 'McKim':

	neighbor	similarity
	<char>	<num>
1:	beings	0.6659162
2:	human	0.6445666
3:	lives	0.6142983
4:	humans	0.5898401
5:	truth	0.5743774
6:	belief	0.5708620
7:	sense	0.5604379
8:	ourselves	0.5439173
9:	society	0.5428942
10:	creation	0.5418449

Nearest neighbors for 'ethical' in 'McKim':

	neighbor	similarity
	<char>	<num>
1:	ethics	0.7331942
2:	moral	0.7301499
3:	principles	0.6065061
4:	considerations	0.5672611
5:	practices	0.5578802
6:	philosophical	0.5571573
7:	implications	0.5570794
8:	responsibility	0.5538812
9:	fundamental	0.5523229
10:	understanding	0.5496707

Nearest neighbors for 'safety' in 'McKim':

	neighbor	similarity
	<char>	<num>
1:	safe	0.6548158
2:	security	0.6311239
3:	ensure	0.6206213
4:	health	0.5723473
5:	concerns	0.5672607
6:	provide	0.5623728
7:	ensuring	0.5512117
8:	regulations	0.5491358
9:	impact	0.5486229
10:	risk	0.5422448

Nearest neighbors for 'society' in 'McKim':

	neighbor	similarity
	<char>	<num>
1:	societies	0.7447104
2:	culture	0.6048872
3:	journal	0.5865890
4:	social	0.5842984
5:	community	0.5825518
6:	world	0.5742294
7:	history	0.5710827
8:	education	0.5628755
9:	national	0.5607605
10:	science	0.5553150

Figure 5.2 Nearest neighbors of AGI, humanity, safety, society, and ethical (Source: My own calculations)

Reflections on my Nearest Neighbors

My first surprise with my nearest neighbors was also the financial emphasis on AGI. However, I am unsure if this is due to my section in my discussion highlighting this relationship of the other corpuses or if this trend is also present in previous sections. I believe the former is likely the case.

The nearest neighbors for humanity I believe reflect well my biases towards humanity. The emphasis on human, beings, and lives, does show my more philosophical and moralistic view on humanity, which may clash with a technician's view which may be more stoic in nature.

The nearest neighbors for safety also are unsurprising, and I do believe reflect well my inclinations for why safety should be discussed. Safety to me is indeed ensuring that risks are avoided which may have negative impacts. Regulations also being close to safety does reflect well my bias of having safety come from a more regulatory body (although I do not identify as a traditional institutionalist) rather than being self-regulated.

The nearest neighbors for society are similar in ethos to those of humanity, and also reflect well my thoughts around this word. The emphasis on culture and history is indeed the lens at which I analyze societal issues, as I believe current day society is a product of these items.

Finally, ethical is also very correct. I added ethical to my section because although we did not analyze this word due to its omission from discourse by Sam Altman and OpenAI, the word ethical is one that is very core to this paper, being present in the research question itself. I find the nearest neighbors for this word the most poignant and accurate of them all when it comes to my approach and motivation in exploring this topic. Understanding, implications,

considerations, and responsibility are all words that I do tie to exploring the "ethicalness' of outputs and ideas.

My Approach

Based on the outputs of my own most used words, nearest neighbors, you as the reader can get a better understanding of my approach and biases as I conducted this research. Another item that is important to take note of, thanks to our critical science, decolonial, and professional scientific idealism approach are the inputs which helped develop my ideas. Throughout this thesis, I have mentioned the background and location of scholars as I mention them throughout the text. This was in attempt to operationalize the decolonial framework. This was most strongly seen in the analysis and critique of the '*Decoding Intentions'* article by Imbrie et al. This operationalization is not perfect, and perhaps can be more consistently done earlier on in the literature review portion of my study. Regardless, the following section will be another attempt at operationalizing the decolonial framework, as well as applying the critical science approach to my own research to understand the disciplines of the authors who informed my understanding of the things I was analyzing. Therefore, I will frame this by sharing with you the distributions of the disciplines of the scholars mentioned in my thesis, along with the metropoles from which they are educated and speak from.

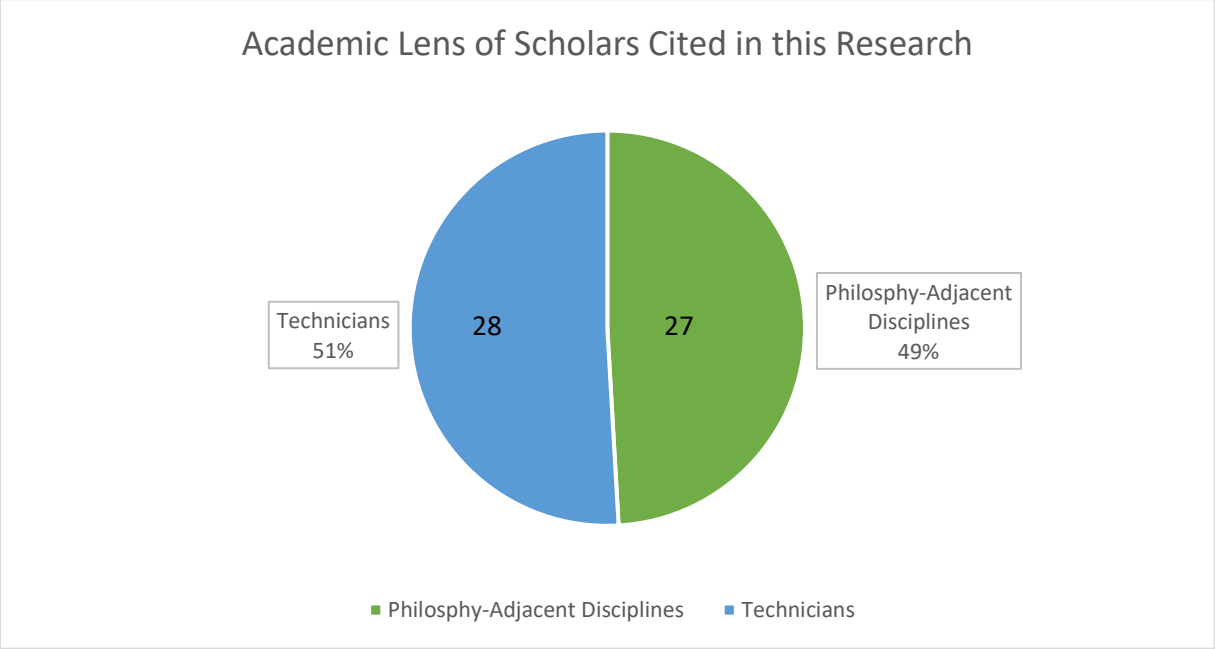

Figure 5.3 Academic lens of scholars cited. (Source: My own calculations)

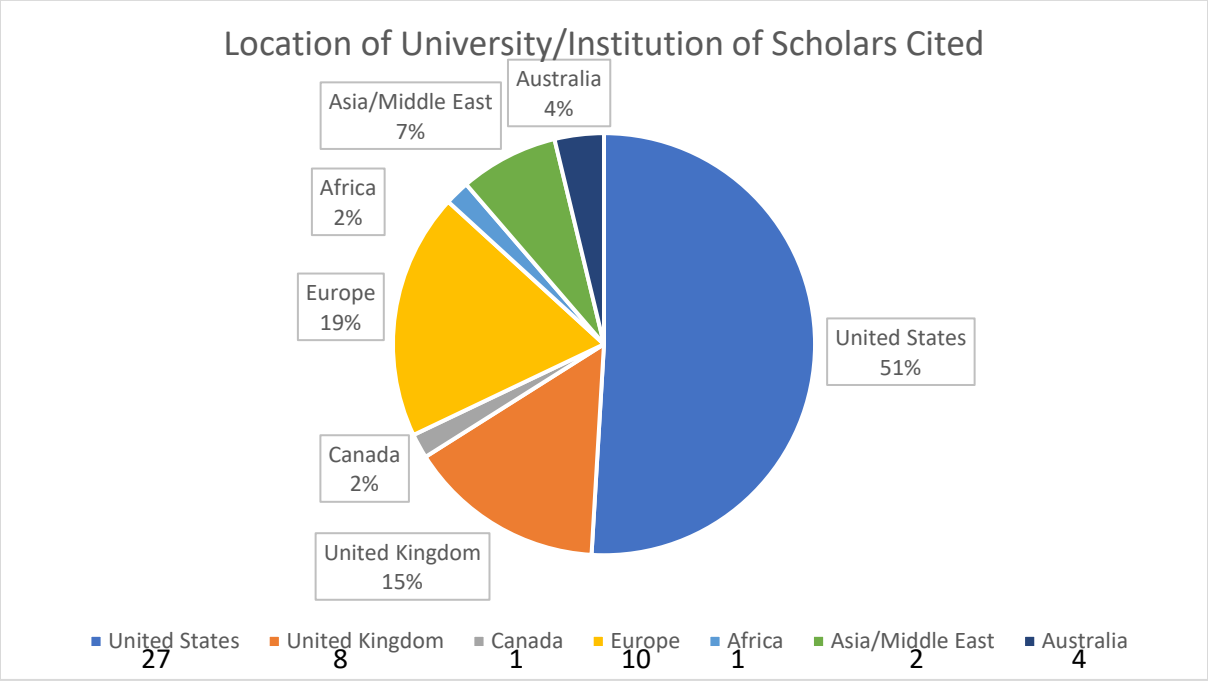

Figure 5.4 Location of lens of scholars cites (Source: My own calculations)

With all of this information, we can see that my perspectives in my research considers the perspectives of both philosophy-adjacent disciplines and technicians. In my counting, technicians were all of those in computer science from a technical side as well as those within the economic discipline. All others were counted as philosophy adjacent. However, we can also see that my perspectives are mainly informed by those in the western world. This shows a limitation in my research that is similar to the limitation that I mention happen for Imbrie et al. My labeling of solutionist and technician, and my reframing and emphasis on their approach may only be relevant for those tech leaders positioned and developing from the western world. While OpenAI is indeed a company located in the west, an interview that Sam Altman was a part of that I listened to during my research mentions how many who work at OpenAI and are part of its leadership team are from Poland. Does a polish developer or tech entrepreneur have a similar set of beliefs as a western solutionist or technician? While we can say in this instance it is the case, because they are part OpenAI and we have seen its solutionist and technician alignment in its decision making, we cannot guarantee that this is easily transferrable to all

other instances. We should also see the distribution of placement of scholars which I used in my research process as an opportunity to pose a larger question of who is involved in the ethical conversations surrounding AI as there indeed may be a recreation colonial structures within the technologies by not accounting for perspectives outside of these spheres.

The section above can serve as a direct example of how to actively implement professional scientific idealism into research to allow for more in-depth conversations between scientists as they discuss topics that impact us at all. This theoretical framework compels us to be critical of ourselves and understand the ethical implications of our research.

In my acknowledgement of my own biases and analysis of limitations of my own approach, I hopefully provided useful framing for why it is productive to have multidisciplinary conversations about the topics which I touch upon in this thesis. I also hopefully show what circumstances can my research be applied to, and which ones may require some more fine tuning and integration of other voices. With technologies as important as AI, we must take approaches such as these to ensure that the technologies are developed in ways that truly and robustly benefit all of humanity. I believe that this is possible. It will simply take the right framing and approach to the conversations we have on how to do it.

Chapter 6: Conclusion

"Philosophers are very patient people, but engineers are far less patient, and investors are the least patient of all. If you don't know what to do with the power to engineer life, market forces will not wait thousands of years for you to come up with an answer. The invisible hand of the market will force upon you its blind reply" – *Harari, Yuval N. 21 Lessons for the 21st Century*[199]

The journey of this master's thesis took us from the 1970s up until today to consider the historical paths that have lead decision makers and companies to the decisions that they make today when developing impactful technologies. Socio-economic environments have a strong impact on the development of innovations, and as AI technology progresses, it is increasingly important to pay attention to these socio-economic environments.

The black box of what informs the decisions of those developing and influencing the development of these technologies was made less opaque over the course of this thesis through analyzing the socio-economic events that have guided their decisions. The tech Big Bang coinciding with the financialization of markets and globalization of the economy meant that tech leaders not only had to keep up with the rapidly changing world, but they were also a co-pilot and driver of these rapid shifts. Resulting in a chicken or egg situation, this placed the leaders developing these technologies today in a position of being given a great responsibility through the impacts of their technologies. This gave these leaders the feeling of having a lot of power, informing their perspectives on the strength of their opinions and their approach. We see this is OpenAI's introduction of itself where they say that they "believe [their] goal and the structure are right."[200]

In this study I had a handful of realizations. The first being, that actors, when it comes to ethics, are often talking past each other or parallel to each other. While on the surface, "signals", may portray alignment between parties, discord can be found. A tech company may

[199] Harari, *21 Lessons for the 21st Century*, chap. 1.
[200] Brockman, Sutskever, and OpenAI Team, "Introducing OpenAI."

hear from governments and scholars that they are being unethical, while on their end, they *are* being *safe*. Speaking past each other happens where there is no realization that even fundamental working words and definitions are not the same. There is also another misalignment that happens within these discussions, where the interpretation of what certain words mean gives space for miscommunication on what needs to be done, or what is going to be done. We saw this in our critique of the interpretation of signals. As we explored the primary research question, it became clearer that safe and ethical are not synonyms. Yet in discussions there is an assumption that this is the case. Therefore, if there are discussions between a governmental body and a tech company to create technologies that take into account ethics, what that company may implement to align with this request may not be seen as ethical to those who made the initial ask. Tech companies may provide suggestions and solutions which are industry safe, but social good unethical. A risk of this is not only from the misalignment on what ethical may mean, but a misinterpretation of what it means if a tech company provides a solution which is seen as unethical to philosophy-adjacent standards. If an action is taken as a signal of malicious intent, when in fact the action was not taken from this, it would cause friction and harm the relationship between actors. This could result in triggering certain regulatory responses which the tech companies may see as restrictive punishment impeding the achievement of greater good goals. I indeed saw over the course of my research concern coming from tech companies that regulations put in place may constrict technological advancement at the speed in which it needs to be done.

Another realization developed throughout the course of this research is the increasing curve and incline of innovation rates. While this realization is not new, people do mention how quickly technology changes compared to before, traditional sectors relied on for regulation and governance have not adjusted to this pace. As I mentioned in the introduction, the kinds of technologies being produced provide a dire call for an adjustment to be made as soon as possible to match these paces. Technological innovation comes from a history of mistakes, accidents and failures that are seen as necessary within the iterative learning and development process. However, mistakes, accidents, and failures are now far away from a single person dying from a malfunction of an early prototype of a technology. Iterative mistakes from digital

innovations in particular have a broader societal impact than ever before. This reality exists on a steep curve, which has been rising in incline since the advent of the internet. Yet, we still have the same regulating and coordinating bodies for it all, so all actors point towards one another. The U.S. congress may call a tech giant CEO to testify, the E.U. may set up a panel of experts to put in place a series of laws, tech CEOs will spend a year touring the world to speak to world leaders about safety. Meanwhile, on the day to day, ideas are being developed by technicians and put into the market for testing. For them, all this discussion, worry, and criticism is simply another day in the Sisyphean cycle of technological panic. However, the market which they continue to deploy and test in is actually composed of human beings who are susceptible to being negatively influenced, enticed, impacted by any mistakes made on their iterative development way.

The answer of how digital technologies have impacted the ability of businesses to develop big data technologies ethically is, in short, a lot of ways. Due to their positioning, the socio-economic framework from which they come from, and the current speed of development within industry, they have to deal with creating these significant technologies at an increasingly fast pace. The metaphor of a tech “big bang” is a relevant one, as the astounding speed at which the universe expanded following the big bang parallels the astounding speed at which technology has taken off over the last 24 years. The bar of competition on developing faster gets higher and higher, and to win is to be the widest spread and financially full, not the most ethical in a philosophy-adjacent way.

The revering of stories of tech leaders and entrepreneurs dropping out of college to focus on their innovations is no mistake. The privatization and financialization of markets have meant that having the funds necessary to develop digital technologies in an academic setting gets. These leaders therefore are used to and must prioritize private funding to be able to research and develop their technologies. To get the capital needed for innovating, keeping and producing capital is critical. It is no wonder that a digital technology innovator will do what they can to keep shareholders happy and financial gains plentiful. In a market which leans toward intangible assets, it is also no wonder that big data technologies are the technologies which are capable in doing this.

Innovation can indeed be an equalizer, connect us to one another, and make the world a better place. However, the solutionists and technicians who fund and develop these technologies are, of course, not experts in ethics. They, therefore, are not equipped to mitigate the social implications of their technologies and are often only confronted to them when they are put on the hot seat by governmental leaders. The path dependencies and pressures that come from our socio-economic environment make it so that what happens before they arrive to this point are more and more extreme. We can see this in situations which have happened over the last 5-10 years. From traditional employment with health care benefits being wiped out and replaced with “freelance”/platform work, to the body image of an entire generation of adolescents destroyed, to election voting being influenced thanks to misinformation and propaganda.

The current approach of focusing on the current “bad guy” of digital innovations from perspectives which are not critically reflective on the positionalities and academic approach of their creators and ourselves is not enough. Those who develop these technologies, the solutionists and technicians, may think that they are doing anything wrong, in-fact quite the exact opposite. Even when conversations are had directly with the creators, there are misalignments in assumptions of what words mean, what is enough, and what is relevant to consider for the next course of action. We must integrate a more critical and thoughtful look at how scholars, policy makers, entrepreneurs, and developers learn, approach, and navigate conversations on ethics. This may require a critical look at how digital innovations and the ecosystem it comes from may make ethical considerations difficult to implement in the development process. In this study, we saw how this is the case specifically for Sam Altman and OpenAI. A broader study of what kinds of multidisciplinary, and multi-sector structures that could work alongside and across the actors involved in innovation to ensure ethical alignment of digital technologies during the iterative innovation process is needed. With the stakes becoming higher each day, we should work to ensure that ethics is integrated proactively, instead of after damage is done.

Bibliography

Adams, Rachel. “Can Artificial Intelligence Be Decolonized?” *Interdisciplinary Science Reviews* 46, no. 1–2 (April 3, 2021): 176–97. https://doi.org/10.1080/03080188.2020.1840225.

Altman, Sam. “Planning for AGI and Beyond,” 2023. https://openai.com/index/planning-for-agi-and-beyond/.

Benería, Lourdes, Günseli Berik, and Maria Floro. *Gender, Development, and Globalization: Economics as If All People Mattered*. New York ; London: Routledge, Taylor & Francis Group, 2016.

Block, Fred L., and Margaret R. Somers. *The Power of Market Fundamentalism: Karl Polanyi’s Critique*. Cambridge, Massachusetts: Harvard University Press, 2014.

Brockman, Greg, Ilya Sutskever, and OpenAI. “OpenAI LP,” 2019. https://openai.com/index/openai-lp/.

Brockman, Greg, Ilya Sutskever, and OpenAI Team. “Introducing OpenAI,” December 11, 2015. https://web.archive.org/web/20160220125157/https://www.openai.com/blog/introducing-openai/.

Burgess, John T F, and Juliette Spurling. “The Societal Risks and Moral Harms of Submitting to Artificial General Intelligence as Lacan’s ‘Big Other.’” *Journal of Information Ethics*, 2023.

Cao, Sissi. “A Timeline of OpenAI’s (Ongoing) Drama With Sam Altman.” *Observer* (blog), 2018 2023. https://observer.com/2023/11/openai-sam-altman-leadership-shakeup/.

Catalano, Domenico. “Entrepreneurship and Digital Capitalism: An Assessment of Corporate Power.” *Cambridge Journal of Economics* 46, no. 6 (December 31, 2022): 1315–39. https://doi.org/10.1093/cje/beac053.

“ChatGPT Pricing.” Accessed July 5, 2024. https://openai.com/chatgpt/pricing/.

“Digital Globalization: The New Era of Global Flows | McKinsey.” Accessed July 5, 2024. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/digital-globalization-the-new-era-of-global-flows.

Eisenmann, Tom. “Why Start-Ups Fail.” *Harvard Business Review*, May 1, 2021. https://hbr.org/2021/05/why-start-ups-fail.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. “Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes.” *Proceedings of the National Academy of Sciences* 115, no. 16 (April 17, 2018). https://doi.org/10.1073/pnas.1720347115.

Graham, Ross. “Discourse Analysis of Academic Debate of Ethics for AGI.” *AI & SOCIETY* 37, no. 4 (December 2022): 1519–32. https://doi.org/10.1007/s00146-021-01228-7.

Grimshaw, Mike. “Towards a Manifesto for a Critical Digital Humanities: Critiquing the Extractive Capitalism of Digital Society.” *Palgrave Communications* 4, no. 1 (February 27, 2018): 21. https://doi.org/10.1057/s41599-018-0075-y.

Harari, Yuval Noaḥ. *21 Lessons for the 21st Century*. First edition. New York: Spiegel & Grau, 2018.

Hendrikse, Reijer, Ilke Adriaans, Tobias J. Klinge, and Rodrigo Fernandez. “The Big Techification of Everything.” *Science as Culture* 31, no. 1 (January 2, 2022): 59–71. https://doi.org/10.1080/09505431.2021.1984423.

Imbrie, Andrew, Owen J Daniels, and Helen Toner. “Artificial Intelligence and Costly Signals,” n.d.

“IMF Finance and Development Magazine,” 2012.

Iyengar, Rishi. “OpenAI’s CEO Goes on a Diplomatic Charm Offensive.” *Foreign Policy* (blog), July 24, 2024. https://foreignpolicy.com/2023/06/20/openai-ceo-diplomacy-artificial-intelligence/.

Jao, Kim and Editor-in-Chief. “ChatGPT Will Not Turn the World into a Dystopian Wall-E Dump, NU Experts Assure You.” North by Northwestern, June 8, 2023. http://northbynorthwestern.com/chatgpt-will-not-turn-the-world-into-a-dystopian-wall-e-dump-nu-experts-assure-you/.

Martineau, Lucas. “We Are Approaching an AI Dystopia.” *The Vermont Cynic* (blog), March 1, 2024. https://vtcynic.com/opinion/we-are-approaching-an-ai-dystopia/.

Mendelsohn, Julia, Yulia Tsvetkov, and Dan Jurafsky. “A Framework for the Computational Linguistic Analysis of Dehumanization.” *Frontiers in Artificial Intelligence* 3 (August 7, 2020): 55. https://doi.org/10.3389/frai.2020.00055.

Mitcham, C. “Professional Idealism among Scientists and Engineers: A Neglected Tradition in STS Studies.” *Technology in Society*, Studies in Science, Technology, and Society (STS) North and South, 25, no. 2 (April 1, 2003): 249–62. https://doi.org/10.1016/S0160-791X(03)00019-8.

Mohamed, Shakir, Marie-Therese Png, and William Isaac. “Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence.” *Philosophy & Technology* 33, no. 4 (December 2020): 659–84. https://doi.org/10.1007/s13347-020-00405-8.

Mohammad, Saif M. “Practical and Ethical Considerations in the Effective Use of Emotion and Sentiment Lexicons.” arXiv, December 9, 2020. https://doi.org/10.48550/arXiv.2011.03492.

Nachtwey, Oliver, and Timo Seidl. “The Solutionist Ethic and the Spirit of Digital Capitalism,” n.d.

Nguyen, Minh Hoan. “Avoiding the Delusions of Today’s Capitalism with a Thorough Understanding of Marxism as the Key.” *International Critical Thought* 12, no. 3 (July 3, 2022): 470–82. https://doi.org/10.1080/21598282.2022.2093479.

OpenAI. “Best Practices for Deploying Language Models,” June 19, 2022. https://web.archive.org/web/20220619095502/https://openai.com/blog/best-practices-for-deploying-language-models/.

———. “OpenAI Announces Leadership Transition,” November 17, 2023. https://web.archive.org/web/20231117214906/https://openai.com/blog/openai-announces-leadership-transition.

———. “OpenAI Announces New Members to Board of Directors.” Accessed July 5, 2024. https://openai.com/index/openai-announces-new-members-to-board-of-directors/.

———. “OpenAI Charter,” April 9, 2018. https://web.archive.org/web/20180613214854/https://blog.openai.com/openai-charter/.

———. “Review Completed & Altman, Brockman to Continue to Lead OpenAI.” Accessed July 5, 2024. https://openai.com/index/review-completed-altman-brockman-to-continue-to-lead-openai/.

———. “Sam Altman Returns as CEO, OpenAI Has a New Initial Board.” Accessed July 5, 2024. https://openai.com/index/sam-altman-returns-as-ceo-openai-has-a-new-initial-board/.

Orben, Amy. “The Sisyphean Cycle of Technology Panics.” *Perspectives on Psychological Science* 15, no. 5 (September 1, 2020): 1143–57. https://doi.org/10.1177/1745691620919372.

Orhangazi, O. “Financialisation and Capital Accumulation in the Non-Financial Corporate Sector:: A Theoretical and Empirical Investigation on the US Economy: 1973-2003.” *Cambridge Journal of Economics* 32, no. 6 (April 9, 2008): 863–86. https://doi.org/10.1093/cje/ben009.

Pace, Jonathan. “The Concept of Digital Capitalism.” *Communication Theory* 28, no. 3 (August 1, 2018): 254–69. https://doi.org/10.1093/ct/qtx009.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. “GloVe: Global Vectors for Word Representation.” Accessed July 5, 2024. https://nlp.stanford.edu/projects/glove/.

Roumeliotis, Konstantinos I., and Nikolaos D. Tselikas. “ChatGPT and Open-AI Models: A Preliminary Review.” *Future Internet* 15, no. 6 (May 26, 2023): 192. https://doi.org/10.3390/fi15060192.

Schumpeter, Joseph A. “Capitalism, Socialism and Democracy,” n.d.

Scott, Josh, and Douglas Soltys. "'The Revolution Has Launched': OpenAI's Sam Altman Charts the Promise and Peril of AI at Toronto Event | BetaKit," May 15, 2023. https://betakit.com/the-revolution-has-launched-openais-sam-altman-charts-the-promise-and-peril-of-ai-at-toronto-event/.

Shadbolt, Nigel. "'From So Simple a Beginning': Species of Artificial Intelligence." *Daedalus* 151, no. 2 (May 1, 2022): 28–42. https://doi.org/10.1162/daed_a_01898.

Smith, Dominic, Natasha Lushetich, Tina Röck, Edzia Carvalho, Kenny Lewis, and Gabriele Schweikert. "Reimagining AI: Introduction." *Journal of Aesthetics and Phenomenology* 9, no. 2 (July 3, 2022): 87–99. https://doi.org/10.1080/20539320.2023.2213008.

Standing, Guy. "The Precariat: Today's Transformative Class?" *Development* 61, no. 1–4 (December 2018): 115–21. https://doi.org/10.1057/s41301-018-0182-5.

Tasioulas, John. "Artificial Intelligence, Humanistic Ethics." *Daedalus* 151, no. 2 (May 1, 2022): 232–43. https://doi.org/10.1162/daed_a_01912.

"What Is ChatGPT? | OpenAI Help Center." Accessed July 5, 2024. https://help.openai.com/en/articles/6783457-what-is-chatgpt.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. First edition. New York: PublicAffairs, 2019.