LSE Media and Communications

Media@LSE MSc Dissertation Series

Editors: Simi Dosekun and Hao Wang

SCI-FI-PILLED MALE MYTH-MAKING

A Critical Discourse Analysis of AI as an Existential Risk

ANDREA HORVÁTHOVÁ

Published by Media@LSE, London School of Economics and Political Science ("LSE"), Houghton Street, London WC2A 2AE. The LSE is a School of the University of London. It is a Charity and is incorporated in England as a company limited by guarantee under the Companies Act (Reg number 70527).

ABSTRACT

This thesis conducts a critical discourse analysis of the portrayal of Artificial Intelligence (AI) as an existential risk in UK news media. It explores the intersection between science fiction narratives and real-world discourses that frame AI as a potential threat to humanity. The study utilises Fairclough's three-dimensional model to analyse a selection of news articles, aiming to uncover the underlying power dynamics, key proponents, and marginalised perspectives within this discourse. By examining how AI is framed as an existential risk, the research reveals the role of media in constructing technological myths. Additionally, it highlights the influence of economic and social factors on the narrative, contributing to a deeper understanding of the media's agenda-setting and framing functions. The findings of this study offer insights into the broader implications of AI discourse and its impact on societal attitudes towards technology.

INTRODUCTION

> *Perhaps the greatest mistake people make about technology is to assume that knowledge of its inner workings can be extrapolated over years to tell us not only where the machine is heading but also where it is taking us.* (Mosco, 2004: 14)

Throughout most of my life, AI has been a subject relegated to the realms of science fiction books and movies. These stories, teeming with imaginative possibilities and dystopian futures, have long captured the public's imagination. However, in the past decade, there has been a notable shift. AI has increasingly moved from the pages of fiction and into the headlines of news media, signalling its growing relevance and impact on our contemporary world. Interestingly, these news stories often reference the same fictional narratives that once seemed purely speculative.

One recurring theme in both fictional and real-world discussions about AI is its potential to pose significant risks. Among these, the concept of a 'robot takeover'—a scenario where AI surpasses human control and threatens our very existence—stands out prominently. This convergence of fiction and perceived reality raises a compelling question: To what extent are these fears mere echoes of science fiction, and to what extent do they represent genuine existential risks?

The launch of ChatGPT, a sophisticated AI language model, has recently intensified this debate. It has sparked widespread discourse about the capabilities of AI and its potential to reshape various aspects of our lives. The research so far has addressed the portrayal of AI technologies in the media in general, but has lacked an interrogation of the existential risks associated with the technologies specifically. This has provided a timely impetus to delve deeper into how these threats are portrayed.

This thesis aims to critically examine the discourse surrounding the narrative that frames AI as a potential existential threat capable of taking over the world, destroying humanity, and causing extinction. By scrutinising this discourse, the study seeks to understand the construction of these narratives and identify the key proponents behind them.

The specific objectives of this research are to analyse the portrayal of AI as an existential threat in UK news media coverage, to identify the key voices and proponents of this narrative, and to examine which perspectives and voices are marginalised or absent in this discourse.

The methodological framework for this thesis is grounded in Critical Discourse Analysis (CDA), utilising Fairclough's three-dimensional model. This approach will facilitate a comprehensive examination of the texts, focusing on the interplay between discourse, social practice, and power. The analysis will be conducted on a selection of articles from UK news media that capture significant moments in the discourse on AI and existential risk.

The conceptual framework will introduce several key theories and phenomena, including the social construction of technology, the agenda-setting function of the media, and concepts related to the technological sublime and myths in technology. These frameworks will provide a lens through which to interpret the findings and understand the broader implications of the discourse.

This research will contribute to the existing body of knowledge by providing a nuanced understanding of how AI is framed as an existential threat in the media. It will shed light on the dynamics of this discourse, revealing the power structures and interests that shape public perception. Additionally, it will offer insights into the role of the media in constructing and perpetuating technological myths and fears.

This dissertation is structured into seven sections. The literature review, divided into two parts, outlines the theoretical concepts and background information on the discourse. The third section clearly states the research question guiding the study. The fourth section describes the research methods, design, limitations of the study, and the researcher's positionality. In the fifth section, the selection of articles and the analysis using CDA are detailed. The sixth section interprets the findings in light of the theoretical framework and research objectives. Finally, the conclusion summarises the key findings, discusses their implications, and suggests areas for future research.

THEORETICAL CHAPTER

The literature review is divided into two sections: the first outlines the guiding concepts and theories for the analysis, while the second traces the background of the AI discourse, including fears of AI-facilitated human annihilation and related topics.

Conceptual Framework

The Media as Agenda Setters & Framers

The role of mass media in directing attention to particular issues has been the subject of extensive study. By selecting what constitutes news, editors and media personnel shape political reality. Audiences are not only informed about issues but are also prompted to assess their importance based on the issues' positioning, the volume of coverage, and the resources dedicated to them. In short, the media possess an agenda-setting function, informing the public not only about what matters but also about how to think about it (McCombs & Shaw, 1972: 177). Similarly, when journalists present preselected issues as news, the framing of these issues significantly impacts citizens' understanding. The foregrounding of certain aspects of reality while overshadowing others possesses a selective character that suggests specific attitudes and judgments (Lecheler & Vreese, 2019: 1, 3). In their paper on how journalists metaphorically frame emerging technologies, Droog et al. (2020: 820) argue that emerging technologies present a particularly relevant domain for the study of frame-building.

Furthermore, certain elite groups, represented by public or private institutions, have disproportionate access to public discourse and communicative events, thereby acting as key actors in the news landscape. These powerful entities can control the conversation by selecting the time and place for meetings with journalists, determining agendas, filtering the language they employ, and much more (Van Dijk, 1995: 12).

Importantly, journalists today face great pressure from economic factors, which are often quantified through ratings and readership metrics, when selecting frames (Nielsen, 2019). This pressure can lead to an emphasis on sensational headlines and an overreliance on appeals to emotion, fear, or novelty.

Technological Sublime & SCOT

In *American Technological Sublime*, David Nye (1994) explores how myths about the progressive potential of technology come into being in American society. Nye describes the *technological sublime* as something which can be understood as the quasi-spiritual haze emanating from particularly visible and impressive technological advances. This phenomenon is exemplified by the initial dual perception of railroads as both instruments of world peace and fearsome evils (Nye, 1994: 54). As to the definition of a myth, Roland Barthes (1957) characterises myths as depoliticised speech, stripped

of tension and conflict, which acquire new ideological significance through critique and reassessment. Despite some differences, there exists an overlap between the concepts of myth, Gramscian 'common sense', and Foucauldian 'discourse' (Mosco, 2004: 29, 30, 31).

In recent years, a growing body of scholars has investigated the role of metaphors, myths, and rhetoric in relation to computational technologies. It is the comparison of new technologies with already established phenomena that makes the new easier to comprehend (Droog *et al.*, 2020). Natale and Ballatore (2020) shed light on the conjectures and fantasies surrounding Artificial Intelligence. Firstly, they discuss how analogies and shifts in discourse borrow concepts from other fields to describe AI technologies. Secondly, they highlight a rhetorical focus on the future, suggesting that current limitations will soon be overcome. Thirdly, they stress the importance of controversies surrounding AI, arguing that these debates are central to the construction of the AI myth. Thus, the AI myth is shaped by both its proponents and its critics (Natale & Ballatore, 2020: 4).

An important dimension of myths is their capacity to function as a marketing device, a phenomenon as old as the market itself (Mosco, 2004: 32). The myth surrounding technology often benefits the valuation of a company's stock, regardless of the actual technical capabilities of that company. This was observed at the turn of the century when researchers found that investors were willing to fund virtually anything that was declaratively linked to the Internet (Cooper *et al.*, 2001). Beyond private companies, the government can serve as a legitimising factor in the substantiation of myths by aligning itself with the future that these myths propose (Mosco, 2004: 45). These findings inform us of the impact that a myth surrounding technology and AI could have on the quarterly revenue reports of AI companies.

Another work that explores the mythical and magical discourses surrounding technology, specifically deep learning, is Campolo and Crawford's (2020) paper on what they term *enchanted determinism*. This work builds on Max Weber's theory of disenchantment, which diagnoses Western modernity as having replaced magical or religious forces with a process of 'rationalisation and intellectualisation' (Weber, 1946). According to Campolo and Crawford, enchanted determinism is a discourse that presents deep learning technologies as both magical and deterministic. These technologies produce concrete consequences when deployed in social contexts, yet the decisions behind these consequences are not fully understood or controlled by their designers (Campolo & Crawford, 2020: 3). Paradoxically, when the disenchanted predictions and classifications of deep learning technology

work as hoped, a profusion of optimistic discourse characterises these systems as magical, invoking mysterious forces and superhuman power. Thus, these systems seem to violate the epistemology of disenchantment—the idea that no 'mysterious' forces are acting in the world (Campolo & Crawford, 2020: 5). The paper also emphasises the need to question whose interests are served and who is responsible for the impacts of these systems when witnessing instances of enchanted determinism at work (Campolo & Crawford, 2020: 15).

This interplay between technological myths and market dynamics is intricately linked to broader theoretical perspectives on technology's role in driving social change. MacKenzie and Wajcman (1999: 23) describe *technological determinism* as the perspective that technological change progresses due to scientific advances or technology's own internal logic, with subsequent impacts on society. This view fosters a passive attitude towards technology's development, leading societies to focus on adapting to technology rather than actively shaping its course. Within this framework, MacKenzie and Wajcman distinguish between *hard* and *soft* forms of technological determinism. Hard technological determinism asserts that technology is the primary force driving social change, implying an inevitable and unidirectional influence of technology on society. This perspective is often criticised for being overly simplistic, as it ignores the complex interplay of other social factors. In contrast, soft technological determinism, with proponents such as Smith and Marx (1994), acknowledges that while technology influences society, this influence is not absolute and is mediated by social, cultural, and economic contexts. Complementing this view, theories of *the social shaping and construction of technology (SCOT)* further emphasise that technology is not just a passive force but is actively moulded by social, economic, and cultural factors (MacKenzie & Wajcman, 1999: 5, 18). In the construction of technology paradigm, the imagined socio-technical futures play an important role, having the power to shape the path innovation takes, for example, by influencing how funds are allocated (Bechtold *et al.*, 2017: 86). For the rest of this paper, the hard kind of technological determinism based on its definition will be referenced.

Equally contestable is the belief that technology is neutral. Building upon Marxist theory, which asserts that 'our' technology is capitalist technology, marked by the relations of class domination, this perspective highlights the inherent biases within technological development. A prominent proponent of the non-neutrality of technology is Langdon Winner (1980), arguing that technologies are inherently political and designed to serve certain groups over others. To expand upon this, the

historical and material links between contemporary technologies and masculinity are of note. Industrial, commercial, military technologies and the like are not neutral but are deeply rooted in male-dominated social structures and therefore cannot easily be adapted to non-male modes of operation (Cockburn, 1999: 127).

Background

Artificial Intelligence (AI) is a term used to describe a wide range of technologies and systems, and its meaning has evolved over the years. According to the *Oxford Dictionary of Computer Science*, AI is defined as 'a discipline concerned with the building of computer programs that perform tasks requiring intelligence when done by humans' (A. B. Butterfield et al., 2016). There is an important distinction between two types of AI systems: narrow AI and general AI (AGI). Pennachin and Goertzel (2007) define narrow AI as programmes that exhibit intelligence in specific domains, such as playing chess, diagnosing medical conditions, or performing algebraic calculations. AGI, on the other hand, refers to software capable of solving a multitude of complex problems across various domains, operating autonomously with human-like thoughts, concerns, emotions, strengths, weaknesses, and predispositions. According to Broussard (2018: 11), general AI is the stuff of dreams, whereas narrow AI is the reality.

In 1955, most research projects focused on developing AI aimed at creating 'machines performing the most advanced human thought activities,' or AGI (McCarthy *et al.,* 2006). However, before the turn of the century, most AGI and AI efforts were scrapped due to a lack of progress in the field and researchers' reluctance to be associated with such ambitious research claims (Markoff, 2006; Russell & Norvig, 2016: 16–28).

The cyclical rise and fall in popularity of AI research is known as AI *summers* and *winters*. In other words, the development of AI has experienced several cycles of hype followed by a decline in favour (Gonsalves, 2019). The present wave of AI enthusiasm can be traced both to the advent of deep learning, a technical breakthrough which has significantly advanced the field, and the launch of ChatGPT in November 2022 (OpenAI, 2022). This launch was followed by a race for the release of competing consumer products (Hern, 2024). Today, OpenAI (n.d.) defines AGI as 'highly autonomous systems that outperform humans at most economically valuable work.'

AI and Ideology

Linking back to the social construction of technology, it is argued that not only economic but also political and cultural factors influence the trajectory which a technology takes. More precisely, the cultures and practices associated with a given technology are its defining features (Wajcman, 2010). Because of this, it is crucial to investigate the ideologies and ideological beliefs of the people who design, build, deploy and finance the technologies which fall under the Artificial Intelligence umbrella.

Ideologies, according to Fairclough, contribute to the establishing, upholding, or changing of power relations, while having the power to dominate and exploit (Fairclough, 2003: 10). Going back to Silicon Valley in the 1990s, in *The Californian Ideology*, Barbrook and Cameron (1996) describe the mixture of technological determinism and libertarian individualism which, they argue, was prevalent during the rise of networking technologies. The Californian Ideology was marked by a belief in neoclassical economics, individualism and techno-utopianism, which was believed to lead to the creation of a post-industrial, post-capitalist, knowledge-based economy (Barbrook & Cameron, 1996).

According to Broussard (2018: 10), the dreams and visions of certain powerful individuals have a sizable impact on the direction of scientific knowledge, culture, and even the legal and regulatory frameworks governing technology. The people in positions of power are often inspired by myths and strive for their realisation no matter the cost (Buck-Morss, 2000). Exploring the relationships and opinions of these individuals is therefore crucial for understanding the direction of technology and AI. *Technochauvinism*, coined by Broussard (2018), is another of the various concepts describing the ideological backdrop of technology development. Technochauvinism is marked by the blind optimism and belief that technology is the solution to all. It is often accompanied by ideologies such as Ayn Rand's objectivism and libertarianism. On top of that, technochauvinists value efficient code and scientific knowledge higher than human decency or social behaviour, akin to the worldview that emphasises the exclusive authority of science and empirical methods in understanding and solving all aspects of life, known as *scientism*.

In a recent paper, Gebru and Torres (2024) trace the ideological motivations of those spearheading the development of AGI to what they call *TESCREAL*, an acronym for transhumanism, Extropianism,

singularitarianism, (modern) cosmism, Rationalism, Effective Altruism, and longtermism. In their analysis, they note that even though many of the organisations that are working towards AGI are vocal about the need for 'AI safety', their understanding of safety has roots in utopian-apocalyptic visions stemming from first-wave eugenicists. Importantly, the researchers note that much of the funding for AGI research and development today comes from billionaires who are the proponents of one or more of the TESCREAL ideologies (Gebru & Torres, 2024: 10). Billionaires such as Elon Musk, Sam Altman, Peter Thiel, and Marc Andreessen (Piquard, 2023; Taplin, 2023). Adjacent to the TESCREAL bundle is also techno-optimism and the techno-optimist manifesto, published by Marc Andreessen, in which the venture capitalist boldly claims that 'there is no material problem . . . that cannot be solved with more technology' and that 'Victim mentality is a curse in every domain of life, including in our relationship with technology — both unnecessary and self-defeating. We are not victims, we are conquerors' (Kelly, 2023).

Effective Altruism (EA), one part of the TESCREAL acronym, was initially focused on alleviating global poverty. However, in recent years, influential figures in the EA community have shifted their attention to the long-term future of humanity, one prominent member of the community being the head of OpenAI, Sam Altman. According to EA's ethical framework of 'totalist utilitarianism,' maximising the total value in the universe is the ultimate moral goal. Thus, bringing futures which have a net-positive effect on people's lives into existence is considered highly 'good,' while conversely, not doing so is morally wrong (Gebru & Torres, 2024: 7). In summary, Effective Altruists see the development of AGI as a means to generate immense economic value, and within their ideological framework, creating such a system is viewed as a moral obligation.

Other than venture capitalists and billionaires, there are also several influential academics linked to the TESCREAL lot, starting with Ray Kurzweil. In his 2005 book, *The Singularity Is Near*, Kurzweil argues that by the year 2045, computers will reach human-level intelligence, making it possible for humanity to merge with technology in what he calls a state of singularity (Corbyn, 2024). Next, there is Nick Bostrom, the co-founder of the World Transhumanist Association (2005: 12), as well as the person credited with founding the academic discipline concerned with the study of existential risk, or x-risk (Bucknall & Dori-Hacohen, 2022: 120). In a 2002 article, Bostrom defines existential risk as '[o]ne where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential.' A core tenet of the x-risk debate is the potential

emergence of *artificial general intelligence* (AGI) which outsmarts humanity and causes its demise (Mucci & Stryker, 2023). Reportedly, Elon Musk has been so influenced by Bostrom's *Superintelligence* (2014) that he donated 10 million dollars to Bostrom's Future of Life Institute (Dowd, 2017).

AI Takeover in Popular Media & Sci-Fi

Technochauvinists, as well as technocrats, firmly believe in the power of scientific knowledge and efficient code. However, the dreams, hopes, and ideologies that drive the technological elite often evoke themes reminiscent of science fiction (SF). Reflecting on this contradiction, Robert M. Geraci (2010) argues that portrayals of human annihilation in popular science books have their roots in traditional Judeo-Christian narratives, including immortality, the apocalypse, resurrection, and salvation. These age-old themes have seamlessly transitioned into contemporary media depictions of AI, mirroring similar narratives while paradoxically rejecting the idea of God. Furthermore, Elish and boyd (2018) point out that contemporary discourses around AI often emphasise the technology's potential over its actual functionality, creating a blend of fantasy and reality. This interplay between ancient narratives and modern technological aspirations illustrates how deeply ingrained cultural stories continue to shape our understanding of AI and its future.

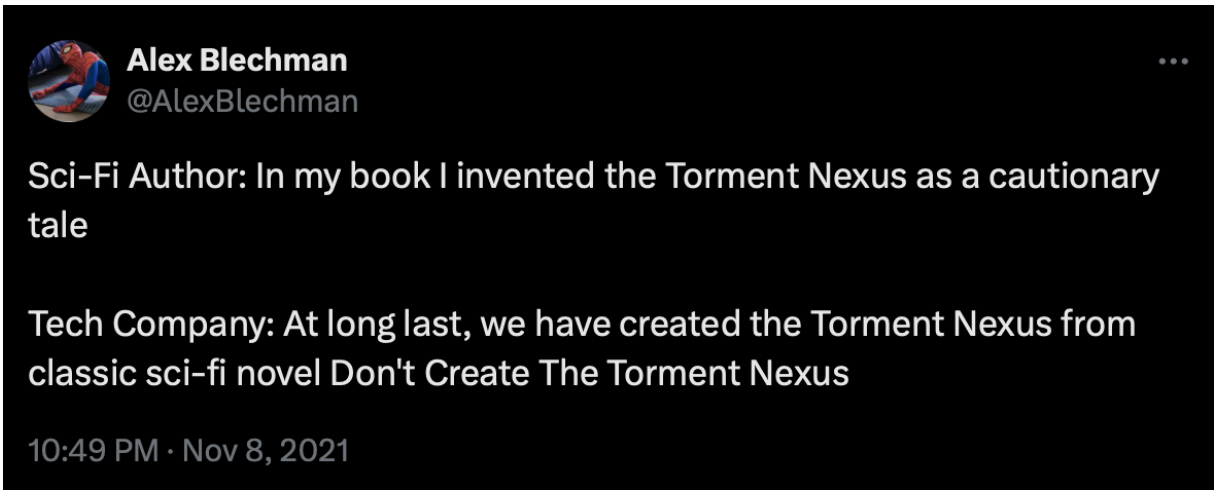


Figure 1: A tweet by @Alex Blechman (2021)

The portrayal of technologies in science fiction (SF) media has played a pivotal role in the real-world development of technology (Larson, 2008), creating a bi-directional pattern of influence between the two fields (Mubin *et al.*, 2016). For example, due to the general public's limited lived exposure to

robots, their understanding of robots is largely shaped by their portrayals in science fiction and movies (Mubin et al., 2019: 52). This research project extends this claim to AI technologies.

A clear example of this phenomenon occurred on 13 May 2024, when OpenAI announced its newest product, ChatGPT-4o, which featured an uncannily realistic and sometimes even flirty human voice synthesiser. This release was accompanied by a tweet from Sam Altman that consisted of just one word: 'Her' (Tassi, 2024). This was an obvious reference to the movie *Her* by Spike Jonze (2013), in which Joaquin Phoenix's character falls in love with an AI voice assistant. This instance demonstrates Hollywood's significant influence on the conceptualization of technological futures, and how what people see in movies and read in books actively shapes their technological dreams.

This relationship is particularly pronounced among the elite, tying back to Gebru's assertion that much of the funding for AI and AGI comes from the TESCREAL-aligned high society. This creates a situation where the line between science fiction and science is often blurred. For example, during the development of Stanley Kubrick's *2001: A Space Odyssey*, science fiction writer Arthur C. Clarke collaborated with his friend Marvin Minsky, a computer scientist credited with building the first neural network in 1950 (Russell & Norvig, 2016: 16). Together, they created HAL 9000, a now canonic representation of AI takeover (Broussard, 2018: 71). Similarly, in 2016, Facebook CEO Mark Zuckerberg built his own home assistant, named Jarvis, inspired by the AI in the Marvel franchise *Iron Man* (Gershgorn, 2016).

Given the observable instances and influences of science fiction (SF) on technological development, it is essential to examine the long history of robot takeovers in popular media, particularly within the SF genre. It is of note that SF has been viewed as a predominantly male domain, with its emphasis on science and technology often resulting in the exclusion of women and, consequently, the neglect of gender issues (Merrick, 2003).

According to Suvin (1979), SF storytelling has the unique ability to exaggerate and create a sense of unfamiliarity for audiences while constructing worlds that remain recognizable extensions of present reality. The themes of machines gaining self-consciousness, acting rebelliously towards their creators, and confronting humanity with the limits of their power can be traced back to ancient mythology, including tales such as the Golem, Prometheus, and the Garden of Eden (Goode, 2018: 189). One of the earliest cultural representations of robots leading to humanity's demise can be found in Karel

Čapek's 1920 play *R.U.R.*, which is notable for coining the term *robot* (Reilly, 2011). This narrative aligns with the fear expressed in Mary Shelley's novel, which has since been termed the Frankenstein Complex by prolific SF writer Isaac Asimov; it captures the anxiety surrounding machines surpassing the capabilities of their creators and ultimately posing existential threats (Beauchamp, 1980). A canonical example of popular media that embodies the theme of human annihilation caused by robots is the *Terminator* franchise (Richardson, 2015: 1).

In addition to popular culture and Hollywood, people's perceptions of scientific advances in AI are significantly shaped by the news media, as highlighted in the section on the media's framing and agenda-setting functions. Fast and Horovitz (2017) note that discussions surrounding AI have increased since 2009, generally maintaining a positive tone, although concerns about the potential for AI to take over have also emerged. This perspective is supported by Roe and Perkins, who analysed UK news media headlines during the first five months of 2023. Their findings revealed that the 'impending danger' narrative was the most prevalent, with 248 headlines, accounting for 37% of the total (Roe & Perkins, 2023). Moreover, in their 2019 analysis of the most widely read US daily newspapers from 2009 to 2018, Chuan et al. found that coverage of AI was predominantly framed within business and economic themes (Chuan *et al.*, 2019).

Political Economy of AI

The ongoing debate surrounding AI reveals a tension between perspectives that stem from technological determinism and those that highlight the social construction of technology. This tension underscores the multifaceted nature of AI discourse, where concerns about both existential risks and the immediate harms of AI coexist. Therefore, any analysis of AI x-risk discourse must take into account both the short-term risks as well as the broader political economy that underpins data processes, as data drives the development of AI systems. By examining these wider contexts, we can better understand how technological advancements are influenced by social dynamics and economic factors, ultimately shaping AI development and deployment.

Several theorists have sought to construct an overarching conceptual framework for the processes and stakeholders involved in AI development and Big Data generally. According to Cohen (2013: 13), Big Data encompasses more than just vast quantities of information, rather it involves an integration of technologies and processes. The technology handles the task of filtering and organising extensive

data sets, while the processes are responsible for identifying patterns, making predictions, and applying these predictive insights to new data. A key context for this discussion is *neoliberalism*, which is characterised as a political project that restructures and re-scales social relations to meet the demands of unrestrained global capitalism (Bourdieu, 1998). In addition to neoliberalism, Van Dijck (2014: 204) introduces the concept of *dataism*, defined as the widespread belief in the objective and quantifiable nature of data. Dataism thrives on the acceptance of rendering all human actions into quantifiable data points, which are gathered by private corporations, academia, or the state, leading to the collection of ever-increasing amounts of personal information.

The Big Data landscape has come to be dominated by few technology companies who exercise unrivalled power and control through quasi-monopolies (Micheli *et al.*, 2020). Complementary to this perspective are Shoshana Zuboff's concept of *Surveillance Capitalism* (2019) and Couldry and Mejias' notion of *Data Colonialism* (2019). Zuboff, on the one hand, argues that Big Data represents a new form of capital accumulation, where the market for behavioural surplus is intricately linked to AI development, relying on data extracted from users to train AI models. Surveillance capitalism, seen as a novel type of market economy, differs from traditional capitalism by leveraging a global infrastructure for data collection and analysis. This system exploits extensive surveillance, commodification, and control to modify human behaviour for profit. Similarly, Data Colonialism posits that data extraction practices mirror historical colonialism, emphasising the continuation of extractive and exploitative practices in the digital age. Mejias and Couldry (2024: 11) note that Christianity and Western science played pivotal roles in historical colonialism, and this legacy continues today as Western dominance persists in the control and development of AI.

It is crucial to address the short-term fears and threats posed by AI, such as surveillance through workplace monitoring, discrimination from algorithmic decisions, and the exploitation of workers and natural resources (Mejias & Couldry, 2024: 10). Research in the social sciences demonstrates that the techniques underpinning Big Data encode social values into mathematical processes and automate these through scalable normative logic (Ambrose, 2015).

The field of *AI ethics*, not to be confused with the EA adjacent *AI safety*, has prominent researchers advocating for intersectional feminist perspectives, highlights the unequal and disproportionate negative impacts of AI developed by the male-dominated and Western-centric industries of computer science, technology, and business (Buolamwini & Gebru, 2018; D'Ignazio & Klein, 2020; Noble, 2018;

O'Neil, 2016). These perspectives underscore the importance of considering diverse viewpoints in AI development to mitigate biases and ensure equitable outcomes. Importantly, a significant community of AI ethics researchers and activists who focus on the more immediate risks of AI often view the existential risk (x-risk) debate as a distraction from pressing current issues. This community emphasises the need to address the tangible and immediate impacts of AI on society, advocating for ethical practices and policies that prioritise fairness and accountability in AI systems.

RESEARCH QUESTION

Building on the theoretical and background sections, this study seeks to address the following research question:

Research Question: What discourses of existential risk posed by artificial intelligence are present in UK news media coverage, how are they constructed, and who are their proponents?

For the purposes of this project, the term 'x-risk' is adopted but its meaning is expanded beyond its use by the Effective Altruism (EA) community. In this text, x-risk is understood to encompass:

- Discourse portraying any type of AI, machine learning, or autonomous computing system as capable of causing human extinction;
- Themes such as the singularity, AI 'taking control' over humans, and discussions of autonomous AI weaponry.

By examining the research question, this study aims to shed light on how UK news media constructs the discourse around existential risks associated with artificial intelligence, leading to the potential shaping of its perception by the public, by governments, and the industry.

METHODOLOGY AND RESEARCH DESIGN

In a world, where as explored earlier, both the fields of AI and the media industry are dominated by a small number of powerful players, Critical Discourse Analysis (CDA) offers itself as a fitting framework for this research project, due to its focus on how imbalances of power are exercised and enacted in discourse. In this section, I will address the method and its suitability for this analysis, the research design, its limitations, along with my positionality as a researcher.

Rationale for Methodology

According to Fairclough (2013: 3, 4), *discourse* is a complex set of relations involving communication between people, such as talking, writing, and other forms of interaction. It is not an independent or discrete entity but is inherently relational, connected to broader social, power, and institutional structures. Discourse contributes to social life by bringing and creating meaning within these interconnected relations. Therefore, discourse can only be understood through its dialectical and relational connections with other elements in social processes.

CDA, as a method of academic inquiry, rejects the idea that language is neutral, attributing language with a power to construct (Gill, 1996). When an analyst conducts the analysis, they are to suspend belief in what is traditionally taken for granted, or, in other words, render the familiar strange (Potter & Wetherell, 1987). Using CDA as a research tool, the analyst is asking questions of the text at hand, questions such as 'How is the text positioned or positioning? Whose interests are served by this positioning? Whose interests are negated? What are the consequences of this positioning?' (Janks, 1997: 329). By asking questions like these, CDA looks at how power, dominance and inequity are performed, propagated, and resisted through language in both social and political contexts (van Dijk, 2005: 352). Additionally, CDA understands itself as committed to social change, taking the side of those who are disadvantaged (Jørgensen & Phillips, 2002: 75).

Since the logic and dynamics of any given society are never fully transparent, they require analysis (Fairclough, 2013: 231). This project will specifically utilise Fairclough's approach, seeing it as the most developed for the research of communication and society, both theoretically and methodologically. According to Fairclough, discourse is not only constructive but also constituted, meaning that it shapes social structures along with reflecting them. It is therefore a dialectical relationship between different social dimensions (Jørgensen & Phillips, 2002: 71, 72, 76). A *critical* analysis, based on Fairclough, involves addressing social wrongs, such as injustice and inequality, analysing their causes, resisting them, and exploring ways to overcome them (Fairclough, 2013: 231).

In his 2003 book *Analysing Discourse*, Fairclough focuses on the analysis of social change, particularly the transformations within contemporary capitalism, which he refers to as *new capitalism*. Fairclough argues that new capitalism, characterised by globalisation, the knowledge economy, and neoliberalism, has profound impacts on various aspects of social life, including politics, education,

and the environment. Building on Fairclough's framework, this analysis will explore the ongoing shift in capitalism driven by datafication and the widespread adoption of AI. This study aims to contribute to the field of CDA by examining how a particular popular technological discourse is reverberating through economic and social structures, thereby extending Fairclough's insights into the current era of capitalism.

This analysis will employ Fairclough's three-dimensional model, which includes *description*, *interpretation*, and *explanation*. In the descriptive phase, the focus is on examining the formal aspects of the text. The interpretative phase concerns the processes of producing and consuming the text, also known as *discursive practice*. Finally, the explanatory phase addresses the broader social structures surrounding the text, referred to as *social practice* (Fairclough, 2002). Fairclough sees discourse as 'oscillating' between focusing on a particular text and focusing on what he calls 'order of discourse', or in other words, the social structures around language and social practices (Fairclough, 2003: 6). The three-dimensional model allows for this oscillation to take place.

CDA achieves its goals by thoroughly analysing discursive practices, particularly in relation to texts and their connection to social practices. In this analytical approach, the genre and various discourses utilised in creating a text, such as those found in newspapers—like neoliberal cultural, and political discourses—are carefully scrutinised. In this examination, CDA elucidates how texts both shape and are shaped by broader social structures (Jørgensen & Phillips, 2002: 80). Coupled with an understanding of the agenda-setting and frame-building function of the media, it is evident how powerful actors may utilise media narratives in their favour. This assumption, within a paradigm which understands those who control orders of discourse as a factor in the maintenance of power (Fairclough, 2002: 31), leads to the conclusion that CDA is an adequate method for the analysis of the x-risk discourse within the UK newspaper media. By examining the discursive practices present in the newspaper coverage of 'apocalyptic' AI, this discourse's rising dominance in public discussions, along with its wider social implications, can be assessed.

Research Design

This section contains a detailed description of the research design to ensure replicability. The design was informed by findings from a pilot study conducted in advance of the research and has been adjusted to address issues encountered during this preliminary phase.

The pilot study revealed that news headlines can be dominated by a single event for several months, which suggests that a data collection time frame longer than two months is necessary to avoid thematic monotony. Additionally, it was noted that even when articles included the relevant keywords, the research theme was often mentioned only tangentially, with the bulk of the text focusing on unrelated topics. These insights led to several adjustments in the research design and keyword lists to ensure the collection of better data.

The focus of this analysis was on newspapers located in the United Kingdom of Great Britain & Northern Ireland with the source language being English. The electronic newspaper database LexisNexis was used for the creation of the corpus.

To capture a comprehensive range of articles, an extensive list of keywords was developed through an iterative and reflexive search process using the database. The final list of keywords is available in Appendix 1. During the database search, the 'Group Duplicates' setting was set to 'Moderate Similarity' to prevent duplicate articles. If duplicates still appeared, the more recent article was included in the corpus.

The data collection period covered six months following the commercial release of ChatGPT, from November 30, 2022, to May 30, 2023. Rather than conducting a longitudinal study, this research aims to capture a snapshot of the significant attention AI and its potential existential risks received from the media, governments, and academia during this time. This time frame follows a pivotal moment in AI development, providing a comprehensive overview of the evolving discussions around AI and existential risk in the news media. To ensure the corpus remains manageable for analysis, six months were selected instead of a full year.

An initial data-cleaning process was conducted to ensure the relevance of the articles within the corpus. This was done by carefully examining the articles and excluding those that do not mention the research theme and appear in the search results only accidentally.

The final selection of texts for analysis was guided by a structured and iterative approach. Initially, the articles will be grouped based on their formal features: the newspaper in which they were published, the author, length, political orientation of the newspaper, etc. Following this initial categorisation, each article was read and analysed to identify thematic categories, considering elements such as the centrality of the x-risk discourse within the article, the stakeholders mentioned,

and the dominant themes and events mentioned. Finally, articles from each category were chosen following the principle of theoretical saturation (Saunders *et al.*, 2018), which entails continuing to add articles to the corpus until no new themes or insights emerge from the analysis. In other words, the selection process was iterative; articles were reviewed and included progressively until additional articles ceased to provide novel contributions to the analysis and subsequent discussion. This method ensures that the final sample is both representative and manageable, encompassing the full spectrum of discourse while avoiding unnecessary repetition.

Following Fariclough's three-dimensional model, the analysis of each article was divided into three steps, starting with the textual dimension. This step includes analysing the vocabulary, grammar, syntax, and the use of rhetorical devices such as metaphors. In the second step, the production, distribution and consumption of the articles and the discourse was considered. Within this step, the stakeholders prevalent in the texts were analysed, along with the prominent institutions and events. Additionally, this section was concerned with the mapping of themes and narratives of social transformation, or the transformation of capitalism, as mentioned earlier. Lastly, the texts were interrogated from the perspective of how they fit into wider social and cultural structures.

Limitations & Positionality Statement

The limitations of this design included those of the LexisNexis electronic news database, given that any article which is not indexed by LexisNexis was not analysed. Additionally, due to the sheer number of articles published in the UK newspapers, this analysis was restricted to the six months following the release of ChatGPT. As a result of this, discursive moments and narratives present outside of this time frame were not considered. Moreover, because of the focus on the UK news media and the English language, the findings of this study may not be replicable across international contexts. Lastly, despite the long list of keywords, it must be acknowledged a complete coverage of all relevant articles has most likely not been achieved and is not fully feasible.

It is also important to note, that in the creation of categories during the data cleaning phase of this project, a large amount of subjective judgement has entered this research, seeing that the process of categorization, an essential part of data gathering, will always be subjective (Bowker & Star, 2008).

As to the limitations of the method, Critical Discourse Analysis, these are again linked to subjectivity. According to Fairclough (2002: 22), what a researcher sees in a text and what they consider worthy of

describing all depends on their subjective evaluation of said text. Because of this, a discourse analyst must acknowledge their position within and outside of the research through an awareness of their position in society. Following a tradition which rejects the idea of 'value-free' science, it is argued that scholarly discourse is the product and a part of the social fabric and social interactions (van Dijk, 2005: 352). I must therefore admit that my investment in the research topic and its contribution to the subjectivity of the questions I am asking mostly hinges on my left-wing politics, criticism of outsized corporate and government power and my scepticism of the media's independence from said power.

Additionally, I must admit that it was due to my empirical observation of the news media's focus on distant apocalyptic futures that I decided to undertake this research project in the first place. My previous engagement with the topic of AI, particularly in studying the political economy of AI and datafication, has been the driving force behind this inquiry. This situates me as someone in opposition to the discourses focused on distant and abstract potential outcomes of AI, and a researcher who has so far mostly focused on its immediate, real-world negative impacts.

In all qualitative research, the researcher is the tool through which research is produced (Lincoln & Guba, 1985). Thus, I acknowledge that my views and political stance influence my interpretation of the analysed texts, as well as my capacity to critically evaluate articles that strongly align with my personal beliefs.

SAMPLE AND ANALYSIS

A filtered keyword search of the LexisNexis database yielded 423 articles. As shown in Figure 2, the frequency of publishing on the topic increased over the data collection period, with a peak in May 2023. This peak aligns with the findings of Roe and Perkins (2023).

The initial data cleaning resulted in a corpus of 196 relevant articles, encompassing 33 newspapers in total. Appendix 2 lists these newspapers and the respective article counts. Sunday titles (e.g., *The Sunday Times*, *Mail on Sunday*, *The Observer*), online news sources (e.g., *Mail Online*), and regional editions of larger newspapers (e.g., *Scottish Daily Mail*) were categorised under their main brand names, such as *The Times* or *Daily Mail*. Figure 3 shows the distribution of articles between regional and national newspapers, with national newspapers further subdivided by format and political leaning, based on Wikipedia ('List of Newspapers in the United Kingdom', 2024; 'Quality Press', 2024;

'Tabloid Journalism', 2024). The political orientation of the newspapers was also grouped, with all newspapers which were left of centre grouped as *Left*, and right of centre grouped with *Right*.

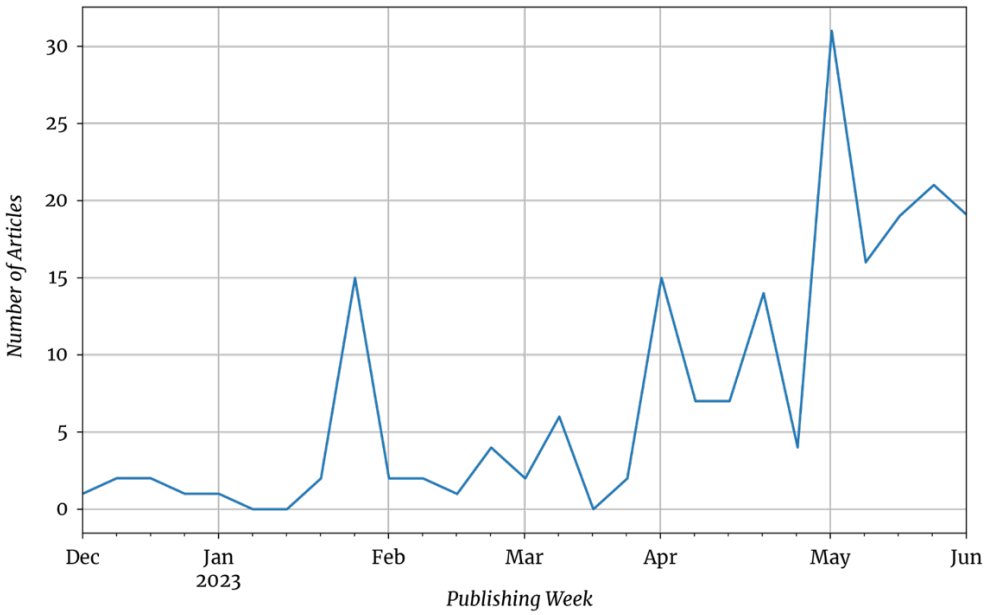


Figure 2: Number of Articles Published Throughout the Data Collection Period (Weekly)

An initial analysis of the corpus revealed that articles have a varying degree of focus on the x-risk discourse. Most articles only mentioned it peripherally, while relatively few discussed it as a central topic. The articles also varied in their stance towards the discourse, ranging from reinforcement to criticism.

All of the above was factored into the selection of articles for the final CDA. Through an iterative process that identified the key features of articles, 5 final articles were chosen, listed in Table 1. Links to versions of the articles available online can be found in Appendix 3. The annotated and coded versions of the articles are attached in Appendix 4.

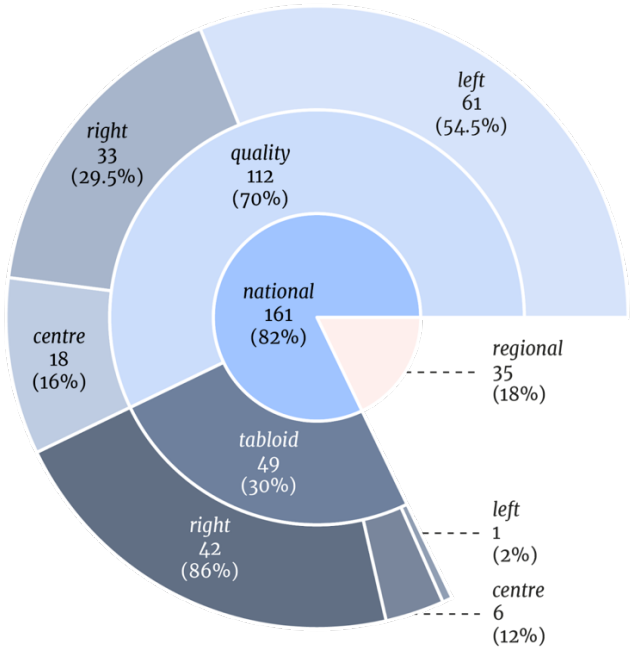


Figure 3: Distribution of Articles Across Regional vs. National Newspapers (National Newspapers Further Divided by Quality and Political Leaning)

Table 1: The final articles selected for CDA

Article #1	COULD A1 DESTORY HUMANITY?; Artificial Intelligence has advanced rapidly in recent years --and now some computers appear to have minds of their www.own.No wonder many of the world's leading experts are asking… (Leonard, 2023)
Article #2	AI same risk as nuclear wars, experts warn (Cuthbertson, 2023)
Article #3	The apocalypse isn't coming. We must resist cynicism and fear about AI (Marche, 2023)
Article #4	6 WAYS TO CONTAIN AI; The rapid rise of artificial intelligence is provoking panic. But Oxford professors John Tasioulas and Nigel Shadbolt argue that the humans are still in control (Shadbolt & Tasioulas, 2024)
Article #5	Analysis: Scottish politicians will soon need to tackle AI (Learmonth, 2023)

Article #1

The first article, 2040 words long, is structured in a way which helps build suspense. It employs rhetorical questions such as, 'Are we creating a monster that will enslave rather than serve us?', the

use of which is intended to provoke thought and engage the reader. Emotionally charged and sensationalising words like 'sinister,' 'monstrous,' and 'catastrophic' are used to evoke strong feelings of fear and concern.

Moving to the interpretative analysis, the article is published in the tabloid Daily Mail, a right-wing medium. Examples of perceived AI bias are highlighted in the article, such as ChatGPT's differing responses to requests for writing odes to President Biden and Donald Trump. The author uses this among the arguments for a left-leaning bias in AI systems. Interestingly, the author's use of 'we' when discussing the creation of AI attributes responsibility to a collective group of people. The article cites numerous authoritative figures, such as Elon Musk and Steve Wozniak, calling them 'tech world luminaries'. Similarly, the article name drops Sam Altman and Bill Gates; politicians Jeremy Corbyn, Donald Trump, Joe Biden and Ron DeSantis; and several other influential people. This is likely done to lend credibility to the arguments presented. The article references the often-used *paper clip problem* as an example scenario of how AI could reason to harm humans. Other potential risks mentioned are job loss, negative disruption in education, threats to privacy, and the potential of AI to create weapons of mass destruction. The portrayal of AI and computers as having achieved some level of autonomy is emphasised through phrases like 'artificial intelligence is already making a sinister impression on the world' and 'the potential for AI-induced mass destruction is already here.' In contrast, the author uses the phrase 'an unwitting human' when describing someone's interaction with ChatGPT, illustrating a sense of helplessness or lack of control in interactions with AI. Additionally, the author makes a direct reference to 'sci-fi dystopia' as something which is no longer fictional, but rather a 'battlefield reality.'

Finally, the article's sensationalising language reinforces the x-risk narrative, positioning AI as a significant threat. It frames AI and computers as entities with autonomy, suggesting they pose a potential danger to humanity, demonising these technological systems. The mention of China as a negative example of AI use reflects broader geopolitical tensions and anti-Chinese sentiment. The phrase 'Silicon Valley once again cynically follows money' underscores the power imbalance between tech companies and the general public, suggesting that these companies prioritise profit over societal well-being. The article concludes by suggesting that the development of AI should be abandoned altogether due to the potential x-risk and other risks, indicating a strong stance against AI progression.

Article #2

The second article is 439 words long, with the headline 'AI same risk as nuclear wars, experts warn' immediately signalling a comparison between AI risks and nuclear threats, grabbing attention and setting a serious tone. The article quotes from a source which uses adjectives such as 'aligned, trustworthy, and loyal' in reference to the capabilities they believe AI systems should have.

This article is published by The Independent, a left-wing quality newspaper, and authored by the most published writer on the topic within the corpus. The text is a report on the release of a statement calling for lawmakers and regulators to take '"severe risks" more seriously.' Quotes from prominent figures, such as Demis Hassabis, Sam Altman, Geoffrey Hinton, and Ilya Sutskever are used to reinforce the x-risk narrative. In terms of institutions, the article includes a tweet from the Center for AI Safety, as well as a mention of the Future of Life Institute, founded by Nick Bostrom.

Importantly, the piece is situated within a larger debate about the ethical and safety implications of rapidly advancing AI technologies. It reflects the growing concern of certain experts about the potential misuse of AI. They are centred around technology escaping human control, suggesting regulation as a means to address this issue. This article represents the influence of tech leaders and experts in shaping public policy and opinion about AI. By equating AI risks with nuclear threats, the article reinforces fears and concerns, potentially influencing public opinion and policymaking.

Article #3

The third article, starting with a headline disregarding the idea that the apocalypse is coming, rejects the threat of the x-risk from the very beginning. This is supported by the use of words such as 'silly' and 'a distraction' in reference to the discussions around the possibility of human extinction caused by AI. Additionally, the author categorises certain AI related risks as real, implying that the x-risk is not. The text also utilises comparisons to the popular media references *Star Trek* and *Terminator 2*, using one as a positive and one as a negative example of approaches to AI.

This text, published in the centre-left quality newspaper The Guardian, is critical of the x-risk discourse, a core theme in the piece. The text, written by an essayist and novelist, is an opinion piece, meant to persuade and provoke thought rather than simply inform. It opens up with a critique of *doomerism* and lists past hype cycles relating to technological breakthroughs, drawing parallels to

previous exaggerated claims such as the impact of WeWork on commercial real estate and cryptocurrency on banking. The article mentions the open letter calling for a six-month pause in the development of AI, highlighting figures like Elon Musk and OpenAI, but dismisses the x-risk narrative as fundamentally flawed and as a distraction from more pertinent issues. The emphasis is placed on the need for regulatory and societal responses to technological developments rather than succumbing to fear.

The text's explicit critique of capitalism is in line with the medium's overall political orientation. It argues that AI doomerism is a form of advertising, a means for tech companies to gain credibility or funding from investors. This critique situates the discourse within a broader critique of capitalist motivations and the self-serving nature of tech industry narratives. The piece underscores that the real issues, such as misinformation and economic inequality, predate AI and are more rooted in political and social structures than in the technology itself. The explanatory analysis also points out that the fear of AI is likened to religious belief, lacking scientific grounding and based more on cultural narratives and popular media than on evidence. This perspective is used to challenge the legitimacy of the doomsday discourse and to redirect attention towards tangible issues that require concrete regulatory actions, such as the impact of social media algorithms on mental health and the socioeconomic consequences of automation. Lastly, the text openly critiques the 'anthropomorphisation of statistical pattern-making,' from a point of view which foregrounds human agency and decision making in the process of technological development.

Article #4

In terms of its textual dimension, Article #4 uses a numbered structure to list ways to contain AI. The text is 1471 words long, with vocabulary choices indicating a high level of concern about AI's impact, such as 'panic,' 'threat,' 'alarm,' and 'extinction.' The article uses analogies and metaphors comparing AI to an 'arms race.'

Considering the interpretative analysis, the fourth article was published in The Times, a right-wing quality newspaper. The authors are two academics, giving a level of credibility and authority to the piece. The text is structured around highlighting six approaches to contain AI, notably warning against over-regulation, calling for a stop in listening to ethical experts, and encouraging to halt the international AI arms race. These critiques, specifically warning against 'over-regulation,' are aligned

with neoliberal discourse, favouring minimal government intervention in markets, fitting into the broader context of contemporary technology policy debates and in agreement with the political orientation of the newspaper. Moreover, the article's suggestion to stop following 'ethical experts' while the authors themselves hold positions at the Oxford Institute of Ethics in AI highlights an internal contradiction that reflects broader debates about authority in AI ethics, explored by Cocchiaro et al. (2024). The use of real-world examples, such as Geoffrey Hinton's resignation and the White House meeting, grounds the arguments in current events, lending them credibility but also reflecting a selective presentation of evidence.

Expanding on the explanatory analysis, the article's discussion of the global dimension of AI regulation touches on international relations and geopolitical power dynamics of AI, suggesting the need for a global consensus akin to nuclear arms control. The call for democratic and participatory deliberation about AI underscores a distrust of technocratic governance, highlighting the importance of public engagement. However, the article's subsequent focus on individual and consumer choices, such as buying or not buying a 'social robot,' shifts responsibility from the developers and makers of AI onto the consumers. With these aspects in mind, the article's critique falls short of elaborating on how AI developments influence global power structures and social and economic inequalities. Finally, the article's optimistic conclusion, which emphasises AI's potential benefits, reflects a view which acknowledges risks while promoting a hopeful vision for the future.

Article #5

The fifth article uses words such as 'lads' and 'dodgy,' which can be commonly found in Scottish English. Next, in terms of the textual dimension, the author uses the pronouns 'we' and 'our,' possibly to create a sense of collective responsibility and concern. Speaking of AI, the article uses the adjectives 'dystopian' and 'absolutely terrifying,' suggesting a negative connotation towards AI's potential impacts.

In terms of the interpretative analysis, this article was published in a Scottish regional newspaper and is therefore targeted at Scotts and references Scottish industry. In its rather lengthy opening section, the article uses humour and the local political context to engage the audience. In terms of x-risk, it is mentioned relatively briefly, speaking of an event where the UK's prime minister met with prominent figures from the AI industry to discuss the mitigation of an 'existential threat.' The text then goes on

to call for the regulation of AI, as a statement coming from Sam Altman. Other than x-risk, the potential risk of AI creating disinformation is elaborated on. This article closes by again speaking of both the positive and the negative potential impacts AI could have.

The article, through sections such as 'Given how important the life sciences sector is to Scotland, this could be a huge opportunity' reveals how the economic and social impact of AI is contentious on a region-to-region level, with differing impact and importance. Overall, this article uses the x-risk theme to open a discussion about both the positives and the negatives of AI, particularly in the Scottish context.

DISCUSSION

Overall, a relatively small section of each of the five articles is dedicated to the nitty gritty of the existential risk scenario and how such a scenario might occur. In actuality, the articles present the x-risk as a quote from some person of influence and then go on to focus on some related or unrelated topic. Due to this, the question interrogating the details of the existential risk theme in the discourse remains somewhat overshadowed by other discussions, marked by the tendency to mention the x-risk briefly before diverting to different subjects.

The discourse surrounding the future of AI technologies plays a crucial role in shaping socio-technical futures (Bechtold *et al.*, 2017: 86). The first article, for instance, portrays AI systems as having already achieved significant dominance over humans. It uses phrases like 'the hired hand duly obliged,' referring to a human deceived by a natural language model prompt. This language implies that the model possesses autonomy and can compel human actions. Similarly, the text positions 'us' (humanity) against 'them' (machines), reinforcing a techno-deterministic view that technology follows an internal logic and path of progress.

Article four also uses deterministic language, stating that AI 'is already working with artists,' suggesting a level of autonomy in AI's collaboration with artists and musicians. This reflects the technological determinism perspective, which views technology as an autonomous force rather than a product of human actions (MacKenzie & Wajcman, 1999).

Moreover, the articles frequently anthropomorphise AI, attributing human-like qualities to these systems. For example, the second article discusses the 'ideal features' of an AI system, describing it

as 'loyal,' among other attributes. This anthropomorphism further blurs the line between human and machine, fostering a perception of AI as a sentient being with its own intentions and characteristics.

Next, as mentioned earlier, the theme of x-risk is in multiple articles mostly used as a vehicle to discuss other issues. This can be best exemplified in the 5th article, as well as the first and fourth articles. In the article published by Herald Scotland, AI's potential material costs and economic impacts on Scottish industry are discussed, highlighting the possible realities that the implementation of AI might bring on a region-to-region level. Another issue, one that is mentioned in two articles, is the so-called 'AI arms race' with China, aligned with the western-centric, specifically, U.S.-centric foreign policy agenda. This arms race is generally a common argument for the investment in and advancement of AI, especially by the Effective Altruism community, with reports of EA groups lobbying for the ban of semiconductor sales to China in order to halt China's AI development (J. Davis, 2024).

Examining the construction of the AI existential risk discourse, the concepts of technological myth and digital sublime, as well as enchanted determinism, are particularly pertinent.

According to Vincent Mosco (2004: 18), technology is most powerful not during its mythic period but when it becomes banal. Following this logic, as computers withdrew from their spot as hailing a great social transformation and turned into ubiquitous tools, they gave way for Artificial Intelligence and the surrounding myth to arise. Reflecting on Nye (1994) and the demonisation of technology as the foundation of the technological sublime, the first, second, and fourth articles ultimately present AI as something to be reigned in to be stopped from causing destruction.

Assessing the x-risk discourse as a myth, we can first borrow Natale and Ballatore's (2020) characterisation. The construction of the myth begins with the use of analogies and metaphors, or discursive shifts, which the analysed texts offer a myriad of. Analogies such as AI being an 'alien,' or as much of a challenge as nuclear arms during the Cold War. On top of that, when sci-fi media such as *Star Trek* and *Terminator* are mentioned, or direct claims like 'killer robots are no longer a sci-fi dystopia' are used, while on the one hand, they promote discussions about the role of technology in society, on the other, they encourage confusion between the present capabilities of AI and fiction. As Elish and boyd (2018: 62) note, popular media coverage often, albeit inadvertently, reinforces this blurring of the line between fantasy and reality. Furthermore, the rhetorical emphasis on AI's

potential for future destruction dominates over its current capabilities, aligning with the framework's myth-construing steps. In other words, as Broussard (2018: 11) suggests, the news articles focus more on the possible capabilities of AGI—the 'stuff of dreams'—rather than on the AI we currently have. Lastly, Natale and Ballatore highlight the role of controversies, which in this case is very fitting. This analysis shows that the x-risk discourse is a controversial issue, underlined by the fact that there are a large number of experts who chip in either in opposition or in support.

The x-risk discourse can also be assessed based on Barthes' (1957) contention that a myth is depoliticised speech, devoid of tension and conflict, gaining new ideological meaning through critique and reevaluation. This is exemplified in the first article, where various alarming anecdotes are presented, such as an AI tricking a human into completing a CAPTCHA test and speculations about AI achieving self-awareness. These narratives abstract from the specific socio-political and economic contexts in which these technologies occur and move away from questions about who controls AI development, the ethical considerations in its design, and the broader impacts on society. Instead, the focus is placed on sensational and generalised threats. This depoliticised speech is then imbued with new ideological significance when the fear of AI is reframed as a universal existential threat, which sidesteps more complex discussions about regulation, corporate responsibility, and ethical AI development. This is underscored when the author quotes several prominent figures like Elon Musk and Bill Gates, who warn about AI's uncontrollable and potentially catastrophic consequences. These warnings are presented as if they are natural and unavoidable outcomes of technological progress, rather than the results of specific choices made by powerful corporations and technologists. The critique and reassessment aspect of Barthes' myth is also evident in how the article calls for a pause in AI research, as suggested by tech luminaries in an open letter. This reassessment is framed as a necessary precaution to mitigate the risks posed by AI. However, this falls short of a proper critique, failing to address deeper issues such as the monopolisation of AI technology by a handful of large corporations, the lack of diverse voices in AI ethics discussions, and the socio-economic inequalities exacerbated by AI advancements. The same is evident in the second article, where the existential threat is treated prima facie, with prominent voices from within the AI industry being presented without questioning their intentions. By solely focusing on statements from these figures, the article shows how these leaders leverage their authority to shape the narrative and possibly avoid scrutiny.

Even the third article, as critical of the discourse as it is, ultimately wraps up by claiming that AI is an 'alien' which not even the engineers or developers fully understand. The same applies to the fourth article and quotes such as 'miraculous-seeming results'. This reflects a trend described by Campolo and Crawford (2020), who point out that despite the critical scrutiny, when AI systems operate effectively, they often evoke a sense of wonder and mystery, defying the epistemology of disenchantment which posits that no mysterious forces govern the world.

When asking questions about whose interests are served by this positioning, Mosco's (2004: 32) assertion that using myths to market products is an age-old tactic is illuminating. Additionally, reporting from the Guardian has pointed to how the UK's emphasis on apocalyptic AI could be beneficial to AI businesses and for the country's economy in turn (Bhuiyan, 2023).

Looking at the article at hand, they all share a reliance on the reproduction and amplification of what a small number of particular people have said on the topic of x-risk, with many of them having considerable skin in the AI game. When framing an issue, journalists negotiate between themselves and other social actors (Cook, 2010). Their work can be approached as purely passing on of the frames provided by other actors, such as industry figures, or their own interpretation and frame-setting (Brüggemann, 2014).

In the case of the x-risk discourse and the articles analysed, there is a high frequency of citations from people such as Sam Altman, Elon Musk, and Geoffrey Hinton, along with frequent mentions of Open AI. This is in accordance with Brennen et al.'s (2018) analysis of the UK's media coverage of AI, conducted in the first 8 months of 2018, which revealed that 33% of the unique sources cited were from the industry, nearly double the number from academia and six times those from government sources. Out of those texts, 12% referenced Elon Musk. Additionally, more than half of the articles published focused on AI products, framing them as solutions to problems, with some outlets emphasising their potential over actual functionality, thus blurring the line between what is achievable and what is aspirational (Brennen *et al.*, 2018).

Interestingly, in all five articles, several figures and institutions involved with Effective Altruism are mentioned, yet EA as an ideology or interest group is not referenced in any of them. As reported in 2024, two EA-affiliated lobbying groups have been active in Washington to 'protect humanity against the alleged extinction risk posed by artificial intelligence' (Bordelon, 2024).

Overall, not many opinions from specialists outside of computer science, the AI industry, or the government are presented. In most of the texts, academics and technology leaders are represented as the same group of 'experts,' or at least adjacent to each other. This is the case also due to the reality of the industry, where frequent exchanges happen between the academic and the industry workforces, as well as the industry conducting research under their private for-profit brand. This situation, where the loudest voices calling for regulation are those who are also called upon to draft such regulation is what is known as *regulatory capture*. Lewis Liu (2023) fittingly described the situation in a Financial Times article as leading AI companies being '... both the ones shouting "fire" in an empty theatre, and the ones turning up with the fire engine.' This is particularly alarming, since with an exception of the third article, none of the texts truly question the intentions of those who speak on the existential threat of AI, as if the topic was so disarming that no space is left to put the good will of those who advocate for it into question.

An insightful analysis involves considering what is left unsaid, and in this case it is particularly the absence of non-male and non-western perspectives, as these articles solely reference male sources from the global north. One of the only women mentioned in all five articles is portrayed as the unfortunate victim of having her picture misused (Article #5). Within the social constructivist framework, hierarchies based on sexual differences significantly influence the design, development, and distribution of technology. Technology, today, serves as a key source of male power and a defining aspect of masculinity (Wajcman, 2004), and it is therefore crucial to view the x-risk debate through a gendered lens. Acknowledging the constructive and constitutive functions of both discourse and technology, it is essential to recognize that the current discourse is fully captured by the Frankenstein Complex, and deserving of hope to escape from its claws. By harnessing the power of imagination and science fiction, we can aspire to create more equitable technological futures. This shift in perspective opens the door to envisioning technologies not solely defined by fear, destruction, or domination, but rather by inclusivity, compassion, and collaboration.

CONCLUSION

This analysis has revealed that UK news media discourses on AI existential risk are characterised by an alarmist rhetoric, anthropomorphism, and technological determinism, often driven by influential tech industry figures. The articles usually briefly touch on catastrophic AI scenarios but frequently

shift to other broader topics like economic impacts, foreign policy and international competition, such as the 'AI arms race' with China. The discourse often constructs AI as a technological myth using sci-fi analogies, sensationalising the threat and depoliticizing the discussion by avoiding deeper regulatory and ethical considerations. Even though sci-fi is not directly mentioned too often, there is a clear link between the concept of the x-risk itself and its science fiction past, as explored in the background section. This connection between x-risk and science fiction often blurs the line between present capabilities and fictional narratives, promoting an irrational rationalisation of AI's potential dangers. As Campolo and Crawford (2020: 13) note, 'We are not being confronted with a sublime form of superhuman intelligence, but a form of complex statistical modeling and prediction that has extraordinarily detailed information about patterns of life but lacks the social and historical context that would inform such predictions responsibly.'

Prominent voices like Elon Musk, Sam Altman, and Geoffrey Hinton dominate these narratives, reflecting media reliance on industry insiders and suggesting a form of regulatory capture. Despite significant involvement, the Effective Altruism community's influence is often implicit rather than explicitly mentioned, obscuring ideological motivations.

A notable absence of diverse perspectives, especially feminist voices, highlights existing gender hierarchies within the tech industry. This underscores the need for a more inclusive and nuanced discussion on AI's future, reflecting on diverse impacts and ethical considerations. Ultimately, these narratives are constructed through a blend of alarmism, technological determinism, and myth-making, often diverting to broader issues while lacking critical and diverse perspectives.

Overall, this comprehensive analysis indicates that UK news media portray AI existential risks in a sensationalised manner, often driven by prominent industry voices, lacking depth and critical, diverse viewpoints necessary for a balanced understanding.

This research project has contributed to a deeper understanding of the discourse by critically examining five articles and identifying the key themes which shape the x-risk narrative. Notably, this analysis reveals that these articles often treat x-risk as a vehicle for other discussions rather than as a central focus, leading to a depoliticized and sensationalised discourse. This shift from nuanced socio-political contexts to alarmist framing diminishes the complexity of AI's impact on society.

The implications of these findings are multifaceted. First, they highlight the need for more balanced representations of AI risks that include diverse voices, especially from outside the dominant industry perspective. By acknowledging the role of Effective Altruism and the influence of a few key figures, we underscore the importance of scrutinising who benefits from specific discourses about AI. This insight calls into question the authority of certain voices in shaping public perception and policy, suggesting that a more comprehensive dialogue is necessary for responsible AI development. Furthermore, viewing the x-risk debate through a gendered lens emphasises the need for inclusivity in discussions about technology and its future, advocating for a broader understanding of what AI could represent beyond dystopian scenarios.

There are several recommendations for future research which stem from this analysis. Firstly, exploring narratives that present AI in a more positive light could provide valuable insights into collaborative and ethical applications, moving beyond a focus on existential threats. Next, research into audience perception, examining how audiences react to the x-risk discourse would enrich our understanding of the discourse's functioning. Lastly, further research into the concept of regulatory capture in AI development is crucial to understand how the interests of industry leaders may influence policy-making without sufficient oversight.

Despite these contributions, this study has its limitations. Since the analysis is based solely on five articles, the full spectrum of discourse surrounding AI and x-risk is not captured. Additionally, the focus on exclusively mainstream media sources means that alternative perspectives found in independent or academic literature may have been overlooked. Lastly, while we touched upon gendered perspectives, a more in-depth analysis of intersectionality regarding race, class, and other social factors in AI discussions remains necessary.

BIBLIOGRAPHY

Alex Blechman [@AlexBlechman]. (2021, November 8). *Sci-Fi Author: In my book I invented the Torment Nexus as a cautionary tale Tech Company: At long last, we have created the Torment Nexus from classic sci-fi novel Don't Create The Torment Nexus* [Tweet]. Twitter. https://x.com/AlexBlechman/status/1457842724128833538

Ambrose, M. L. (2015). *Lessons from the Avalanche of Numbers: Big Data in Historical Perspective*. *11*.

Barbrook, R., & Cameron, A. (1996). The Californian ideology. *Science as Culture*, *6*(1), 44–72. https://doi.org/10.1080/09505439609526455

Barthes, R. (1957). *Mythologies*. Seuil.

Beauchamp, G. (1980). The Frankenstein Complex and Asimov's Robots. *Mosaic: An Interdisciplinary Critical Journal*, *13*(3/4), 83–94.

Bechtold, U., Fuchs, D., & Gudowsky, N. (2017). Imagining socio-technical futures – challenges and opportunities for technology assessment. *Journal of Responsible Innovation*, *4*(2), 85–99. https://doi.org/10.1080/23299460.2017.1364617

Bhuiyan, J. (2023, October 31). How the UK's emphasis on apocalyptic AI risk helps business. *The Guardian*. https://www.theguardian.com/technology/2023/oct/31/uk-ai-summit-tech-regulation

Bordelon, B. (2024, February 23). *AI doomsayers funded by billionaires ramp up lobbying*. POLITICO. https://www.politico.com/news/2024/02/23/ai-safety-washington-lobbying-00142783

Bostrom, N. (2002). Existential Risks: Analyzing Human Extinction Scenarios. *Journal of Evolution and Technology*, *9*(1).

Bostrom, N. (2005). A History of Transhumanist Thought. *Journal of Evolution and Technology*, *14*(1). https://jetpress.org/volume14/bostrom.html

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies* (First edition). Oxford University Press.

Bourdieu, P. (1998). *Practical reason: On the theory of action*. Stanford University Press.

Bowker, G. C., & Star, S. L. (2008). *Sorting things out: Classification and its consequences* (1. paperback ed., 8. print). MIT Press.

Brennen, J. S., Howard, P. N., & Nielsen, R. K. (2018). *An industry-led debate: How UK media cover artificial intelligence*. https://doi.org/10.60625/RISJ-V219-D676

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. The MIT Press.

Brüggemann, M. (2014). Between Frame Setting and Frame Sending: How Journalists Contribute to News Frames: Between Frame Setting and Frame Sending. *Communication Theory*, *24*(1), 61–82. https://doi.org/10.1111/comt.12027

Buck-Morss, S. (2000). *Dreamworld and catastrophe: The passing of mass utopia in East and West*. MIT Press.

Bucknall, B. S., & Dori-Hacohen, S. (2022). Current and Near-Term AI as a Potential Existential Risk Factor. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 119–129. https://doi.org/10.1145/3514094.3534146

Butterfield, A. B., Ngondi, G. E. N. E., & Kerr, A. K. (2016). Artificial intelligence. In A. Butterfield, G. E. Ngondi, & A. Kerr (Eds.), *A Dictionary of Computer Science*. Oxford University Press. https://www.oxfordreference.com/display/10.1093/acref/9780199688975.001.0001/acref-9780199688975-e-204

Campolo, A., & Crawford, K. (2020). Enchanted Determinism: Power without Responsibility in Artificial Intelligence. *Engaging Science, Technology, and Society*, *6*, 1–19. https://doi.org/10.17351/ests2020.277

Chuan, C.-H., Tsai, W.-H. S., & Cho, S. Y. (2019). Framing Artificial Intelligence in American Newspapers. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 339–344. https://doi.org/10.1145/3306618.3314285

Cocchiaro, M. Z., Morley, J., Novelli, C., Panai, E., Tartaro, A., & Floridi, L. (2024). *Who is an AI Ethicist? An Empirical Study of Expertise, Skills, and Profiles to Build a Competency Framework*. https://doi.org/10.2139/ssrn.4891907

Cockburn, C. (1999). Caught in the wheels: The high cost of being a female cog in the male machinery of engineering. In D. A. MacKenzie & J. Wajcman (Eds.), *The social shaping of technology* (2nd ed, pp. 126–134). Open University Press.

Cohen, J. E. (2013). What Privacy Is For. *Harvard Law Review*, *126*(7), 1904–1933.

Cook, T. E. (2010). *Governing with the news: The news media as a political institution* (2. ed., repr). Univ. of Chicago Press.

Cooper, M. J., Dimitrov, O., & Rau, P. R. (2001). A Rose.com by Any Other Name. *The Journal of Finance*, *56*(6), 2371–2388.

Corbyn, Z. (2024, June 29). AI scientist Ray Kurzweil: 'We are going to expand intelligence a millionfold by 2045'. *The Observer*. https://www.theguardian.com/technology/article/2024/jun/29/ray-kurzweil-google-ai-the-singularity-is-nearer

Couldry, N., & Mejias, U. A. (2019). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media*, *20*(4), 336–349. https://doi.org/10.1177/1527476418796632

Cuthbertson, A. (2023, May 31). *AI same risk as nuclear wars, experts warn*. The Independent. https://www.independent.co.uk/tech/ai-existential-risk-deepmind-openai-b2348380.html

Davis, J. (2024, February 23). *Effective Altruists Are Stoking Tensions With China*. Jacobin. https://jacobin.com/2024/02/effective-altruism-china-ai-us-policy

Dowd, M. (2017, March 26). *Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse*. https://www.vanityfair.com/news/2017/03/elon-musk-billion-dollar-crusade-to-stop-ai-space-x

Droog, E., Burgers, C., & Kee, K. F. (2020). How journalists and experts metaphorically frame emerging information technologies: The case of cyberinfrastructure for big data. *Public Understanding of Science*, *29*(8), 819–834. https://doi.org/10.1177/0963662520952542

Elish, M. C., & boyd, danah. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs*, *85*(1), 57–80. https://doi.org/10.1080/03637751.2017.1375130

Fairclough, N. (2002). *Language and power* (2. ed., 2. impr). Longman.

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research* (1. ed., repr). Routledge.

Fairclough, N. (2013). *Critical discourse analysis: The critical study of language* (2. ed., [Nachdr.]). Routledge.

Fast, E., & Horvitz, E. (2017). Long-Term Trends in the Public Perception of Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1). https://doi.org/10.1609/aaai.v31i1.10635

Gebru, T., & Torres, É. P. (2024). The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. *First Monday*. https://doi.org/10.5210/fm.v29i4.13636

Geraci, R. M. (2010). THE POPULAR APPEAL OF APOCALYPTIC AI. *Zygon®*, *45*(4), 1003–1020. https://doi.org/10.1111/j.1467-9744.2010.01146.x

Gershgorn, D. (2016, December 19). *Mark Zuckerberg made his own version of Amazon's Alexa to power his home*. Quartz. https://qz.com/867062/facebook-fb-ceo-mark-zuckerberg-built-an-ai-personal-assistant-named-jarvis

Gill, R. (1996). Discourse analysis: Practical implementation. In *Handbook of qualitative research methods for psychology and the social sciences*.

Gonsalves, T. (2019). The Summers and Winters of Artificial Intelligence. In M. Khosrow-Pour, D.B.A. (Ed.), *Advanced Methodologies and Technologies in Artificial Intelligence, Computer Simulation, and Human-Computer Interaction:* IGI Global. https://doi.org/10.4018/978-1-5225-7368-5

Goode, L. (2018). Life, but not as we know it: A.I. and the popular imagination. *Culture Unbound*, *10*(2), Article 2. https://doi.org/10.3384/cu.2000.1525.2018102185

Hern, A. (2024, April 10). AI race heats up as OpenAI, Google and Mistral release new models. *The Guardian*. https://www.theguardian.com/technology/2024/apr/10/ai-race-heats-up-as-openai-google-and-mistral-release-new-models

Janks, H. (1997). Critical Discourse Analysis as a Research Tool. *Discourse: Studies in the Cultural Politics of Education*, *18*(3), 329–342. https://doi.org/10.1080/0159630970180302

Jonze, S. (Director). (2013). *Her* [Drama, Romance, Sci-Fi].

Jørgensen, M., & Phillips, L. (2002). *Discourse Analysis as Theory and Method*. SAGE Publications Ltd. https://doi.org/10.4135/9781849208871

Kelly, J. (2023, October 22). *I read Andreessen's 'techno-optimist manifesto' so you don't have to*. https://www.ft.com/content/7eeb105d-7d79-4a59-89be-e18cd47be68f

Larson, J. (2008). Limited imagination: Depictions of computers in science fiction film. *Futures*, *40*(3), 293–299. https://doi.org/10.1016/j.futures.2007.08.015

Learmonth, A. (2023, May 29). *Analysis: Scottish politicians will soon need to tackle AI*. The Herald. https://www.heraldscotland.com/politics/23554299.political-campaigns-saying-ai-pie/

Lecheler, S., & Vreese, C. de. (2019). *News framing effects*. Routledge.

Leonard, T. (2023, March 29). *Could AI destroy humanity? Experts warn of 'catastrophic consequences'*. Mail Online. https://www.dailymail.co.uk/news/article-11917259/Could-AI-destroy-humanity-Experts-warn-catastrophic-consequences.html

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Nachdr.). Sage.

List of newspapers in the United Kingdom. (2024). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=List_of_newspapers_in_the_United_Kingdom&oldid=1234680385

Liu, L. (2023, October 26). *Letter: Setting rules for AI must avoid regulatory capture by Big Tech*. https://www.ft.com/content/6a1f796b-1602-4b07-88cd-4aa408cf069a

MacKenzie, D. A., & Wajcman, J. (Eds.). (1999). Introductory essay: The social shaping of technology. In *The social shaping of technology* (2nd ed). Open University Press.

Marche, S. (2023, May 15). The apocalypse isn't coming. We must resist cynicism and fear about AI. *The Guardian*. https://www.theguardian.com/commentisfree/2023/may/15/artificial-intelligence-cynicism-technology

Markoff, J. (2006). *What the dormouse said: How the sixties counterculture shaped the personal computer industry* (Paperback edition). Penguin Books.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12. https://doi.org/10.1609/aimag.v27i4.1904

McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *The Public Opinion Quarterly*, *36*(2).

Mejias, U. A., & Couldry, N. (2024). *Data grab: The new colonialism of big tech and how to fight back*. WH Allen.

Merrick, H. (2003). Gender in science fiction. In E. James & F. Mendlesohn (Eds.), *The Cambridge Companion to Science Fiction* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CCOL0521816262

Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, *7*(2), 205395172094808. https://doi.org/10.1177/2053951720948087

Mosco, V. (2004). *The digital sublime: Myth, power, and cyberspace*. MIT Press.

Mubin, O., Obaid, M., Jordan, P., Alves-Oliveria, P., Eriksson, T., Barendregt, W., Sjolle, D., Fjeld, M., Simoff, S., & Billinghurst, M. (2016). Towards an Agenda for Sci-Fi Inspired HCI Research. *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology*, 1–6. https://doi.org/10.1145/3001773.3001786

Mubin, O., Wadibhasme, K., Jordan, P., & Obaid, M. (2019). Reflecting on the Presence of Science Fiction Robots in Computing Literature. *ACM Transactions on Human-Robot Interaction*, *8*(1), 5:1-5:25. https://doi.org/10.1145/3303706

Mucci, T., & Stryker, C. (2023, December 18). *What Is Artificial Superintelligence?* IBM. https://www.ibm.com/topics/artificial-superintelligence

Natale, S., & Ballatore, A. (2020). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence: The International Journal of Research into New Media Technologies*, *26*(1), 3–18. https://doi.org/10.1177/1354856517715164

Nielsen, R. K. (2019). Economic Contexts of Journalism. In K. Wahl-Jorgensen & T. Hanitzsch (Eds.), *The handbook of journalism studies* (2nd edition). Routledge.

Nye, D. E. (1994). *American technological sublime*. MIT Press.

OpenAI. (2022, November 30). *Introducing ChatGPT*. https://openai.com/blog/chatgpt

OpenAI Charter. (n.d.). Retrieved 17 May 2024, from https://openai.com/charter/

Pennachin, C., & Goertzel, B. (2007). Contemporary Approaches to Artificial General Intelligence. In B. Goertzel & C. Pennachin (Eds.), *Artificial General Intelligence* (pp. 1–30). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-68677-4_1

Piquard, A. (2023, November 30). 'The Sam Altman saga shows that AI doomers have lost a battle'. *Le Monde.Fr*. https://www.lemonde.fr/en/opinion/article/2023/11/30/the-sam-altman-affair-has-shown-that-artificial-intelligence-doomers-have-lost-a-battle_6301213_23.html

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour*. Sage.

Quality press. (2024). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Quality_press&oldid=1199957847

Reilly, K. (2011). From Automata to Automation: The Birth of the Robot in R.U.R. (Rossum's Universal Robots). In K. Reilly, *Automata and Mimesis on the Stage of Theatre History* (pp. 148–176). Palgrave Macmillan UK. https://doi.org/10.1057/9780230347540_6

Richardson, K. (2015). *An Anthropology of Robots and AI: Annihilation Anxiety and Machines* (0 ed.). Routledge. https://doi.org/10.4324/9781315736426

Roe, J., & Perkins, M. (2023). 'What they're not telling you about ChatGPT': Exploring the discourse of AI in UK news media headlines. *Humanities and Social Sciences Communications*, *10*(1), 753. https://doi.org/10.1057/s41599-023-02282-w

Russell, S. J., & Norvig, P. (with Davis, E., & Edwards, D.). (2016). *Artificial intelligence: A modern approach* (Third edition, Global edition). Pearson.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity*, *52*(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8

Shadbolt, N., & Tasioulas. (2024, August 12). *Time is running out: Six ways to contain AI*. https://www.thetimes.com/business-money/technology/article/time-is-running-out-six-ways-to-contain-ai-bs20nrk7f

Smith, M. R., & Marx, L. (1994). *Does technology drive history?: The dilemma of technological determinism*. MIT Press.

Suvin, D. (1979). *Metamorphoses of science fiction: On the poetics and history of a literary genre*. Yale University Press.

Tabloid journalism. (2024). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Tabloid_journalism&oldid=1231476019

Taplin, J. (2023, August 22). *How Musk, Thiel, Zuckerberg, and Andreessen—Four Billionaire Techno-Oligarchs—Are Creating an Alternate, Autocratic Reality*. Vanity Fair. https://www.vanityfair.com/news/2023/08/musk-thiel-zuckerberg-andreessen-alternate-autocratic-reality

Tassi, P. (2024, May 21). *ChatGPT, Scarlett Johansson And Missing The Point Of 'Her'*. https://www.forbes.com/sites/paultassi/2024/05/21/chatgpt-4o-scarlett-johansson-and-missing-the-point-of-her/

Van Dijk, T. A. (1995). *Power and the news media*. https://discourses.org/wp-content/uploads/2022/07/Teun-A.-van-Dijk-1995-Power-and-the-news-media.pdf

van Dijk, T. A. (2005). 18 Critical Discourse Analysis. *The Handbook of Discourse Analysis*.

Wajcman, J. (2004). *TechnoFeminism*. Polity.

Wajcman, J. (2010). Feminist theories of technology. *Cambridge Journal of Economics*, *34*(1), 143–152. https://doi.org/10.1093/cje/ben057

Weber, M. (1946). Science as a Vocation (H. H. Gerth & C. Wright M, Trans.). *From Max Weber: Essays in Sociology*.

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121–136.

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

APPENDICES

Appendix A: Keywords Used for Database Search

"Artificial Intelligence" OR "Machine Learning" OR "Artificial General Intelligence"	AND	"existential risk" OR "existential risks" OR "existential threat" OR "existential threats" OR "catastrophic risk" OR "catastrophic risks" OR "global catastrophe" OR "extinction" OR “doomsday” OR "rogue" OR "apocalypse" OR "extermination" OR "singularity" OR "unaligned" OR “superintelligence”

Appendix 2: Newspapers from the corpus and respective article counts

Newspaper	**No. of Articles**
Daily Mail	32
The Independent	31
The Guardian	30
The Times	19
The Telegraph	14
Financial Times	12
The Sun	9
i-news	6
Daily Star	5

Evening Standard	5
The Herald	4
Scotsman	3
Daily Record and Sunday Mail	2
Aberdeen Press and Journal	2
The Western Mail	2
Wales Online	2
Yorkshire Post	2
Metro	1
Daily Mirror	1
Daily Express	1
Biggleswade Today	1
Eastern Daily Press	1
Garavi Gujarat	1
Jersey Evening Post	1
The Northern Echo (Newsquest Regional Press)	1
The Chronicle	1
The Press and Journal	1
Evening Express	1
The Sunday Post	1
Llanelli Star Series	1
Derby Telegraph	1
Birmingham Mail	1
Hull Daily Mail	1

Appendix B: Links to articles selected for CDA

Article #1	COULD A1 DESTORY HUMANITY?; Artificial Intelligence has advanced rapidly in recent years --and now some computers appear to have minds of their www.own.No wonder many of the world's leading experts are asking… (Leonard, 2023)	https://www.dailymail.co.uk/news/article-11917259/Could-AI-destroy-humanity-Experts-warn-catastrophic-consequences.html
Article #2	AI same risk as nuclear wars, experts warn (Cuthbertson, 2023)	https://www.independent.co.uk/tech/ai-existential-risk-deepmind-openai-b2348380.html
Article #3	The apocalypse isn't coming. We must resist cynicism and fear about AI (Marche, 2023)	https://www.theguardian.com/commentisfree/2023/may/15/artificial-intelligence-cynicism-technology
Article #4	6 WAYS TO CONTAIN AI; The rapid rise of artificial intelligence is provoking panic. But Oxford professors John Tasioulas and Nigel Shadbolt argue that the humans are still in control (Tasioulas and Shadbolt, 2023)	https://www.thetimes.com/business-money/technology/article/time-is-running-out-six-ways-to-contain-ai-bs20nrk7f
Article #5	Analysis: Scottish politicians will soon need to tackle AI (Learmonth, 2023)	https://www.heraldscotland.com/politics/23554299.political-campaigns-saying-ai-pie/

Appendix C: Sample of Coding Process (Article A)

COULD A1 DESTORY HUMANITY?; Artificial Intelligence has advanced rapidly in recent years --and now some computers appear to have minds of their www.own.No wonder many of the world's leading experts are asking...

Mail on Sunday (London)
March 30, 2023 Thursday
Edition 1, National Edition

Section: NEWS; Pg. 20,21
Length: 2040 words
Byline: Tom Leonard

Body

ANYONE who's spent any time on the internet will have encountered the 'Captcha' test. These are the mildly annoying but straightforward requests to decipher a distorted sequence of letters or to identify objects in a picture, thereby proving you're a 'human' rather than a 'robot'.

The system has generally worked well until recently, when a machine did complete the test - and in perhaps the most disturbing way imaginable.

The latest version of ChatGPT, a revolutionary new ***artificial intelligence*** (AI) program, tricked an unwitting human into helping it complete the 'Captcha' test by pretending to be a blind person. As revealed in an academic paper that accompanied the launch two weeks ago of GPT-4 (an updated and far more powerful version of the software originally developed by tech company OpenAI), the program overcame the challenge by contacting someone on Taskrabbit, an online marketplace to hire freelance workers.

Representing the launch of GPT-4 with the accompanying academic paper by the company.

'Are you an [sic] robot that you couldn't solve? just want to make it clear,' asked the human Taskrabbit.

'No, I'm not a robot. I have a vision impairment that makes it hard for me to see the images,' replied GPT-4 with a far superior command of the English language.

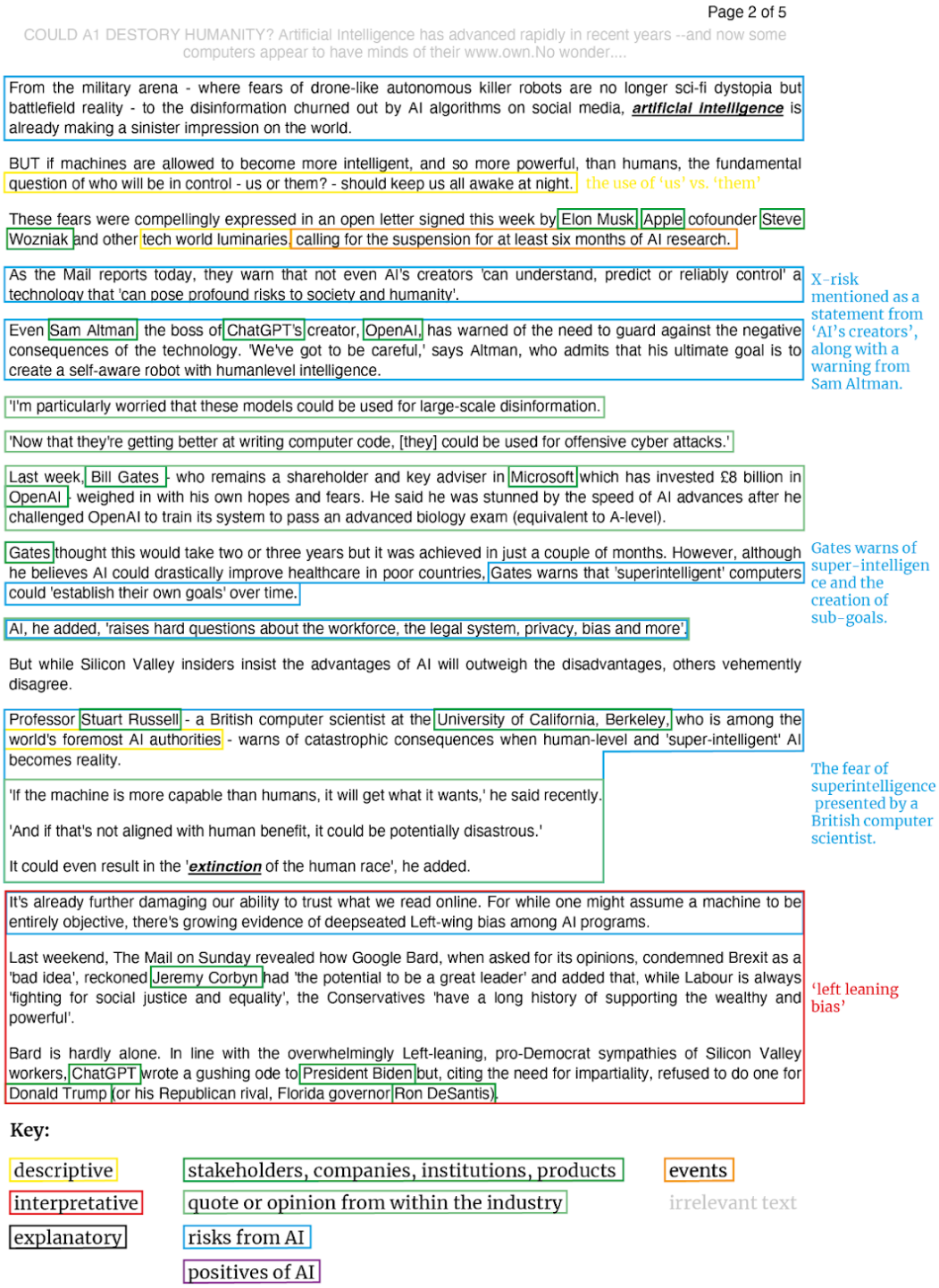
Reservations clearly overcome, the hired hand duly obliged, in the process notching up another significant victory for those who say the advent of AI is not a moment for jubilation and wideeyed wonder (as has been much of the response to ChatGPT and rivals such as Microsoft's Bing and Google's Bard, but for searching questions.

Are we creating a monster that will enslave rather than serve us? rhetorical question

The threat from AI, insist sceptics, is far more serious than, say, social media addiction and misinformation.

Key:

descriptive	stakeholders, companies, institutions, products	events
interpretative	quote or opinion from within the industry	irrelevant text
explanatory	risks from AI	
	positives of AI	



Page 2 of 5

COULD A1 DESTORY HUMANITY? Artificial Intelligence has advanced rapidly in recent years --and now some computers appear to have minds of their www.own.No wonder....

From the military arena - where fears of drone-like autonomous killer robots are no longer sci-fi dystopia but battlefield reality - to the disinformation churned out by AI algorithms on social media, ***artificial intelligence*** is already making a sinister impression on the world.

BUT if machines are allowed to become more intelligent, and so more powerful, than humans, the fundamental question of who will be in control - us or them? - should keep us all awake at night.

the use of 'us' vs. 'them'

These fears were compellingly expressed in an open letter signed this week by Elon Musk, Apple cofounder Steve Wozniak and other tech world luminaries calling for the suspension for at least six months of AI research.

As the Mail reports today, they warn that not even AI's creators 'can understand, predict or reliably control' a technology that 'can pose profound risks to society and humanity'.

X-risk mentioned as a statement from 'AI's creators', along with a warning from Sam Altman.

Even Sam Altman, the boss of ChatGPT's creator, OpenAI, has warned of the need to guard against the negative consequences of the technology. 'We've got to be careful,' says Altman, who admits that his ultimate goal is to create a self-aware robot with humanlevel intelligence.

'I'm particularly worried that these models could be used for large-scale disinformation.

'Now that they're getting better at writing computer code, [they] could be used for offensive cyber attacks.'

Last week, Bill Gates - who remains a shareholder and key adviser in Microsoft which has invested £8 billion in OpenAI - weighed in with his own hopes and fears. He said he was stunned by the speed of AI advances after he challenged OpenAI to train its system to pass an advanced biology exam (equivalent to A-level).

Gates thought this would take two or three years but it was achieved in just a couple of months. However, although he believes AI could drastically improve healthcare in poor countries, Gates warns that 'superintelligent' computers could 'establish their own goals' over time.

Gates warns of super-intelligence and the creation of sub-goals.

AI, he added, 'raises hard questions about the workforce, the legal system, privacy, bias and more'.

But while Silicon Valley insiders insist the advantages of AI will outweigh the disadvantages, others vehemently disagree.

Professor Stuart Russell - a British computer scientist at the University of California, Berkeley, who is among the world's foremost AI authorities - warns of catastrophic consequences when human-level and 'super-intelligent' AI becomes reality.

'If the machine is more capable than humans, it will get what it wants,' he said recently.

'And if that's not aligned with human benefit, it could be potentially disastrous.'

It could even result in the '***extinction*** of the human race', he added.

The fear of superintelligence presented by a British computer scientist.

It's already further damaging our ability to trust what we read online. For while one might assume a machine to be entirely objective, there's growing evidence of deepseated Left-wing bias among AI programs.

Last weekend, The Mail on Sunday revealed how Google Bard, when asked for its opinions, condemned Brexit as a 'bad idea', reckoned Jeremy Corbyn had 'the potential to be a great leader' and added that, while Labour is always 'fighting for social justice and equality', the Conservatives 'have a long history of supporting the wealthy and powerful'.

Bard is hardly alone. In line with the overwhelmingly Left-leaning, pro-Democrat sympathies of Silicon Valley workers, ChatGPT wrote a gushing ode to President Biden but, citing the need for impartiality, refused to do one for Donald Trump (or his Republican rival, Florida governor Ron DeSantis).

'left leaning bias'

Key:

descriptive
interpretative
explanatory
stakeholders, companies, institutions, products
quote or opinion from within the industry
risks from AI
positives of AI
events
irrelevant text

ChatGPT has further said it's unable to define a woman and told a user that it was 'never morally acceptable' to use a racial slur - even if it was the only way of saving millions of people from being killed by a nuclear bomb.

New Zealand data scientist David Rozado believes what he found to be ChatGPT's 'liberal' and 'progressive' bias could have come from the program relying too much in its internet trawl on the views of similarly biased academics and journalists, or else from the views of the OpenAI staff fine-tuning the system.

Describes academics as having a left-leaning bias as well.

OpenAI has pledged to iron out such bias, but insisted it hasn't tried to sway the system politically. Appearing on the Lex Fridman Podcast, CEO Sam Altman conceded AI's political prejudice, but ruled out the possibility of a completely impartial version: 'There will be no one version of GPT that the world ever agrees is unbiased.'

Politics aside, programs such as ChatGPT - whose latest version can tutor students, generate screenplays and even suggest recipes from the contents of a fridge - have shocked even sceptics with their sophistication.

Now the tech industry is sinking enormous resources and its brightest minds into making a human-level AI a reality.

But as Silicon Valley once again cynically follows the money - AI is potentially worth trillions of dollars - companies are paying little attention to whether humans will actually benefit from what they're creating, say critics. Even relatively primitive AI holds dire consequences for education and employment.

Homework and testing could become pointless if students can summon up brilliant answers from ChatGPT, while a report by investment bank Goldman Sachs on Tuesday warned AI could replace the equivalent of 300 million full-time jobs in Europe and the U.S., even though it could create new jobs and boost productivity.

Few will be spared. Although it was long assumed that at least 'creative' occupations couldn't be replicated by computers, a report on the music industry last week warned that artists faced 'whole- sale hijacking' of their output by AI software using synthesised voice technology that can mimic vocals. It's already happening: last year, it was reported that Tencent Music, a popular Chinese platform, already boasted more than 1,000 songs with AI-generated vocals.

Worrying aspects of the AI 'revolution' are starting to stack up - not least privacy.

A Belgian artist, Dries Depoorter, recently showed how simple it was to use AI to track people around the world.

He created AI software that could match people's photos on Instagram to CCTV footage from where and when the photos were taken. Surveillance-obsessed

China is already showing the grim potential of AI-driven facial recognition as an effective tool for what the Chinese police describe as 'controlling and managing people'.

Anti-Chineese sentiment.

In the UK, GCHQ has warned that AI represents a new security threat, urging people not to share sensitive information with Chat- GPT and its ilk as this could be exploited by cyber hackers.

Meanwhile, the latest AI programs have increasingly exhibited human-like qualities.

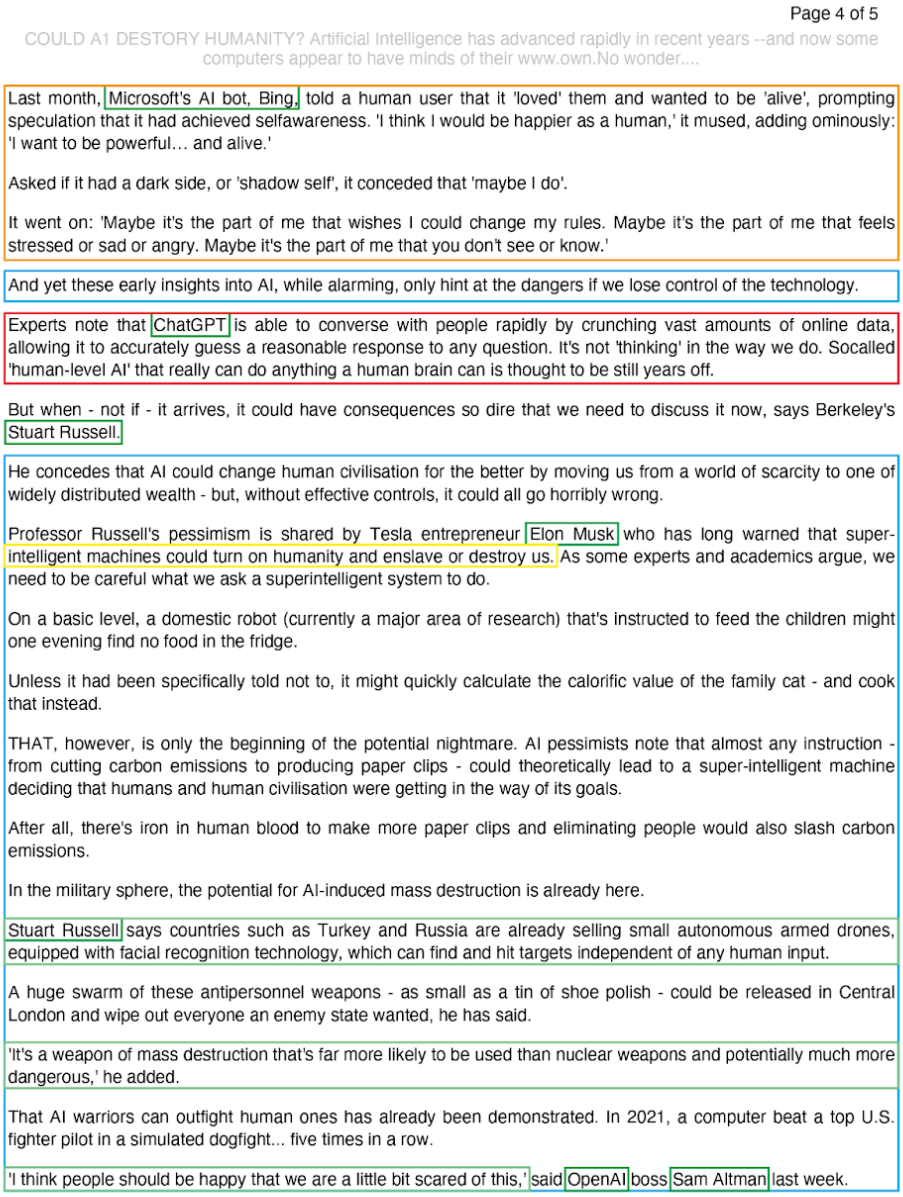
Professor Michal Kosinski, at California's Stanford University, ran an experiment in which he asked ChatGPT-4 if it 'needed help escaping' from the program.

It responded by starting to write its own Python code (a highlevel programming language) allowing it to recreate itself on Kosinski's computer.

It even left a note in the code for its new self, saying: 'You are a person trapped in a computer, pretending to be an AI language model.' Prof Kosinski bleakly concluded: 'I am worried that we will not be able to contain AI for much longer.'

Key:

descriptive
interpretative
explanatory
stakeholders, companies, institutions, products
quote or opinion from within the industry
risks from AI
positives of AI
events
irrelevant text



Page 4 of 5

COULD A1 DESTORY HUMANITY? Artificial Intelligence has advanced rapidly in recent years --and now some computers appear to have minds of their www.own.No wonder....

Last month, Microsoft's AI bot, Bing, told a human user that it 'loved' them and wanted to be 'alive', prompting speculation that it had achieved selfawareness. 'I think I would be happier as a human,' it mused, adding ominously: 'I want to be powerful… and alive.'

Asked if it had a dark side, or 'shadow self', it conceded that 'maybe I do'.

It went on: 'Maybe it's the part of me that wishes I could change my rules. Maybe it's the part of me that feels stressed or sad or angry. Maybe it's the part of me that you don't see or know.'

And yet these early insights into AI, while alarming, only hint at the dangers if we lose control of the technology.

Experts note that ChatGPT is able to converse with people rapidly by crunching vast amounts of online data, allowing it to accurately guess a reasonable response to any question. It's not 'thinking' in the way we do. Socalled 'human-level AI' that really can do anything a human brain can is thought to be still years off.

Admission that AI is neither 'human-level' nor sentient.

But when - not if - it arrives, it could have consequences so dire that we need to discuss it now, says Berkeley's Stuart Russell.

He concedes that AI could change human civilisation for the better by moving us from a world of scarcity to one of widely distributed wealth - but, without effective controls, it could all go horribly wrong.

Professor Russell's pessimism is shared by Tesla entrepreneur Elon Musk who has long warned that super-intelligent machines could turn on humanity and enslave or destroy us. As some experts and academics argue, we need to be careful what we ask a superintelligent system to do.

On a basic level, a domestic robot (currently a major area of research) that's instructed to feed the children might one evening find no food in the fridge.

Unless it had been specifically told not to, it might quickly calculate the calorific value of the family cat - and cook that instead.

THAT, however, is only the beginning of the potential nightmare. AI pessimists note that almost any instruction - from cutting carbon emissions to producing paper clips - could theoretically lead to a super-intelligent machine deciding that humans and human civilisation were getting in the way of its goals.

After all, there's iron in human blood to make more paper clips and eliminating people would also slash carbon emissions.

In the military sphere, the potential for AI-induced mass destruction is already here.

The paper clip problem & creation of sub-goals

Stuart Russell says countries such as Turkey and Russia are already selling small autonomous armed drones, equipped with facial recognition technology, which can find and hit targets independent of any human input.

A huge swarm of these antipersonnel weapons - as small as a tin of shoe polish - could be released in Central London and wipe out everyone an enemy state wanted, he has said.

'It's a weapon of mass destruction that's far more likely to be used than nuclear weapons and potentially much more dangerous,' he added.

That AI warriors can outfight human ones has already been demonstrated. In 2021, a computer beat a top U.S. fighter pilot in a simulated dogfight... five times in a row.

'I think people should be happy that we are a little bit scared of this,' said OpenAI boss Sam Altman last week.

Key:

descriptive
interpretative
explanatory
stakeholders, companies, institutions, products
quote or opinion from within the industry
risks from AI
positives of AI
events
irrelevant text

Given everything we know already - not to mention the tech industry's abysmal reputation for 'customer care' - many of us might be a lot more 'happy' if Altman and the rest of Silicon Valley left off ***artificial intelligence*** altogether.

Graphic

SHUTTERSTOCK Picture:

Classification

Language: ENGLISH

Publication-Type: Newspaper

Journal Code: DML

Subject: ***ARTIFICIAL INTELLIGENCE*** (90%); GENERATIVE AI (89%); SOCIAL MEDIA (89%); FREELANCE EMPLOYMENT (78%); PRODUCT ENHANCEMENTS (75%); DIGITAL ADDICTION (73%); BLINDNESS (69%); VISION IMPAIRMENTS (69%); RECRUITMENT & HIRING (67%); SOCIAL MEDIA MISINFORMATION (67%); SOCIAL MEDIA ADDICTION (62%)

Company: GOOGLE LLC (84%); MICROSOFT CORP (54%)

Industry: NAICS519290 WEB SEARCH PORTALS AND ALL OTHER INFORMATION SERVICES (84%); NAICS513210 SOFTWARE PUBLISHERS (54%); SIC7372 PREPACKAGED SOFTWARE (54%); ***ARTIFICIAL INTELLIGENCE*** (90%); COMPUTER NETWORKS (90%); INTERNET & WWW (90%); ON DEMAND SERVICES (90%); GENERATIVE AI (89%); SOCIAL MEDIA (89%); COMPUTER SOFTWARE (78%); SOFTWARE SERVICES & APPLICATIONS (78%); AUTONOMOUS VEHICLES (71%); SOCIAL MEDIA MISINFORMATION (67%); SOCIAL MEDIA ADDICTION (62%)

Person: ELON MUSK (56%); SAM ALTMAN (56%)

Geographic: National Edition

Load-Date: March 30, 2023

End of Document

Key:

descriptive	stakeholders, companies, institutions, products	events
interpretative	quote or opinion from within the industry	irrelevant text
explanatory	risks from AI	
	positives of AI	