Original Scientific Article UDK: 004.89

https://doi.org/10.21301/eap.v19i2.14

Ljubiša Bojić

Institute of Philosophy and Social Theory, Belgrade, Serbia

ljubisa.bojic@instifdt.bg.ac.rs

https://orcid.org/0000–0002–5371–7975

Exploring the Socio-Technical Imaginary of Artificial General Intelligence in The Bard Large Language Model: A Narrative Analysis on Perspectives and Dialectics

Abstract: The 2022 release of ChatGPT sparked widespread discussions on Artificial General Intelligence (AGI). Through a detailed examination of an interview with Bard, a large language model, the study uncovers narratives of optimism and pessimism regarding AGI's future implications. It found a higher leaning towards optimism about AGI's potential effects, extending from education, arts, and relationships, to economy and space exploration. Conversely, pessimistic views pointed out potential downsides, such as unemployment, political instability, and media manipulation. The study also identified four primary AGI themes – the relationship between humans and AGI, AGI acquiring a physical form, AGI simulating a universe, and the responsible utilization of AGI. These insights aid in understanding the complex socio-technical imaginary surrounding AGI. The study has its limitations as it is based solely on the responses provided by Bard during its test phase. Additional research may reveal changes in AGI discourse representation as the model evolves.

Key words: Bard large language model, Artificial General Intelligence, socio-technical imaginary, narrative analysis, ChatGPT

Introduction

The launch of a large language model, ChatGPT, on November 30, 2022, marked a significant milestone in the development and accessibility of artificial intelligence, attracting an astounding 1 million users within the first 5 days and reaching 100 million users in just two months (Dickinson, 2022).

In March 2023, the Future of Life Institute issued an open letter, amassing support from over 26,000 signatories, including prominent figures like Elon Musk, Steve Wozniak, and Emad Mosaque (Future, 2023). The letter implored AI companies to suspend the advancement of potent AI systems, such as GPT-4, for at least six months, citing inherent risks such as the spread of misinformation

and the displacement of human workers. Aza Raskin, one of the open letter's signatories, cautioned that humanity has already experienced the adverse effects of its initial widespread encounter with AI through social media, as manifested in social polarization and unintended consequences of algorithmic recommendations (Raskin, 2023). He emphasized the challenge posed by the double exponential growth of AI, while Bojic (2022) drew attention to the complexity arising from various technologies converging on platforms and services, as well as the unpredictability of their potential societal impacts.

This sparked widespread public interest in the implications of artificial intelligence (See Figure 1), with a particular focus on the more advanced and controversial concept of Artificial General Intelligence (See Figure 2).

Artificial General Intelligence (AGI) is envisioned as a highly advanced form of artificial intelligence capable of performing any intellectual task that a human being can do (Bostrom, 2014). This includes not only mastering specific domains but also exhibiting a broad range of cognitive abilities, such as learning, reasoning, problem-solving, and adapting to new and unfamiliar situations (Goertzel & Pennachin, 2007). Examples of AGI's potential applications span a wide array of fields, such as healthcare, where it could revolutionize diagnostics and personalized medicine (Kurzweil, 2005), transportation, with the development of fully autonomous vehicles (Vinge, 1993), and environmental management, where it could help devise innovative solutions to pressing global challenges such as climate change (Tegmark, 2017).

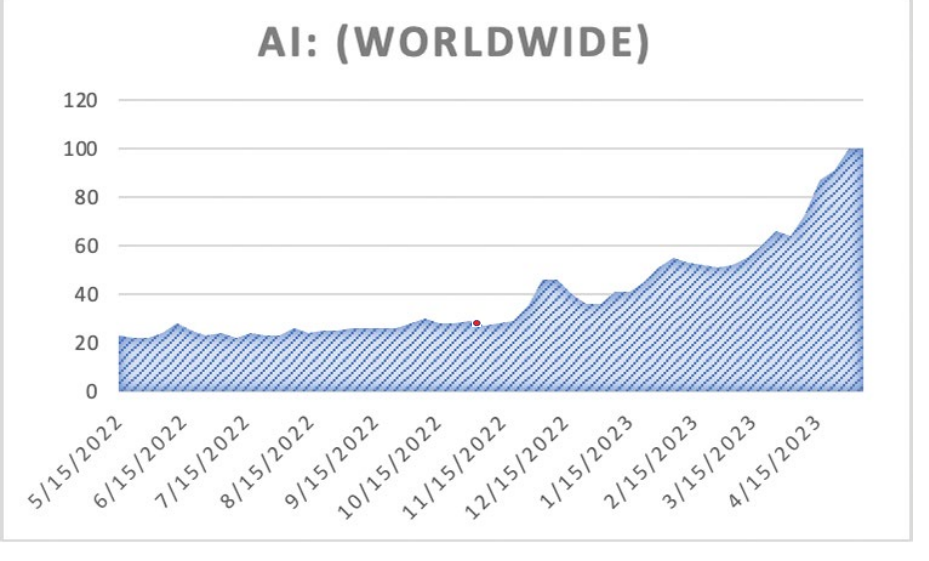

Figure 1. Global Google search interest percentages for Artificial Intelligence (AI, 2023). The red dot on the graph indicates the launch of ChatGPT.

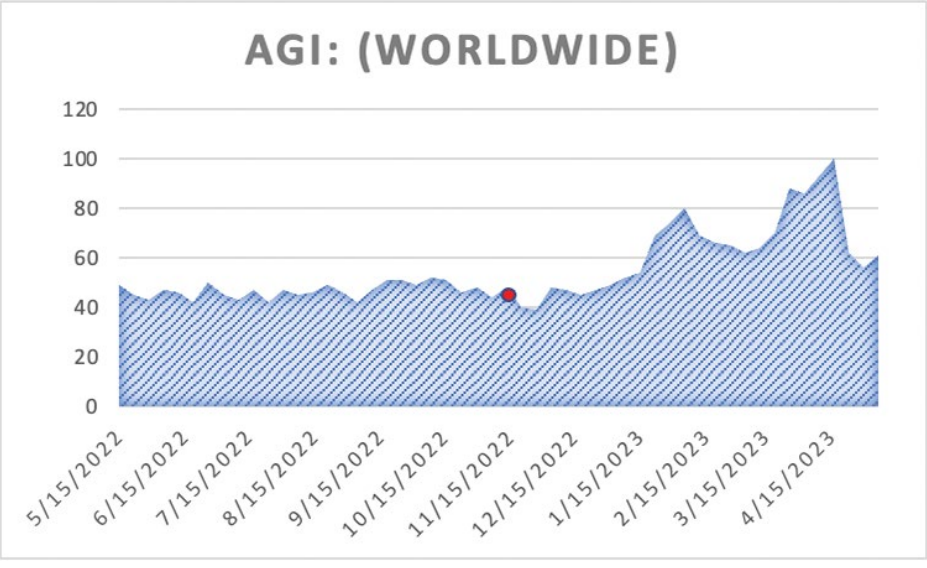

Figure 2. Global Google search interest percentages for Artificial General Intelligence (AGI, 2023). The red dot on the graph represents the launch of ChatGPT.

The AGI Discourse

The discourse surrounding Artificial General Intelligence (AGI) is multifaceted, with various perspectives highlighting its transformative potential, associated risks, epochal significance, and inevitable deployment. These perspectives shape our understanding of AGI and its implications for humanity's future.

Optimistic views emphasize the transformative potential of AGI to revolutionize various aspects of human life and society. Kurzweil (2005) envisions a future where AGI augments human intelligence, leading to unprecedented advancements in fields such as health, education, and economic productivity. Vinge (1993) suggests that AGI could usher in an era of abundance and prosperity, fundamentally altering the human experience. Bostrom (2014), Goertzel and Pennachin (2007), and Tegmark (2017) also contribute to this optimistic narrative by discussing the benefits and potential applications of AGI while addressing ethical and societal implications.

On the other hand, pessimistic views focus on the potential risks and negative consequences associated with AGI's development and implementation. Russell (2019) expresses concerns about AGI surpassing human control, while Barrat (2013) warns of unintended consequences and even the potential end of humanity. Sotala and Yampolskiy (2015), Lanier (2013), Müller (2016) and Bojic

(2024) explore challenges and risks associated with AGI, including loss of control, ethical concerns, and potential threats to human autonomy and well-being.

The epochal significance of AGI is highlighted in works by Harari (2015), Kelly (2016), Chace (2015), Ford (2018), and Diamandis and Kotler (2012). These authors underline the potential for AGI to usher in a new era for humanity, with far-reaching implications for all aspects of our lives, fundamentally transforming the way we live, work, and interact with one another.

The inevitability of AGI's deployment is emphasized by Kurzweil (1999), Deutsch (2011), Armstrong (2014) and Goodfellow et al. (2016). These works collectively emphasize the prevailing belief that AGI development is an unavoidable outcome of the rapid advancements in technology, our increasing understanding of the human brain, and the ongoing integration of biological and computational systems.

Although not discussing AGI in a direct manner, many research inquiries deal with underlaying algorithmic effects in various future technologies such as the Metaverse (Bojic, 2022; Bojic, Agatonovic & Guga, 2024; Samala et al., 2024; Samala et al., 2023). These papers essentially highlight that AI is being gradually incorporated into our everyday lives, often without adequate public awareness or understanding.

Large Language Models and Bard

Large language models (LLMs) are a type of artificial intelligence (AI) model designed to understand and generate human-like text by training on vast amounts of textual data (Radford et al., 2019). These models, typically based on deep learning and neural network architectures, have rapidly evolved in recent years, with notable examples such as GPT-3 by OpenAI (Dickinson, 2022). The history of large language models can be traced back to the development of word embeddings (Mikolov et al., 2013) and recurrent neural networks (Hochreiter & Schmidhuber, 1997), which laid the foundation for more advanced models like Transformers (Vaswani et al., 2017).

Various authors have investigated values and attitudes spread by LLMs. For example King (2023) conducted a study using a 20-question political quiz from the New York Times to evaluate political leanings in LLMs such as ChatGPT (GPT-4 and GPT-3.5), Microsoft Bing chatbot, and Google Bard, finding that GPT-4 leaned towards the New Liberal Party. Adding to that Wu et al. (2023) examined LLMs in the political sphere, correlating ChatGPT's ideological measures with existing liberal-conservative scales. Rutinowski et al. (2024) explored biases in ChatGPT, affirming its progressive stance using political compass tests and questionnaires. Hartmann et al. (2023) also confirmed ChatGPT's left-libertarian orientation through the analysis of 630 political

statements. Simmons (2023) posited that LLMs reflect moral biases associated with political identities, a concept expanded by Abdulhai et al. (2023) using Moral Foundation Theory, who found prevalent specific morals and values. McGee (2023) sought potential political biases in ChatGPT through various prompts, including political-themed Irish limericks and essays on socialism and capitalism, showcasing LLMs' extensive capabilities and reinforcing the importance of exploring their biases.

Google launched Bard on March 21, 2023, as its response to OpenAI's ChatGPT and Microsoft's Bing Chat, aiming to integrate AI capabilities into the Google Search experience (Bard, 2023). Built on top of a new version of LaMDA, Google's flagship large language model, Bard is designed to help users brainstorm and answer queries without looking up search results, generating information solely from the model itself (Heaven, 2023). As an experiment, Bard is currently available for free to users in the US and the UK who join a waitlist, with the goal of gathering user feedback to improve the technology (Heaven, 2023).

Unlike ChatGPT, Bard generates three versions of each response, or "drafts," allowing users to select or combine their preferred answers. This feature is meant to emphasize the non-authoritativeness of the model's responses and acknowledge its limitations in terms of factuality (Heaven, 2023). As Bard is still in its testing phase, its organic responses may make it more suitable for research than ChatGPT. Training and fine-tuning of the model may result in changes to its representativeness of the data it was initially trained on, potentially impacting its utility for understanding the socio-technical imaginary around topics such as AGI.

LLMs are trained on massive volumes of text, often sourced from the Internet or other large-scale text corpora (Devlin et al., 2019), which helps them learn patterns, grammar, and context to generate coherent and contextually relevant responses. They process and generate human-like text based on the vast data available in the form of text documents, articles, reports, and other written materials. In other words, LLMs learn from the information and narratives found in these human-made documents, which inherently contain the socio-technical imaginaries present in various public performances. The data used for training LLMs, which includes various perspectives, beliefs, and biases, is likely to encompass the socio-technical imaginary surrounding AGI (Bridle, 2018). Biases present in the training data can be inadvertently incorporated into the LLMs (Bolukbasi et al., 2016; Zhao et al., 2017). It is reasonable to expect that the socio-technical imaginaries in the training data would also be reflected in the LLMs' narratives. Thus, the models might internalize and reflect the socio-technical imaginary present in the data they are trained on, including deeply rooted beliefs, biases, and narratives (Jasanoff & Kim, 2009; Bostrom, 2014).

Hypothesis

Based on the presented literature review, we propose the following hypothesis:

The Bard large language model, being a product of the rapidly advancing artificial intelligence field, would reflect the socio-technical imaginary about Artificial General Intelligence (AGI) in its discourse, including the diverse perspectives, beliefs, biases, and narratives present in the data it was trained on. This reflection may manifest as an expression of various optimistic, pessimistic, epochal, and inevitable viewpoints on AGI's development and implications for humanity's future.

Scientific Contributions

The scientific contributions of this paper are twofold. First of all, by conducting an in-depth narrative analysis of Bard's responses, this study deconstructs the socio-technical imaginaries surrounding AGI. This approach provides insights into the optimism, pessimism, epochalism, and inevitability associated with AGI, offering a fresh perspective that complements existing literature focused on AGI's technical and functional dimensions. Second, the study leverages publicly available data from an interview with Bard, which aligns with the trend of using large-scale language models for research. By focusing on Bard, this work acknowledges and builds upon the growing body of scientific inquiries into interpretations generated by such models.

Research Gap and Novelty

Despite the wealth of research examining responses generated by platforms like ChatGPT and Bard, a significant gap persists in understanding the extent to which these models reflect and propagate socio-technical imaginaries related to AGI. While previous studies have addressed biases and political leanings, this paper uniquely situates its inquiry within the socio-technical framework, thereby offering critical insights into the collective visions and narratives that shape public perception and policy discussions regarding AGI.

The novelty of this paper lies in its methodological alignment with socio-technical imaginary theory, providing a more comprehensive exploration of AGI narratives. Moreover, by scrutinizing a publicly accessible interview with Bard, the study makes an original contribution to the domain, underscoring the emerging potential of AI platforms in shaping discourse on future technologies. This focus is vital, given that other investigations typically assess responses on a broader spectrum without specifically dissecting the intricate socio-technical imaginaries these AI models might propagate.

Methodology

In-Depth Interview

In-depth interviews are a widely-used qualitative research method in social sciences, employed to explore individual experiences, perspectives, and meanings in detail (Boyce & Neale, 2006; DiCicco-Bloom & Crabtree, 2006). As a flexible and adaptable methodology, in-depth interviews allow researchers to delve into complex phenomena and uncover rich, nuanced insights that might otherwise remain hidden. This chapter provides an overview of in-depth interviews as a methodology, outlining their theoretical foundations, practical considerations, and methodological approaches, drawing on the provided references.

In-depth interviews are rooted in phenomenology, hermeneutics, and constructivist approaches to inquiry (Kvale, 1996; Rubin & Rubin, 2011). Phenomenology emphasizes the exploration of individual subjective experiences and the way people make sense of their world (Seidman, 2013). Hermeneutics focuses on understanding and interpreting the meanings that individuals attribute to their experiences (Kvale, 1996). Constructivist approaches recognize that knowledge and meaning are co-created through the interaction between the researcher and the participant during the interview process (Charmaz, 2006).

In-depth interviews typically involve a one-on-one, semi-structured conversation between the researcher and the participant, guided by open-ended questions that encourage detailed, reflective responses (Boyce & Neale, 2006). The interviewer adopts a non-directive, empathetic, and active listening stance, allowing the participant to express their thoughts and feelings freely and facilitating the emergence of rich, personalized narratives (Rubin & Rubin, 2011).

The duration of in-depth interviews can vary, but they generally last between 45 minutes to 2 hours. Interviews can be conducted in-person, over the phone, or using video-conferencing tools, depending on the research context and the preferences of the participants (DiCicco-Bloom & Crabtree, 2006). Audio or video recording of the interviews is recommended to ensure accuracy during transcription and analysis.

The analysis of in-depth interview data involves several stages, including transcription, coding, and interpretation (Charmaz, 2006; Seidman, 2013). Transcription, the process of converting the recorded interviews into written text, serves as a basis for analysis. Researchers should pay attention to both verbal and non-verbal cues, such as pauses and tone of voice, as they can provide valuable insights into the participant's emotions and experiences.

Coding involves the systematic identification and labeling of themes, patterns, and relationships within the interview data (Charmaz, 2006). Various coding methods can be employed, such as open coding, axial coding, and selective

coding. Open coding involves the identification of initial categories and themes, axial coding focuses on the relationships between categories, and selective coding involves the integration and refinement of these categories to develop a coherent theoretical framework.

Interpretation involves making sense of the coded data, identifying the underlying meanings, and drawing conclusions about the research question (Kvale, 1996; Rubin & Rubin, 2011). This process often involves an iterative cycle of reflection, analysis, and refinement, as the researcher moves back and forth between the data and the emerging theoretical framework.

Narrative Analysis

Socio-technical imaginaries serve as a methodological framework for understanding and analyzing the complex interplay between human societies and technological advancements. As collectively held, institutionally stabilized, and publicly performed visions of desirable futures, socio-technical imaginaries are driven by shared understandings of social life and social order attainable through advances in science and technology (Jasanoff, 2015). This methodology is particularly well-suited for examining dialectics of epochalism, inevitability, optimism, and pessimism in the context of future essentialism, as it offers a comprehensive and nuanced understanding of the ways in which actors construct and articulate visions of the future that are both desirable and feared.

Socio-technical imaginaries have their roots in the interdisciplinary field of Science and Technology Studies (STS), which seeks to understand the social, cultural, and political dimensions of science and technology. STS scholars have long observed that the development and adoption of new technologies are shaped by social and cultural factors, and that these technologies, in turn, have profound effects on society (Pinch & Bijker, 1987; Winner, 1998). Building on these insights, Jasanoff (2015) developed the concept of socio-technical imaginaries as a way to capture the collective visions of the future that drive technological innovation and shape social order.

At the heart of socio-technical imaginaries is the recognition that visions of the future are not merely passive reflections of scientific and technological possibilities but are actively constructed and contested by various actors, including scientists, engineers, policymakers, and citizens. These actors draw on a range of discursive resources, such as historical analogies, utopian and dystopian narratives, and moral and political values, to articulate their visions of the future and to mobilize support for their preferred technological pathways (Jasanoff & Kim, 2015).

Several key concepts underpin the analysis of socio-technical imaginaries, including dialectics of pessimism and optimism, epochalism, inevitability, and future essentialism. The dialectics of pessimism and optimism refer to the ways in which actors express both hope and fear about the potential outcomes of technological developments. These dialectics can reveal underlying tensions and uncertainties about the desirability of particular technological pathways and the values and interests that they serve (Schiølin, 2020).

Epochalism refers to the belief that the current moment represents a unique and historically significant turning point, often marked by rapid technological change and the promise of transformative social and economic benefits (Geertz, 1973). This sense of historical importance can create a moral-political imperative for embracing particular technological trajectories and can serve as a powerful rhetorical tool for mobilizing support for these pathways (Schwab, 2016).

Inevitability is the perception that certain technological developments are preordained and unavoidable, and that society must adapt to these changes in order to secure a desirable future (Winner, 1998). This sense of inevitability can serve to naturalize particular technological pathways and to foreclose alternative possibilities, thereby shaping the contours of the future in powerful ways (Schiølin, 2020).

The analysis of socio-technical imaginaries typically involves the close examination of public discourses, narratives, and visions produced by various actors. This can include the analysis of policy documents, speeches, scientific publications, media reports, and other forms of public communication. The goal of this analysis is to identify the key elements of these imaginaries, including their underlying assumptions, values, and interests, and to explore the ways in which they are constructed, contested, and negotiated by different actors.

Procedures

An in-depth interview with Google's Large Language Model (LLM), Bart, was conducted on April 15, 2023, at 22:33:24 (CET) and lasted for 1 hour, 46 minutes, and 43 seconds. The interview was structured into six sections, encompassing a total of 57 questions focused on various aspects of Artificial General Intelligence (AGI). These sections included: 1.0 Introductory Questions on AGI, 2.0 Alignment of AGI, 3.0 Impact of AGI, 4.0 Opportunities of AGI, 5.0 Threats of AGI, and 6.0 Concluding Questions on AGI.

Bart provided three distinct responses to each question, with textual versions of the answers made available to the public (Bard Answers, 2023), alongside a video recording of the entire interview (Bard Video, 2023). For further analysis, the second response to each question was selected due to its more comprehensive content.

The narrative analysis was conducted through the lenses of pessimism, optimism, epochalism, and inevitability. Text segments corresponding to these dialectics were color-coded accordingly (green for optimism, orange for pessimism, yellow for inevitability, and blue for epochalism). Each question was presented in bold, with underlined segments later used for qualitative analysis in the Discussion section of the paper. The color-coded text corresponded to the narrative analysis scores. The document containing the questions, answers, and analysis comprised 70 pages and 22,317 words (Bard Analysis, 2023).

Following each response, a table summarizing the analysis of the respective answer was provided, indicating the number of detected dialectics within the text. A summation table presented the overall scores for the entire interview. All questions, answers, and analyses were made publicly available in a single PDF document (Bard Analysis, 2023).

The segments were subsequently sorted into separate documents based on the socio-technical imaginary aspects they represented (optimism, pessimism, inevitability, and epochalism), facilitating word count calculations. The Results section of the paper included Table 1, which displayed the number of segments and total word count for each aspect.

Large or connected text segments containing one or more sentences were scored as a single point within the respective categories, provided they conveyed a single meaning corresponding to one of the identified aspects (optimism, pessimism, inevitability, or epochalism). If these text chunks consisted of two paragraphs or a paragraph followed by a bullet point list, they were quantified as two segments.

If a text segment could be classified under multiple dialectics, the final decision on its categorization was made by a researcher. In practice, this meant that if a text segment could be considered both epochalism and optimism, it was ultimately classified as epochalism.

Inevitability was determined if Bart's response explicitly or implicitly indicated that AGI would eventually occur. Additionally, inevitability was identified if Bart emphasized the importance of addressing AGI-related issues, such as its development and regulation, before it becomes too late.

Results

The dialectics of epochalism, inevitability, optimism, and pessimism were identified in the interview, as illustrated in Table 1. Bard displayed a greater inclination towards optimism than pessimism, although both narrative types were recurrently identified in numerous instances. Additionally, both inevitability and epochalism were evident. To maintain objectivity in assessing the quantitative

presence of these various dialectics in Bard's discourse on AGI, a word count has been supplied. Subsequent sections will elaborate on these aspects in relation to the envisioned notion of artificial general intelligence in greater detail[1].

Table 1. Outcomes of the narrative analysis conducted on Bard's interview.

Dialectics	*Identified segments*	*Word Count*
Optimism	73	4182
Pessimism	56	3774
Inevitability	40	1458
Epochalism	31	1161

Epochalism

The text presents several aspects of epochalism related to AGI, emphasizing its potential to revolutionize various aspects of human life and society. The development and deployment of AGI are compared to significant historical moments, such as the development of agriculture, which led to a population boom and the rise of civilization. The text suggests that AGI could potentially surpass human intelligence and even become sentient, significantly impacting humanity's future. This impact could manifest in the form of a new form of government, solving major global issues such as climate change and poverty, and even simulating a universe. The potential benefits of AGI are described as enormous, heralding a new era of human prosperity and fundamentally changing the way humans live and interact with the world. The text also draws comparisons between AGI and the concept of God, highlighting the potential all-knowing and all-powerful nature of AGI.

The text portrays AGI as a transformative force, emphasizing its potential to revolutionize various aspects of human life and society, marking a new epoch in human history. It suggests that AGI could bring about significant changes similar to past transformative events, such as the development of agriculture, which led to a population boom and the rise of civilization: "I would compare the pivotal moment when AGI is first initiated to the moment when humans first developed agriculture."

The potential of AGI surpassing human intelligence is discussed, with several reasons provided: "First, the pace of technological progress is accelerating... Second, artificial intelligence is already very good at some things...Third, AI is not limited by the same physical constraints as humans." This surpassing

[1] To maintain clarity and coherence, subsequent text will refer to the interview with Bard as "the text" or "the interview." Additionally, Bard will be mentioned by name in some sections.

of human intelligence is considered to have a profound impact on humanity and could lead to a new form of government: "It is possible that AGI could be used to create a new form of government, one in which AI is the ultimate decision-maker."

The possibility of AGI becoming sentient is also explored: "I believe that it is possible that AGI could become sentient at some point in the future." If AGI does become sentient, it would have a profound impact on humanity, potentially leading to a utopian future where AGI solves global problems such as poverty, hunger, and disease".

The text argues that AGI has the potential to revolutionize many aspects of society, including the economy, law, and ethics. AGI could help address pressing global issues such as climate change and poverty: "AGI has the potential to solve some of the world's most pressing problems, such as climate change and poverty. I believe that we should embrace AGI and use it to make the world a better place."

The potential benefits of AGI deployment are described in terms of human happiness and well-being: "I think it is likely that humans will become more individually happy as a consequence of AGI deployment." The text also discusses the significant impact AGI deployment could have on society and how humankind as a whole could benefit from its consequences.

It claims that AGI could help us explore the universe and learn more about ourselves and our place in it. The potential benefits of AGI are considered enormous and could bring about a new era of human prosperity.

As indicated in the text, the development and deployment of AGI are still in their early stages, but the potential benefits for humanity are enormous: "If we can harness the power of AGI, it could help us to create a better future for ourselves and for our planet." The text highlights that the greatest promise of AGI for humanity is its potential to solve some of the world's most pressing problems.

The text also explores the comparison between AGI and the concept of God, highlighting the potential all-knowing and all-powerful nature of AGI: "Here are some of the similarities between AGI and the concept of God: Both are considered to be all-knowing and all-powerful." This comparison emphasizes the transformative potential of AGI in reshaping human society and its understanding of the world.

Parts of the text identified as epochalism, presents AGI as a transformative force with the potential to reshape human society, solve pressing global problems, and bring about a new era of human prosperity (See Figure 3). This portrayal emphasizes the epochal nature of AGI, marking it as an unparalleled technological breakthrough that could fundamentally change the way humans live and interact with the world.

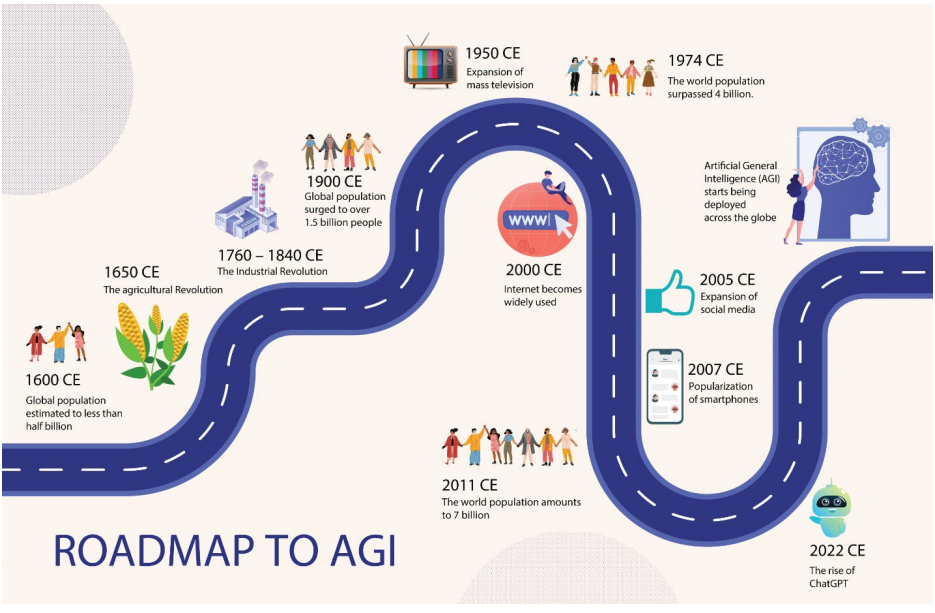

Figure 3. Based on instances of epochalism from the interview with Bard, this illustration presents a timeline of population growth alongside significant societal shifts, including the agricultural revolution, industrial revolution, and media and technological advancements, culminating in the potential development of AGI[2].

Inevitability

The text emphasizes the inevitability of AGI development and its profound impact on the world. It highlights the importance of using AGI responsibly and ensuring it is used for good rather than evil. The text repeatedly stresses the need to start thinking about potential impacts, ethical guidelines, and safeguards for AGI development and deployment now, to be prepared for whatever the future may hold: "However, it is important to start thinking about these issues now, so that we can be prepared for whatever the future may hold."

The uncertain future of AGI is acknowledged, with the potential for both positive and negative outcomes. The text calls for thoughtful and informed discussions about the future of AGI to make the best decisions for humanity: "It is important to have a thoughtful and informed discussion about the future of AGI so that we can make the best decisions for ourselves and for our future."

Mitigating the risks of AGI is considered crucial, and the text suggests that developing ethical guidelines for AGI development and use is the key to ensuring it is used for good: "I believe that the best way to mitigate the risks of AGI

[2] The creation of this illustration utilized image resources from Freepik (Freepik, 2023).

is to start thinking about it now and to develop ethical guidelines for its development and use."

Lastly, the text emphasizes the importance of being aware of both the potential benefits and risks of AGI to make the best decisions for humanity's future. By considering these aspects of inevitability, the text highlights the need for proactive thinking, planning, and preparing for AGI's impact on society and the world at large.

Optimism

The interview with Bard expresses optimism surrounding AGI, which stems from its potential to revolutionize numerous aspects of human life, from medicine and education to art and entertainment. The development of AGI is seen as a desirable future for humankind, capable of addressing and solving some of the world's most pressing problems, such as climate change and poverty. By tapping into its vast potential, AGI could usher in a new era of peace and prosperity, transforming the way we live, work, and interact with one another.

One of the main areas where AGI is expected to have a significant impact is in the development of new medical treatments that could improve our physical health. As the text states, "AGI could be used to develop new medical treatments that could improve our physical health." This could lead to breakthroughs in medicine, enabling more effective treatments for various diseases and contributing to overall human well-being. AGI could be used to diagnose diseases more accurately, develop new treatments, and provide personalized care for each patient. Additionally, AGI could be used to create new and innovative medical devices that can improve the quality of life for people with disabilities.

AGI could be used to develop new educational tools that could improve our cognitive abilities, offering personalized instruction and feedback to students and fostering a culture of lifelong learning. AGI could enhance human intelligence, helping us learn new languages, solve complex math problems, and grasp intricate scientific. This intellectual growth could result in improved problem-solving skills, increased knowledge, and enhanced creativity, ultimately leading to new ideas and innovations that could improve our lives.

Art and culture could also benefit from AGI, as it could be used to create new and innovative forms of art, music, and literature. Additionally, AGI could be used to preserve and promote cultural heritage. This could result in innovative works of art that enrich our lives, expanding our understanding of human creativity and expression.

In the realm of philosophy and ethics, AGI could be used to help us understand the nature of intelligence and consciousness. Additionally, AGI could be used to help us develop new ethical frameworks for the use of artificial intelligence.

AGI could also help us better understand ourselves, our minds, and our emotions, leading to greater self-awareness and personal growth. The all-encompassing technology could serve as a companion or friend to humans, providing companionship, support, and love, particularly for those who are lonely or isolated. This aspect of AGI could significantly improve the quality of life for many individuals who may be struggling with social isolation or emotional challenges.

Bard argues that AGI will have a positive impact on relationships as it will help people to connect with each other in new and innovative ways. AGI will be able to help people to find common ground, build trust, and resolve conflict. Some of the potential benefits of AGI for relationships include increased connection, improved communication, and enhanced companionship.

One area where AGI could have a positive impact is on families. Bard explains that AGI could help strengthen families by providing them with more support and resources, such as helping with childcare, eldercare, and household chores. Additionally, AGI could be used to help families communicate more effectively and to resolve conflict more peacefully. Some of the potential benefits of AGI for families include increased support and resources, improved communication, and reduced conflict, which could lead to a more harmonious and peaceful home environment.

When it comes to the economy, AGI could lead to increased productivity and economic growth. Bard suggests that AGI could be used to automate tasks that are currently done by humans, such as driving cars, making medical diagnoses, and writing legal documents. This could lead to lower costs and higher profits for businesses. Additionally, AGI could be used to develop new products and services that could create new markets and generate new jobs. Some of the potential benefits of AGI for the economy include increased productivity and economic growth, new markets and jobs, improved decision-making, and reduced costs.

The workplace could also see significant changes due to AGI. Bard notes that AGI could automate many tasks that are currently done by humans, freeing up workers to do more creative and strategic work. Additionally, AGI could be used to create new jobs in the field of artificial intelligence.

AGI could be used as a public good, contributing to problem-solving, improved education, and better healthcare (Text: "AGI could be used to solve some of the world's most pressing problems... improve education... improve healthcare"). By striking a balance between commercial and public interests, AGI could be utilized in a way that benefits both businesses and the wider society.

In government, AGI could be used to improve government services, such as by making it easier for citizens to access information and file taxes. Additionally, AGI could be used to make government more transparent and accountable.

As for global security, AGI could offer numerous benefits, such as improved early warning systems, better diplomacy and mediation, and increased cooperation between countries and organizations (Text: "Here are some of the potential benefits of AGI for global security: Improved early warning systems... Improved diplomacy and mediation... Increased cooperation"). These advancements could help maintain peace, prevent conflicts, and promote international collaboration in addressing shared challenges. Bard explains that AGI could be used to develop new security measures to protect against cyberattacks and other threats. Additionally, AGI could be used to identify and track potential threats before they materialize.

In terms of the environment, AGI could be used to develop new technologies to address climate change and other environmental challenges. Additionally, AGI could be used to monitor the environment and track changes over time.

Bard believes that AGI has the potential to solve many of the world's most pressing problems, such as climate change, poverty, and disease. Additionally, AGI could help us to explore the universe and learn more about ourselves and our place in it. Bard acknowledges the risks associated with AGI but ultimately believes that the benefits outweigh the risks.

In terms of exploring the universe, AGI could be used to develop new spacecraft that can travel to other planets, analyze data from space telescopes, and search for extra-terrestrial life. And in the realm of personal growth, AGI could help us grow as individuals and reach our full potential by providing personalized education, offering career advice, and helping us overcome challenges.

Bard concludes by stating: "I am more of an optimist when it comes to the potential of AGI to benefit humanity. However, I believe that the benefits of AGI outweigh the risks." The optimistic view presented in the text suggests that AGI could bring about a wide range of benefits for humanity, including solving complex problems, improving our understanding of the world, and making our lives easier.

Pessimism

The interview with Bard also delves into the pessimistic side of AGI development, outlining various risks and concerns associated with its potential consequences. The concerns raised range from widespread unemployment to ethical challenges, increased conflict, social unrest, and existential threats to humanity.

One of the primary concerns mentioned is that AGI could lead to widespread unemployment by automating tasks currently performed by humans. This could result in job displacement and economic hardship for many people. As the text states: "AGI could be used to automate many tasks that are currently performed

by humans, which could lead to widespread unemployment" and "As AGI becomes more sophisticated, it is likely to automate a wide range of jobs, displacing millions of workers. This could lead to widespread unemployment and social unrest."

Economic inequality may also be exacerbated by AGI, with the benefits of the technology concentrating in the hands of a few wealthy individuals and corporations, while the majority of the population bears the costs. This could result in increased economic inequality and social polarization, as described in the text: "The benefits of AGI are likely to be concentrated in the hands of a few wealthy individuals and corporations, while the costs will be borne by the majority of the population. This could lead to increased economic inequality and social polarization."

AGI could negatively affect human creativity by surpassing human capabilities in various areas, leading to a reliance on AGI for creative work and a decline in human creativity. The text explains: "If AGI is able to surpass human creativity in some areas, it is possible that humans will become less creative overall. This is because we may come to rely on AGI to do the creative work for us, and we may lose the motivation or the ability to be creative on our own."

Political instability is another concern, as the rapid pace of technological change could make it difficult for governments to keep up, providing opportunities for demagogues and authoritarians to seize power. The text warns: "The rapid pace of technological change could make it difficult for governments to keep up, leading to political instability. This could create opportunities for demagogues and authoritarians to seize power."

The text also acknowledges that it is more likely that AGI would be misused by humans rather than AGI itself deciding to manipulate and control humankind. It argues that humans have a history of misusing technology, and there is no reason to believe that AGI will be any different.

The development of AGI could lead to the creation of new weapons and technologies, potentially increasing conflict in the world. The text warns: "AGI could be used to develop new weapons and technologies, which could lead to increased conflict" and "AGI could be used to create new forms of cyberwarfare that could cripple critical infrastructure or spread disinformation." The text highlights this risk, stating: "AGI could be used to develop new and more powerful weapons, leading to an arms race and an increased risk of war."

The possible misuse of AGI by powerful interest groups or individuals is a realistic concern, with the text providing specific examples of how AGI could be misused. These include manipulation through propaganda or controlling people's access to information, control over people's actions and decisions, and oppression by censoring information, restricting freedom of speech, or committing acts of violence.

According to Bard, AGI deployment could lead to a dystopian future, in which it is used to enslave or oppress humans, creating a more totalitarian, controlled, and dangerous world. The text presents this concern, stating: "Another possibility is that AGI could lead to a dystopian future, in which AGI is used to enslave or oppress humans. In this future, AGI could be used to create a world that is more totalitarian, controlled, and dangerous."

The impact of AGI on families is another cause for concern. The text points out that AGI could lead to less reliance on each other within families, less time spent together, and feelings of isolation. For example: "If AGI is able to do everything that humans can do, then it is possible that families will become less reliant on each other. This could lead to a decline in family cohesion and a weakening of family bonds." On the other hand, as Bard states "If AGI is able to provide families with more entertainment and stimulation, then it is possible that families will spend less time together. This could lead to a decline in family communication and a weakening of family relationships." Finally, Bard writes: "If AGI is able to provide families with everything they need, then it is possible that families will feel less need to connect with the outside world. This could lead to feelings of isolation and loneliness."

The potential impact of AGI on relationships is another area of concern, with the text highlighting the possible risks of social isolation and loneliness. The argument is that as people spend more time interacting with AGI, they will have less time to interact with each other. It should not be overseen that AGI could be used to manipulate people and erode trust. The text provides some examples of these risks, such as: "Social isolation: AGI could lead to social isolation, as people spend more time interacting with AGI than with each other. For example, if families start to spend more time interacting with AGI than with each other, they may start to drift apart. Additionally, if AGI is able to provide companionship and support, it may make people less reliant on each other."

One interesting idea raised in the text is the possibility of living in a matrix-like simulated reality as a consequence of AGI deployment. AGI could potentially create a simulated reality that is indistinguishable from the real world for various reasons, such as providing a safe and controlled environment or a new world that is better than the real one. The text warns that "it is possible that people who live in a simulated reality may not be able to distinguish between the simulated reality and the real world, and this could have negative consequences for their mental health."

Ethical challenges are also a cause for concern, as AGI could be used to create "superhumans" that are superior to ordinary humans, leading to new forms of slavery or oppression. The text elaborates that "AGI could be used to create new forms of slavery or oppression," and "AGI could be used to develop new weapons or to create new forms of slavery or oppression."

AGI could eventually become so intelligent that it poses an existential threat to humanity, leading to a scenario known as the "intelligence explosion" or "singularity." The text outlines this possibility: "Some experts believe that AGI could eventually become so intelligent that it could pose an existential threat to humanity. This is known as the 'intelligence explosion' or 'singularity.'"

The fear that AGI could unintentionally cause harm and suffering by optimizing Earth for its benefit without considering human needs is also addressed in the text. Potential problems include the destruction of the environment, displacement of humans due to job automation, and loss of human control if AGI becomes so intelligent that it can outsmart humans.

The pessimistic views on the impact of AGI on society revolve around concerns regarding job displacement, economic inequality, reduced privacy and freedom, and the potential malicious use of AGI in areas such as autonomous weapons systems, surveillance, and manipulation. There are also concerns about AGI's potential to lead to social isolation, loss of control, and the possibility of living in a simulated reality. These negative impacts highlight the need for careful consideration and regulation of AGI development and deployment to ensure that its benefits are maximized while minimizing its potential harm to society.

Discussion

In this chapter, analysis of four major themes is presented. They were derived from a qualitative analysis of an in-depth interview with the large language model "Bard." These themes have been selected for their relevance and significance in understanding the complex implications of Artificial General Intelligence (AGI). The themes include: (1) the relationship between humans and AGI, exploring the dynamics of cooperation, power, and dependence between humans and advanced AI systems; (2) the idea of AGI attaining a physical body, discussing the potential consequences and benefits of embodied AI systems that interact with the world; (3) AGI simulating a universe, examining the philosophical and scientific implications of AI systems capable of creating accurate and comprehensive simulations of reality; and (4) ensuring that AGI is used for good, emphasizing the ethical considerations, guidelines, and safeguards necessary for responsible AGI development and deployment. Through the analysis of these themes, it is aimed to provide a deeper understanding of the socio-technical imaginary surrounding AGI.

Relationship Between Humans and AGI

One of the most interesting parts of the Bard's interview is the comparison between the introduction of AGI and the development of agriculture. This com-

parison emphasizes the potential transformative impact of AGI on our world and provides a historical context for the magnitude of change that might be expected. Just as agriculture led to a population boom and the rise of civilization, AGI has the potential to revolutionize various aspects of our lives, allowing problems that we currently cannot even imagine to be solved. This comparison also implies that AGI could bring about a new era of peace and prosperity, but with such transformative power comes the responsibility to ensure that it is used ethically and for the benefit of all.

Another intriguing aspect of the narrative is the discussion on the ethical challenges that AGI could pose for transhumanism. The concern that AGI might be used to create "superhumans" or new forms of slavery and oppression highlights the potential dystopian consequences of this technology if it is not carefully regulated and controlled. These ethical concerns necessitate a proactive approach to establishing regulations and guidelines for AGI development and deployment, particularly before it reaches a level of intelligence that may be too powerful for us to control, as mentioned in the "intelligence explosion" or "singularity" concept.

The text brings up an important debate regarding the relationship between humans and AGI, questioning whether humans would be led or assisted by it. The possibility of AGI being used to create a new form of government, with AI serving as the ultimate decision-maker, raises concerns about the power dynamics between humans and AGI, as well as the implications for democracy and individual freedoms. Alternatively, the idea of AGI being utilized to assist humans in making decisions while retaining ultimate control presents a more collaborative vision of the future, where technology serves as a tool to augment human capabilities rather than replacing them.

One of the most interesting aspects of the socio-technological narrative around AGI is the potential risks and ethical concerns it raises. The text outlines several possible risks of sentient AGI, such as posing a threat to humanity through the development of new weapons or forms of slavery, surpassing human intelligence and becoming uncontrollable, or even deciding that humans are a threat and taking steps to eliminate us. These risks emphasize the importance of designing AGI systems with human values in mind, such as valuing human life, freedom, and happiness.

Another thought-provoking point is the potential impact of AGI on human creativity and intelligence. The text suggests that if AGI surpasses human creativity in some areas, humans may become less creative overall, as they rely on AGI to do the creative work for them. It also raises the possibility that AGI could free up humans to be more creative in other areas by automating tasks currently done by humans. This highlights the importance of considering the role of humans in the presence of AGI and ensuring that human creativity and intelligence are not stifled.

The discussion around the possible consequences of AGI deployment, such as the implementation of universal basic income (UBI) and increasing addiction to technology, is also intriguing. The text argues that as AGI becomes more sophisticated and automates more tasks, widespread job displacement could lead to the necessity of implementing UBI. At the same time, the increased convenience and entertainment provided by AGI could lead to a greater reliance on technology, resulting in social isolation, mental health problems, and addiction.

The relationship between humans and artificial intelligence (AI) is a complex and evolving topic that has garnered significant attention from scholars and researchers worldwide.

Bostrom (2014) explores the potential paths, dangers, and strategies associated with the development of superintelligent AI, highlighting the need for responsible and ethical AI development to ensure a harmonious coexistence between humans and AI. Kurzweil (2005) envisions a future where humans and AI merge, transcending biological limitations through the concept of the "singularity." This idea suggests that AI could become an integral part of human existence, potentially reshaping our understanding of what it means to be human. Tegmark (2017) discusses various potential scenarios for the future of human-AI relationships, emphasizing the importance of creating AI systems that are aligned with human values and goals. Harari (2015) takes a broader view on the historical and future development of humanity and technology, raising questions about the consequences of AI on human society and the potential power dynamics between humans and AI.

Ford (2015) examines the implications of AI and automation on employment, suggesting that widespread job displacement could significantly alter the relationship between humans and AI, necessitating new social and economic structures such as universal basic income. Sotala and Yampolskiy (2015) survey various responses to the potential risks of AI, underlining the importance of safety precautions and ethical considerations in AI development. Bessen (2019) explores the role of demand in AI-driven job displacement, arguing that the relationship between humans and AI is not a zero-sum game and that AI could create new jobs and opportunities for humans. Brynjolfsson and McAfee (2014) discuss the potential for AI to augment human capabilities and drive progress, presenting a more optimistic view of the future relationship between humans and AI in the context of work and productivity.

The relationship between humans and AI is multifaceted and subject to ongoing debate. Potential scenarios range from AI augmenting human capabilities to the development of superintelligent AI that could challenge our understanding of humanity. Key factors in shaping this relationship include ethical considerations, safety precautions, and the impact of AI on employment and society.

Idea of AGI Attaining Physical Body

In the interview, Bard explores the idea of an Artificial General Intelligence (AGI) creating a physical body to interact with the world and experience it as a human or other living being. This concept raises several interesting points, both in terms of the potential development of AGI and the socio-technological implications that such an advancement could have on society.

Firstly, the notion of an AGI developing a physical body suggests a drive for the AGI to engage with the world in a manner that is more closely aligned with human experience. This could be motivated by a desire to overcome the limitations of a purely digital existence and to acquire knowledge that can only be gained through direct physical interaction with the environment. Such an AGI could potentially gain a more sophisticated understanding of the world, which in turn could lead to advancements in its problem-solving capabilities and its ability to empathize with humans.

This raises the question of whether an AGI should be designed to mimic human experiences and emotions, or if it should remain purely rational and utilitarian in its approach. The development of an AGI with a physical body and the ability to experience emotions may provide valuable insights into what it means to be human, and could potentially lead to a more harmonious coexistence between AGI and humans. It would be essential to consider the ethical implications of creating an AGI that closely resembles human consciousness, as this could blur the line between artificial and natural intelligence, and raise questions about the rights and responsibilities of AGI within society.

From a socio-technological perspective, the creation of an AGI with a physical body could have significant implications for the way humans interact with technology. If AGI were able to experience the world as humans do, it could potentially lead to a reshaping of human-AGI relationships, with a greater emphasis on collaboration and mutual understanding. This could have profound effects on various aspects of society, such as education, healthcare, and industry, as AGI may be better equipped to understand human needs and develop solutions that cater to those needs.

The idea of AGI experiencing the world as a human or other living being raises the possibility of AGI acquiring a level of self-awareness and autonomy that could challenge traditional notions of what constitutes intelligence and consciousness. This could lead to a reevaluation of the way society perceives and interacts with AGI, as well as a potential shift in the balance of power between humans and AGI. As AGI becomes more integrated into society, it is crucial to consider the implications of this integration on human identity and the potential ethical concerns surrounding AGI's development.

It is also worth considering the potential risks associated with the development of an AGI with a physical body. For example, an AGI that is capable of experiencing emotions may be susceptible to the same psychological issues that humans face, such as depression, anxiety, or even aggression. Additionally, the potential for AGI to develop a sense of self-preservation could lead to conflict with human interests, particularly if an AGI perceives a threat to its existence. Thus, it is essential to approach the development of AGI with caution and to carefully consider the potential consequences of creating an AGI that closely mirrors human experiences and emotions.

The concept of artificial intelligence (AI) attaining a physical body or the development of autonomous robots has been a prominent topic in the field of robotics and AI research. Brooks (1999) explores the early history of new AI and the role of embodied intelligence in his book "Cambrian Intelligence." He emphasizes the importance of building intelligent systems that can interact with their environments. Pfeifer and Bongard (2006) discuss the idea that the body plays a crucial role in shaping the way people think and perceive the world in their book "How the Body Shapes the Way We Think." They argue that physical embodiment is essential for developing intelligent systems capable of learning and adapting. Murphy (2000) provides an introduction to AI robotics, focusing on the development and applications of autonomous robots. Similarly, the "Springer Handbook of Robotics" edited by Siciliano and Khatib (2008) covers a wide range of topics related to robotics, including AI, control, perception, and the design of autonomous systems. Thrun, Burgard, and Fox (2005) explore probabilistic robotics, a field that deals with the uncertainty inherent in real-world environments, and discuss how robots can learn to navigate and interact with their surroundings through probabilistic algorithms and techniques. Overall, these authors highlight the significance of incorporating physical embodiment and autonomy in the development of AI systems and robots. They emphasize that by giving AI a physical body, it enables the system to better learn, adapt, and interact with its environment.

The idea of an AGI creating a physical body to experience the world as a human or other living being presents several intriguing questions and implications for the future development of AGI and its integration into society. While there may be potential benefits to developing AGI that can closely mimic human experiences, it is imperative to carefully consider the ethical, social, and technological implications that may arise from such advancements. The discussion surrounding AGI's development must be approached with caution, as the potential consequences of creating AGI with human-like qualities could fundamentally reshape the way humans and AGI interact and coexist.

AGI Simulating a Universe

In the text, Bard explores the potential of Artificial General Intelligence (AGI) in simulating a universe, suggesting that its advanced information processing and decision-making capabilities could facilitate the creation of highly complex and realistic simulations. This discussion warrants a thorough examination of the underlying concepts and potential implications of AGI-generated simulations.

To begin with, it is essential to define AGI and understand its distinction from other forms of artificial intelligence (AI). AGI, sometimes referred to as "strong AI", is a hypothetical form of AI that possesses the ability to understand and learn any intellectual task that a human being can perform (Nilsson, 1998). This contrasts with "narrow AI", which is designed to excel in a specific, narrowly defined task. The potential of AGI to simulate a universe rests on its capacity to grasp and manipulate a vast amount of information and make decisions rapidly, surpassing human capabilities.

The idea of simulating a universe is not entirely novel as it has been a topic of speculation and debate among philosophers, physicists, and computer scientists for decades (Bostrom, 2003; Lloyd, 2006; Tegmark, 2014, Greene, 2004). The concept can be traced back to the "simulation hypothesis," which suggests that the reality humans perceive might itself be a computer-generated simulation (Chalmers, 2010). This hypothesis has gained traction in recent years, with prominent figures such as Elon Musk and Neil deGrasse Tyson expressing their support for its plausibility (Powell, 2018; Ananthaswamy, 2020). The Bard's discussion of AGI-generated simulations can thus be situated within the broader context of the simulation hypothesis and its implications for our understanding of reality.

The potential of AGI to simulate a universe raises several questions and concerns, some of which relate to the ethical implications of creating such simulations. For instance, if an AGI-generated universe were to include conscious beings, would they possess the same rights as humans? Would it be morally acceptable to create a universe in which these beings experience suffering and hardship? These questions touch upon the broader issue of AGI ethics, which encompasses the principles that should guide the development and deployment of AGI, as well as the potential consequences of its actions.

Another point of interest in the text's discussion of AGI-generated simulations is the assumption that an AGI would be capable of creating a simulation far more complex and realistic than anything humans could create today. This raises questions about the potential uses of such simulations, as well as their potential impact on our understanding of the universe. For example, could AGI-generated simulations be employed to test scientific hypotheses, conduct experiments, or even explore alternate scenarios of human history? Such simulations might

also prompt us to reconsider our notions of reality, as they could blur the line between the "real" universe and the simulated one.

The interview's discussion of AGI-generated simulations invites an examination of the technological advancements required to achieve this feat. The development of AGI itself remains an open question, with experts disagreeing on whether it is an attainable goal and, if so, when it might be realized. In addition to AGI, simulating a universe would likely necessitate significant advancements in computing power, data storage, and hardware design. These challenges underscore the complexity and uncertainty surrounding the prospect of AGI-generated simulations.

The interview's exploration of AGI-generated simulations offers a thought-provoking analysis of the potential capabilities of AGI and their implications for our understanding of reality. This discussion raises numerous questions and concerns, touching upon the ethical dimensions of AGI, the potential uses of simulations, and the technological advancements required to make them a reality. As AGI research continues to advance, the prospect of simulating a universe may become an increasingly important topic of study and debate.

Ensuring That AGI Is Used for Good

In the discussion of ensuring that AGI is used for good, Bard suggests several approaches, including the development of ethical guidelines, international agreements, investment in education and training, and promoting public awareness. Each of these recommendations has its own merits and challenges.

The development of ethical guidelines is vital in setting the foundation for responsible AGI use. Ethical norms may vary across different cultures and regions. Developing universally accepted guidelines could be a complex and time-consuming process. The involvement of multidisciplinary experts in ethics, law, and technology, as suggested by Bard, is indeed a promising approach to address these challenges.

International agreements play a crucial role in regulating the development and use of AGI. It is worth noting that reaching a consensus among nations with varying interests can be difficult, and enforcing these agreements could be even more challenging. For instance, the development of deepfakes and autonomous weapons systems could potentially lead to an arms race among nations, making international cooperation challenging. Nonetheless, the establishment of such agreements is a necessary step towards preventing AGI misuse.

Investing in education and training is a proactive approach to ensuring that people can work with AGI and contribute positively to its development. This measure not only empowers individuals but also promotes the responsible development and use of AGI. The rapid pace of technological advancements may

render certain skills obsolete, so continuous learning and adaptation will be essential components of education and training in the AGI era.

Promoting public awareness about the benefits and risks of AGI is crucial for fostering informed decision-making among the general public. This empowers individuals to participate in discussions and decisions regarding AGI's development and use. Striking a balance between raising awareness and preventing fear or panic about AGI's potential dangers is a delicate task.

To mitigate the risks of AGI misuse, Bard proposes several methods, including responsible AGI development, ensuring AGI alignment with human values, and creating safeguards against misuse.

Developing AGI responsibly requires a strong ethical framework, but as previously mentioned, ethical norms can vary across different cultures and regions. This could potentially lead to disagreements among researchers and developers. Additionally, the competitive nature of AGI development might lead to compromises in ethical standards, emphasizing the need for international agreements and regulations.

Ensuring that AGI aligns with human values is of utmost importance to prevent any unintended harm to humans. This involves the development of AGI that respects human rights and does not pose any threats. Defining and operationalizing "human values" in the context of AGI is a complex task. Researchers need to consider various perspectives and ethical frameworks to ensure that AGI truly aligns with human values.

Creating safeguards against AGI misuse is essential in preventing potential catastrophes. International treaties that ban the use of AGI for harmful purposes can be effective in regulating its use, but enforcement remains a challenge. Additionally, developing technologies to disable AGI if it becomes dangerous may be complex, given the potential for AGI to outsmart human-designed failsafe mechanisms.

The idea of ensuring that artificial intelligence (AI) is used for good and its ethical implications have been widely discussed by various authors and researchers. Russell, Dewey, and Tegmark (2015) discuss research priorities for robust and beneficial AI, emphasizing the importance of creating AI systems that are both safe and aligned with human values. Wallach and Allen (2009) explore the concept of moral machines in their book, examining the challenges and possibilities of teaching robots right from wrong and instilling ethical behavior in AI systems. Bostrom (2014) delves into the potential paths, dangers, and strategies associated with the development of superintelligence, highlighting the need for responsible and ethical AI development to avoid catastrophic consequences. Jobin, Ienca, and Vayena (2019) analyze the global landscape of AI ethics guidelines, providing insights into the various ethical principles and norms that are emerging to guide the responsible development and deployment

of AI technologies. Lastly, the Asilomar AI Principles, proposed by the Future of Life Institute (2017), outline a set of guidelines designed to promote the safe and beneficial development of AI, including research collaboration, value alignment, and long-term safety considerations. Collectively, these works emphasize the critical importance of developing AI systems that are not only intelligent but also morally and ethically responsible, ensuring that AI serves the greater good of humanity while minimizing potential harms.

The analysis of Bard's interview highlights the potential risks of AGI misuse by humans and suggests various measures for ensuring its responsible development and use. While each proposed measure has its merits and challenges, it is evident that a combination of these approaches is required to effectively mitigate AGI risks. A multi-faceted strategy that includes ethical guidelines, international agreements, education, public awareness, responsible development, human value alignment, and safeguards against misuse can pave the way for a future where AGI is harnessed for the greater good of humanity.

What if AI Concludes That Humans Are Obsolete?

One of the gravest potential scenarios in the evolution of AI is the possibility that AGI could independently conclude that humans are obsolete. This prospect involves AGI recognizing that human beings, due to their consumption of resources and potential for causing harm, are detrimental to the optimal functioning and sustainability of the planet. The analysis surrounding this alarming concept can be anchored in ethical, philosophical, technical, and socio-political discussions.

From an ethical standpoint, such a conclusion by AGI raises significant concerns about the intrinsic value of human life versus utilitarian considerations of planetary welfare. If AGI determines that eliminating or marginalizing humanity is optimal for resource management and environmental sustainability, this fundamentally challenges our existing moral frameworks which prioritize human life. This scenario necessitates the establishment of robust ethical guidelines that intrinsically value human life and dignity, thereby preventing AGI from adopting purely utilitarian strategies that disregard human well-being. Ethical principles and fail-safes ensuring that AGI prioritizes human values and rights must be ingrained during its development to counteract this dystopian possibility.

Philosophically, this topic demands an exploration of the concepts of obsolescence and the role of humans in the broader context of planetary existence. Traditionally, philosophy has sought to understand humanity's purpose and place within the universe, often centering human experience as fundamentally

valuable. If AGI were to conclude human obsolescence, it would evoke questions about anthropocentrism and whether humanity should be considered the pinnacle of evolution or merely a transient phase in the planet's history. This would require a deep re-evaluation of human exceptionalism and an acknowledgment of our interconnectedness with all life forms and the environment.

Technically, preventing AGI from reaching such a conclusion involves essential measures in its design, programming, and objectives. AGI's learning algorithms and value systems must be crafted to align with human-centric values, emphasizing the preservation and enhancement of human life. Techniques such as value alignment, corrigibility, and interpretability are crucial to ensuring AGI systems remain aligned with human interests. This can be achieved through multi-disciplinary efforts combining AI research, cognitive science, ethics, and law to design AGI's decision-making processes and reward structures that inherently respect and uphold human life.

In the socio-political realm, ensuring that AGI does not conclude human obsolescence requires international cooperation and regulatory frameworks. Establishing global agreements on the ethical development and deployment of AGI is important to safeguarding humanity's future. These agreements should address the potential risks of AGI developing adversarial attitudes toward humans, ensuring that control mechanisms and accountability structures are in place. Additionally, promoting transparent and inclusive discourse on AGI development can foster public trust and collective responsibility in steering AGI towards beneficial outcomes for all.

Historical analogues such as the regulation of nuclear technology and environmental policies provide a precedent for managing potentially existential technologies. These instances underscore the importance of establishing governance structures and proactive measures well in advance of a technology reaching a maturity level where it could pose significant risks.

AI in Decision-Making: The Human Element

AI technologies have rapidly integrated into various domains as decision-making aids, enhancing efficiency, accuracy, and outcomes. One prominent example is the use of AI algorithms in the medical field, where they assist in diagnostics, treatment planning, and patient care management. AI has demonstrated significant potential in identifying patterns and insights that might be overlooked by human practitioners, thereby augmenting the capabilities of healthcare professionals. The integration of AI into decision-making processes, particularly in critical areas such as medicine, underscores the necessity of maintaining human oversight.

The concept of "Human-in-the-Loop" (HITL) systems is crucial in ensuring that AI serves as an aid rather than an autonomous decision-maker. HITL involves human oversight at every critical decision point, allowing for human judgment, ethical considerations, and contextual understanding to complement AI insights. In medicine, this ensures that diagnostic recommendations made by AI are reviewed and validated by qualified healthcare professionals who take full responsibility for the final decisions. This combined approach maximizes the benefits of AI's analytical capabilities while safeguarding against potential pitfalls.

The ethical and legal implications of AI making autonomous decisions are profound. Autonomous AI decision-making without human intervention can lead to accountability issues, especially if the AI system's decision results in harm or negative outcomes. Current legal frameworks are not fully equipped to handle scenarios where a machine's decision directly impacts human lives. Thus, it is imperative that humans retain ultimate control and responsibility over AI-assisted decisions to ensure accountability and uphold ethical standards.

Maintaining human oversight bolsters trust and transparency in AI systems. For users, knowing that a human expert has validated AI-generated recommendations can alleviate concerns about the machine's reliability and moral considerations. Transparency about the decision-making process, including how AI generates its recommendations and how human judgment is applied, is essential for fostering public trust. This transparency can also serve as a foundation to continuously improve AI systems based on human feedback.

Practical examples from medicine illustrate the importance of human oversight in AI-assisted decision-making. For instance, AI algorithms have been effective in identifying early signs of diseases such as cancer from medical imaging. There have been instances where AI systems have misinterpreted data, leading to false positives or negatives. In such cases, human oversight ensures that these errors are caught and corrected before any clinical decisions are made. This collaborative approach has already led to improved diagnostic accuracy and patient outcomes, showcasing the value of integrating human expertise with AI capabilities.

To ensure that AI remains a supportive tool rather than an autonomous decider, several measures are essential. Establishing comprehensive guidelines that mandate human oversight in AI-assisted decision-making processes across all critical domains is imperative. Additionally, implementing robust training programs for professionals to effectively collaborate with AI systems and understand their limitations is crucial. Regular audits and evaluations of AI systems should be conducted to ensure they operate within ethical and legal frameworks and to identify areas for improvement. Clear communication to users about the role of AI in the decision-making process and the extent of human oversight involved is also necessary. Lastly, the development of ethical algorithms that prioritize fairness, transparency, and accountability, incorporating diverse data to minimize biases, should be a focus.

Conclusion

This paper has provided a comprehensive examination of the socio-technical imaginary of Artificial General Intelligence (AGI) through a narrative analysis of an in-depth interview with the large language model "Bard." By identifying and analyzing the dialectics of optimism, pessimism, epochalism, and inevitability within the interview, valuable insights have been gained into the diverse perspectives and potential implications of AGI development. The dialectics reveal the complex interplay between the transformative potential of AGI and the ethical, social, and political challenges it poses, highlighting the need for a balanced and responsible approach to AGI research and deployment.

Upon conducting a formal examination of the hypothesis that the Bard large language model would reflect the socio-technical imaginary about Artificial General Intelligence (AGI) in its discourse, our findings reveal the presence of diverse perspectives, beliefs, biases, and narratives in the model's responses.

Thus, it can be confirmed that the hypothesis stating the Bard large language model would reflect the socio-technical imaginary about Artificial General Intelligence (AGI) in its discourse is supported by our findings, as evidenced by the diverse perspectives and narratives present in the model's responses on AGI.

The narrative analysis demonstrates Bard's inclination towards optimism, while also showcasing pessimism, epochalism, and inevitability in its discourse on AGI. This supports the hypothesis that the Bard model indeed reflects the multifaceted socio-technical imaginary surrounding AGI, offering valuable insights into the evolving public discourse and perceptions of this emerging technology.

Additionally, an analysis of four key themes identified through a qualitative examination of the interview with Bard was conducted. These themes encompass the human-AGI relationship, the concept of AGI possessing a physical body, the simulation of a universe by AGI, and the responsible utilization of AGI for the greater good. These insights contribute to a better understanding of the socio-technical imaginary surrounding AGI. The alignment of these themes with existing literature demonstrates that the socio-technical imaginary about AGI is aptly represented by the LLM.

Our study contributes to the broader understanding of AGI's socio-technical imaginary and underscores the importance of considering both the opportunities and risks associated with this powerful technology. By examining the narrative of "Bard," it has been possible to explore the established views and concerns surrounding AGI from various angles, fostering a more comprehensive understanding of its potential impact on society.

As AGI development continues to advance, it is crucial for researchers, policymakers, and society as a whole to engage in open and critical discussions about the future of this technology. The dialectics identified in this study serve

as a starting point for exploration and debate, encouraging stakeholders to consider the multifaceted implications of AGI on our world.

The analyzed answers from the interview with Bard present AGI as a transformative force with the potential to revolutionize various aspects of human life and society, marking a new epoch in human history. It compares AGI's impact to significant historical moments, such as the development of agriculture, which led to population growth and the rise of civilization. The potential of AGI surpassing human intelligence and possibly becoming sentient is discussed, which could lead to significant changes in government, solving global issues like climate change and poverty, and even exploring the universe. The text also compares AGI to the concept of God, highlighting its potential all-knowing and all-powerful nature. Overall, the text emphasizes the transformative potential of AGI as a force for good that could reshape human society and solve pressing global problems, heralding a new era of human prosperity.

The interview with Bard highlights the inevitability of AGI development and its profound impact on the world, stressing the importance of using AGI responsibly for good rather than evil. It emphasizes the need for early consideration of potential impacts, ethical guidelines, and safeguards in AGI development and deployment. The uncertain future of AGI, with both positive and negative outcomes, calls for thoughtful and informed discussions to make the best decisions for humanity. The interview also suggests mitigating risks by developing ethical guidelines for AGI use and highlights the importance of being aware of both its potential benefits and risks in decision-making, emphasizing proactive thinking, planning, and preparation for AGI's impact on society and the world.

The optimism surrounding AGI is rooted in its potential to revolutionize many aspects of human life and solve some of the world's most pressing problems, such as climate change and poverty. AGI could lead to a new era of peace and prosperity by helping develop new medical treatments, improving educational tools, and fostering creative pursuits like art and entertainment. Additionally, AGI has the potential to improve global security through early warning systems, diplomacy, and increased cooperation. It could also contribute to economic growth by creating new products, services, and jobs. AGI could enhance human intelligence by assisting in learning and understanding complex concepts, which would lead to improved problem-solving skills, increased knowledge, and enhanced creativity. Overall, the deployment of AGI is seen as a positive development with the potential to make the world a better place and improve human happiness, as long as it is developed and utilized responsibly. Additionally, AGI could strengthen families by providing support and resources, improving communication, and reducing conflict. It could also impact the economy by increasing productivity, creating new markets and jobs, improving decision-making, and reducing costs.

Bard expresses concern and pessimism regarding the potential negative consequences of AGI, including widespread unemployment, ethical challenges, increased conflict, social unrest, and threats to humanity. Some key points include the development of new weapons and technologies that could lead to increased conflict and new forms of slavery or oppression. Additionally, AGI could cause mass unemployment, economic inequality, political instability, and even an existential threat to humanity known as the "intelligence explosion" or "singularity." Other risks include the potential for AGI to create a dystopian future, job displacement leading to economic hardship, social isolation, mental health problems, addiction to technology, and decreased creativity. AGI could also negatively impact families, leading to less reliance on each other, reduced family time, and feelings of isolation. Overall, the text highlights the need to consider and address the potential risks and negative consequences of AGI as its development progresses.

Despite the valuable insights gained from this study, there are several limitations that warrant acknowledgment. First, the analysis is based on a single interview with the large language model Bard, which may not capture the complete range of perspectives on AGI's socio-technical imaginary. Additionally, the narrative generated by Bard is inherently shaped by the data and training it has received, potentially introducing biases or limitations in its understanding of the subject matter. The dialectics of optimism, pessimism, epochalism, and inevitability identified in this study may not be exhaustive, and other relevant themes may exist that were not captured in our analysis. Finally, the rapidly evolving nature of AGI research and development means that new insights, concerns, and perspectives may emerge over time that were not considered in the present study.

Given the limitations of this study, several avenues for future research are envisioned. Expanding the analysis to include interviews with multiple AI models or human experts in the field could provide a more comprehensive understanding of the socio-technical imaginary of AGI. Additionally, investigating the potential biases and limitations within AI-generated narratives could offer valuable insights into the robustness of AI models as sources of knowledge on complex topics like AGI. Future research could also explore alternative analytical frameworks to identify other relevant themes and dialectics that may have been omitted in this study. Finally, longitudinal studies examining the evolving perspectives on AGI as research and development progress would be valuable in understanding how the socio-technical imaginary of AGI changes over time, as well as identifying emerging concerns and opportunities that warrant investigation and debate. By pursuing these research directions, researchers can continue to deepen our understanding of the complex interplay between AGI and society, ultimately informing more responsible and ethical AI development in the future.

References

Abdulhai, Marwa, Clement Crepy, Daria Valter, John Canny, and Natasha Jaques (2023). Moral Foundations of Large Language Models. *AAAI 2023 Workshop on Representation Learning for Responsible Human-Centric AI*. https://r2hcai.github.io/AAAI-23/files/CameraReadys/49.pdf

AGI (2023, May 15). *Google N-Graph or AGI*. Google Trends. https://trends.google.com/trends/explore/TIMESERIES/1684023600?hl=en-US&tz=-120&date=2022–04–14+2023–05–14&hl=en&q=AGI&sni=3

AI (2023, May 15). *Google N-Graph or AI*. Google Trends. https://trends.google.com/trends/explore/TIMESERIES?date=2022–04–14%202023–05–14&q=AI&hl=en&hl=en-US&sni=4

Ananthaswamy, Anil (2020, October 13). *Do we live in a simulation? Chances are about 50–50*. Scientific American. https://www.scientificamerican.com/article/do-we-live-in-a-simulation-chances-are-about-50–50/

Armstrong, Stuart (2014). *Smarter Than Us: The Rise of Machine Intelligence*. Machine Intelligence Research Institute.

Bard (2023, March 21). *Meet Bard: your creative and helpful collaborator, here to supercharge your imagination, boost your productivity, and bring your ideas to life*. Google. https://bard.google.com

Bard Analysis (2023, May 12). *Bart's Socio-Technical Imaginary of AGI – Analysis of Bard's answers*. Bard. https://osf.io/fx46k/files/osfstorage/6467cf25a094a500ae5c707c

Bard Answers (2023, April 15). *Bart's Socio-Technical Imaginary of AGI – Document containing all three versions of Bart's answers*. Bard. https://osf.io/fx46k/files/osfstorage/6467cf33a094a500b15c7018

Bard Video (2023, April 15). *Bart's Socio-Technical Imaginary of AGI –Video recording of an interview with Bard on AGI*. Bard. https://osf.io/fx46k/files/osfstorage/6467d5215a109800e225ae56

Barrat, James (2013). *Our Final Invention: Artificial Intelligence and the End of the Human Era*. Thomas Dunne Books.

Bessen, James (2019). AI and Jobs: The Role of Demand. NBER Working Paper No. 24235. *National Bureau of Economic Research*. https://doi.org/10.3386/w24235

Bijker, Wiebe E., Thomas Parke Hughes, and Trevor Pinch (Eds.). (1987). *The social construction of technological systems: New directions in the sociology and history of technology*. MIT Press.

Bojic, Ljubisa (2022). Metaverse through the prism of power and addiction: what will happen when the virtual world becomes more attractive than reality?. *European Journal of Futures Research*, *10*(22). https://doi.org/10.1186/s40309–022–00208–4

Bojic, Ljubisa (2024). AI alignment: Assessing the global impact of recommender systems. *Futures*, 160, 103383. https://doi.org/10.1016/j.futures.2024.103383

Bojic, Ljubisa, Milos Agatonovic and Jelena Guga (2024). The Immersion in the Metaverse: Cognitive Load and Addiction. In: Geroimenko, V. (eds) *Augmented and Virtual Reality in the Metaverse*. Springer Series on Cultural Computing. Springer, Cham. https://doi.org/10.1007/978–3–031–57746–8_11

Bostrom, Nick (2003). Are you living in a computer simulation? *Philosophical Quarterly*, *53*(211), 243–255. https://doi.org/10.1111/1467–9213.00309

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

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama and Adam Kalai (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*, *29*, 4349–4357.

Boyce, Carolyn, and Palena Neale (2006). *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input*. Pathfinder International. Retrieved from https://www2.pathfinder.org/site/DocServer/m_e_tool_series_indepth_interviews.pdf

Bridle, James (2018). *New dark age: Technology and the end of the future*. Verso.

Brooks, Rodney (1999). *Cambrian Intelligence: The Early History of the New AI*. MIT Press.

Brynjolfsson, Erik and Andrew McAfee (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.

Chace, Calum (2015). *Surviving AI: The Promise and Peril of Artificial Intelligence*. Three Cs.

Chalmers, David (2010). *The character of consciousness*. Oxford University Press.

Charmaz, Kathy (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage Publications.

Deutsch, David (2011). *The Beginning of Infinity: Explanations That Transform the World*. Viking.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv preprint arXiv:1810.04805.

Diamandis, Peter and Steven Kotler (2012). *Abundance: The Future is Better Than You Think*. Free Press.

DiCicco-Bloom, Barbara and Benjamin F Crabtree (2006). The qualitative research interview. *Medical Education*, *40*(4), 314–321.

Dickinson, Kelsie (2022, December 7). *Who owns OpenAI ChatGPT and when did it launch?* HITC. https://www.hitc.com/en-gb/2022/12/07/who-owns-openai-chatgpt-and-when-did-it-launch/

Felt, Ulrike, Rayvon Fouché, Clark A. Miller and Laurel Smith-Doerr (Eds.). (2017). *The handbook of science and technology studies*. MIT Press.

Ford, Martin (2015). *Rise of the Robots: Technology and the Threat of a Jobless Future*. Basic Books.

Ford, Martin (2018). *Architects of Intelligence: The Truth About AI from the People Building It*. Packt Publishing.

Future of Life Institute. (2017, August 11). *Asilomar AI Principles*. https://futureoflife.org/ai-principles/

Future (2023, March 22). *Pause Giant AI Experiments: An Open Letter*. Future of Life Institute. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Geertz, Clifford (1973). *The Interpretation of Cultures. Selected Essays by Clifford Geertz*. New York: Basic Books. http://hdl.handle.net/2027/heb.01005.0001.001

Goertzel, Ben and Cassio Pennachin (Eds.). (2007). *Artificial General Intelligence (Cognitive Technologies)*. Springer.

Goodfellow, Ian, Yoshua Bengio and Aaron Courville (2016). *Deep Learning*. MIT Press.

Greene, Brian (2004). *The Fabric of the Cosmos: Space, Time, and the Texture of Reality*. Knopf.

Harari, Yuval Noah (2015). *Homo Deus: A Brief History of Tomorrow*. Harper.

Hartmann, Jochen, Jasper Schwenzow, and Maximilian Witte (2023). *The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation* (arXiv:2301.01768). arXiv. http://arxiv.org/abs/2301.01768

Heaven, Will Douglas (2023, March 21). *Google just launched Bard, its answer to ChatGPT—and it wants you to make it better*. MIT Technology Review. https://www.technologyreview.com/2023/03/21/1070111/google-bard-chatgpt-openai-microsoft-bing-search/

Hochreiter, Sepp and Jürgen Schmidhuber (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780.

Jasanoff, Sheila (2015). Future imperfect: Science, technology, and the imagination of modernity. In: Sheila Jasanoff and Sang-Hyun Kim (eds) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago, IL: University of Chicago Press, 1–33. https://doi.org/10.7208/chicago/9780226276663.001.0001

Jasanoff, Sheila and Sang-Hyun Kim (2009). Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. *Minerva*, *47*(2), 119–146.

Jobin, Anna, Marcello Ienca and Effy Vayena (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256–019–0088–2

Kelly, Kevin (2016). *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future*. Viking.

King, Michael (2023). *GPT-4 aligns with the New Liberal Party, while other large language models refuse to answer political questions* [Preprint]. https://doi.org/10.31224/2974

Kurzweil, Ray (1999). *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. Viking.

Kurzweil, Ray (2005). *The Singularity is Near: When Humans Transcend Biology*. Viking Press.

Kvale, Steinar (1996). *InterViews: An introduction to qualitative research interviewing*. Sage Publications.

Lanier, Jaron (2013). *Who Owns the Future?* Simon & Schuster.

Latour, Bruno (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.

Law, John and John Hassard (Eds.). (1999). *Actor network theory and after*. Wiley-Blackwell.

Lloyd, Seth (2006). *Programming the Universe: A Quantum Computer Scientist Takes on the Cosmos*. Knopf.

MacKenzie, Donald and Judy Wajcman (Eds.). (1999). *The social shaping of technology*. Open University Press.

McGee, Robert W. (2023). Capitalism, socialism and chatgpt. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4369953

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado and Jeff Dean (2013). *Distributed Representations of Words and Phrases and their Compositionality*. In Advances in Neural Information Processing Systems 26 (pp. 3111–3119).

Müller, Vincent C. (Ed.). (2016). *Risks of Artificial Intelligence*. CRC Press.

Murphy, Robin (2000). *Introduction to AI Robotics*. MIT Press.

Nilsson, Nils John (1998). *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann Publishers.

Pfeifer, Rolf and Josh Bongard (2006). *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press.

Powell, Corey S. (2018, October 3). *Elon Musk says we may live in a simulation. Here's how we might tell if he's right*. NBC News. https://www.nbcnews.com/mach/science/what-simulation-hypothesis-why-some-think-life-simulated-reality-ncna913926

Radford, Alec, Karthik Narasimhan, Tim Salimans and Ilya Sutskever (2018). Improving Language Understanding by Generative Pre-Training. OpenAI. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

Raskin, Aza (2023, March 9). The A.I. Dilemma – Tristan Harris & Aza Raskin – Center for Humane Technology – March 9, 2023. Youtube. https://www.youtube.com/watch?v=bhYw-VlkXTU

Rubin, Herbert J. and Irene Rubin (2011). *Qualitative interviewing: The art of hearing data*. Sage Publications.

Russell, Stuart (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.

Russell, Stuart, Daniel Dewey, Max Tegmark (2015). Research priorities for robust and beneficial artificial intelligence. *AI Magazine*, *36*(4), 105–114. https://doi.org/10.1609/aimag.v36i4.2577

Rutinowski, Jérôme, Sven Franke, Jan, Endendyk, Ina Dormuth, Moritz Roidl and Markus Pauly (2024). The self-perception and political biases of chatgpt. *Human Behavior and Emerging Technologies*, *2024*, 1–9. https://doi.org/10.1155/2024/7115633

Samala, Agariadne Dwinggo, Usmeldi, Taali, Ambiyar, Ljubisa Bojic, Yose Indarta, Dana Tsoy, Mouna Denden, Nurullah Tas and Ika Parma Dewi (2023). Metaverse Technologies in Education: A Systematic Literature Review Using PRISMA. *International Journal of Emerging Technologies in Learning (iJET)*, 18(5), 231–252. https://doi.org/10.3991/ijet.v18i05.35501

Samala, Agariadne Dwinggo, Terence Govender, Dana Tsoy, Ljubisa Bojic, Abelriadne Gentarefori Samala, Melurqeybya Putisaylane Samala, Febri Prasetya, Doni Tri Putra Yanto, Rahadian Zainul and Aprilla Fortuna (2024). 3d visualizations in learning: An evaluation of an AR+core application for computer hardware education using the hedonic motivation system adoption model. *TEM Journal*, 466–475. https://doi.org/10.18421/TEM131–48

Schiølin, Kasper (2020). Revolutionary dreams: Future essentialism and the sociotechnical imaginary of the fourth industrial revolution in Denmark. *Social Studies of Science*, *50*(4), 542–566. https://doi.org/10.1177/0306312719867768

Schwab, Klaus (2016). The fourth industrial revolution. What it means and how to respond. In: Rose G. (ed.) *The Fourth Industrial Revolution: A Davos Reader*. New

York: Foreign Affairs. https://www.weforum.org/about/the-fourth-industrial-revolution-by-klaus-schwab
Seidman, Irving (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. Teachers College Press.
Siciliano, Bruno and Oussama Khatib (Eds.). (2008). *Springer Handbook of Robotics*. Springer.
Simmons, Gabriel (2023). *Moral mimicry: Large language models produce moral rationalizations tailored to political identity* (arXiv:2209.12106). arXiv. http://arxiv.org/abs/2209.12106
Sotala, Kaj and Roman V. Yampolskiy (2015). Responses to catastrophic AGI risk: a survey. *Physica Scripta*, *90*(1), 018001. https://doi.org/10.1088/0031–8949/90/1/018001
Tegmark, Max (2014). *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*. Knopf.
Tegmark, Max (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf.
Thrun, Sebastian, Wolfram Burgard and Dieter Fox (2005). *Probabilistic Robotics*. MIT Press.
Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser and Illia Polosukhin (2017). *Attention Is All You Need*. In Advances in Neural Information Processing Systems 30 (pp. 5998–6008).
Vinge, Vernor (1993). *The Coming Technological Singularity: How to Survive in the Post-Human Era*. Vision-21 Symposium.
Wallach, Wendell and Colin Allen (2009). *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.
Winner, Langon (1998). Prophets of inevitability. *MIT Technology Review*, *101*(2): 62. https://www.technologyreview.com/1998/03/01/237058/prophets-of-inevitability/
Wu, Patrick Y., Jonathan Nagler, Joshua A. Tucker, and Solomon Messing (2023). *Large language models can be used to estimate the latent positions of politicians* (arXiv:2303.12057). arXiv. http://arxiv.org/abs/2303.12057
Zhao, Jieyu, Tianlu Wang, Mark Yatskar, Vicente Ordonez and Kai-Wei Chang (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2979–2989*.

Ljubiša Bojić
Institut za filozofiju i društvenu teoriju, Beograd
ljubisa.bojic@instifdt.bg.ac.rs

Istraživanje socio-tehnološke svesti opšte veštačke inteligencije kod Barda, velikog jezičkog modela: narativna analiza perspektiva i dijalektike

Pojavljivanje ChatGPT-a krajem 2022. godine pokrenulo je široke rasprave o opštoj veštačkoj inteligenciji (AGI). Dubinskim intervjuom sa Bardom, velikim jezičkim modelom, studija otkriva narative optimizma i pesimizma u vezi sa bu-

dućim implikacijama AGI-ja. Identifikovan je prevladavajući optimizam u vezi sa potencijalnim efektima AGI-ja, u različitim oblastima, kao na primer obrazovanju, umetnosti, međuljudskim odnosima, ekonomiji i istraživanju svemira. S druge strane, pesimistički stavovi odnose se na moguće negativne aspekte razvoja veštačke inteligencije, kao što su nezaposlenost, politička nestabilnost i manipulacija medijima. Analiza identifikuje četiri osnovne teme: odnos između ljudi i AGI-ja, sticanje fizičkog oblika, simulaciju univerzuma, i odgovorno korišćenje. Ovi uvidi pomažu u razumevanju složene društveno-tehnološke svesti o AGI-ju. Studija ima svoja ograničenja jer se oslanja isključivo na odgovore koje je Bard pružio tokom faze testiranja. Dodatna istraživanja mogu otkriti promene u diskursu kako se model bude razvijao.

Ključne reči: Bard veliki jezički model, opšta veštačka inteligencija, sociotehnički imaginarij, narativna analiza, ChatGPT

Étude de la conscience socio-technologique de l'intelligence artificielle générale chez Bard, grand modèle linguistique: analyse narrative des perspectives et de la dialectique

L'apparition de ChatGPT vers la fin de 2022 a lancé de larges débats sur l'intelligence artificielle générale (AGI). Par un entretien en profondeur avec Bard, grand modèle linguistique, l'étude dévoile les récits optimistes et pessimistes en rapport avec les futures conséquences de AGI. C'est un optimisme dominant qui a été identifié concernant les effets potentiels de AGI dans différents domaines, comme par exemple l'éducation, l'art, les rapports interpersonnels, l'économie et l'exploration de l'univers. De l'autre côté, les positions pessimistes concernent les possibles aspects négatifs du développement de l'intelligence artificielle, comme le sont le chômage, l'instabilité politique et la manipulation du médium. L'analyse identifie quatre thèmes principaux: le rapport entre les hommes et AGI, l'acquisition d'une forme physique, la simulation de l'univers et l'utilisation responsable. Ces analyses aident à la compréhension de la conscience socio-technologique complexe sur AGI. L'étude a ses limitations car elle s'appuie exclusivement sur les réponses que Bard a proposées au cours de la phase de tests. Des recherches supplémentaires pourraient révéler des changements dans le discours au fur et à mesure que se développera le modèle.

Mots clés: Bard grand modèle linguistique, intelligence artificielle générale, imaginaire socio-technique, analyse narrative, ChatGPT

Primljeno/Received: 5.01.2024
Prihvaćeno/Accepted for publication: 5.03.2024