Available online at: http://journal.uinsgd.ac.id/index.php/kt
Khazanah Theologia, Vol. 6 No. 1 (2024): 1-12

DOI: 10.15575/kt.v6i1.33559

Cybertheology and the Ethical Dimensions of Artificial Superintelligence: A Theological Inquiry into Existential Risks

Ted Peters*
Professor emeritus at the Graduate Theological Union, Berkeley, United States
Email: tedfpeters@gmail.com

Abstract

Purpose: This study explores the role of cybertheology in addressing the ethical and societal challenges posed by Artificial Superintelligence (ASI), which has the potential to surpass human cognitive capabilities, heralding a profound cultural and existential crisis. It integrates theological anthropology to assess the implications of a posthuman future. **Methodology**: Utilising a comprehensive literature review, the research examines technological, philosophical, and theological perspectives through primary and secondary sources, including influential works by futurists and ethicists. The methodology aims to uncover the nuanced discourse surrounding the development of ASI and its potential impacts. **Findings**: The analysis reveals a narrative marked by speculative optimism and significant existential concerns regarding ASI. A critical gap in the existing ethical discourse is identified, highlighting the necessity for a grounded ethical framework that addresses the profound implications of superintelligent entities on human dignity and societal norms. **Research Implications**: The findings emphasise the urgent need to incorporate robust ethical considerations into the development and deployment of ASI. Cybertheology is presented as a vital framework for ensuring that ASI technologies align with human values and theological insights, thus providing a valuable lens through which to view the integration of superintelligence into society. **Originality/Value**: This paper contributes to academic and policy discussions on ASI by promoting cybertheology as a crucial perspective in ethical deliberations. It enriches scholarly dialogues by linking technological advancements with theological and ethical evaluations, proposing that cybertheology can play a pivotal role in shaping policies that govern ASI technologies.

Keywords: Artificial Superintelligence; Cybertheology; Ethical Implications; Existential Risk.

Abstrak

Tujuan: Penelitian ini mengeksplorasi peran siberteologi dalam menangani tantangan etis dan sosial yang ditimbulkan oleh Artificial Superintelligence (Kecerdasan Super Artifisial), yang memiliki potensi untuk melampaui kemampuan kognitif manusia, mengarah pada krisis budaya dan eksistensial yang mendalam. Penelitian ini mengintegrasikan antropologi teologis untuk menilai implikasi dari masa depan pasca-manusia. **Metodologi**: Dengan menggunakan tinjauan literatur yang komprehensif, penelitian ini mengkaji perspektif teknologi, filsafat, dan teologi melalui sumber primer dan sekunder, termasuk karya-karya berpengaruh dari para futuris dan etikus. Metodologi ini bertujuan untuk mengungkap diskursus yang rumit seputar pengembangan ASI dan dampak potensialnya. **Temuan**: Analisis mengungkapkan narasi yang ditandai oleh optimisme spekulatif dan kekhawatiran eksistensial yang signifikan mengenai ASI. Sebuah kesenjangan kritis dalam wacana etis yang ada diidentifikasi, menyoroti perlunya kerangka etis yang kokoh yang menangani implikasi mendalam dari entitas superintelligent terhadap martabat manusia dan norma sosial. **Implikasi Penelitian**: Temuan ini menekankan kebutuhan mendesak untuk memasukkan pertimbangan etis yang kuat dalam pengembangan dan penerapan ASI. Siberteologi dipresentasikan sebagai kerangka penting untuk memastikan bahwa teknologi ASI selaras dengan nilai-nilai manusia dan wawasan teologis, sehingga menyediakan lensa berharga untuk melihat integrasi superintelligence ke dalam masyarakat. **Orisinalitas/Nilai**: Penelitian ini berkontribusi pada diskusi akademik dan kebijakan tentang ASI dengan mempromosikan siberteologi sebagai perspektif penting dalam pertimbangan etis. Ini memperkaya dialog ilmiah dengan menghubungkan kemajuan teknologi dengan evaluasi teologis dan etis, mengusulkan bahwa siberteologi dapat memainkan peran penting dalam membentuk kebijakan yang mengatur teknologi ASI.

Kata Kunci: Kecerdasan Super Artifisial; Siberteologi; Implikasi Etis; Risiko Eksistensial.

*Corresponding Author
Received: January 30, 2024; Revised: April 21, 2024; Accepted: May 18, 2024

INTRODUCTION

Artificial Intelligence (AI) has revolutionised the religious landscape, facilitating global connectivity and enabling individuals to engage with religious teachings and rituals regardless of their physical location. This transformation redefines religious practices, allowing for the formation of virtual communities that span geographical and cultural boundaries (Florea & Gilder, 2024). Additionally, AI challenges traditional religious principles by reshaping religious symbols and creating new spiritual meanings that may conflict with existing doctrines (Umbrello, 2023). These changes not only expand the scope of religious influence but also question the boundaries of religious identity and beliefs in the digital era.

Religious perspectives on AI emphasise the importance of considering the social impacts of this technology, with an approach that prioritises relationality and fluidity (Trothen, Kwok, & Lee, 2024). This concern stems from the understanding that technology should be harmonious with social values and communal ethics, not merely a tool for economic advancement. This perspective offers a significant contrast to the dominant narratives that often explore AI from the standpoint of individual gain and technological sophistication, providing an alternative viewpoint that can enrich the global discussion on the integration of AI in human life.

Building upon the transformative influence of Artificial Intelligence (AI) on religion and its broader social implications, a crucial question arises: should the cybertheologian inquire about God in this context? Sonny Zaluchu (2024) asserts, "Dialogue about God occurs in new forms as a manifestation of cyber theology." Indeed, while the role of God remains paramount, there is an acute need for insights drawn from theological scholarship on the human condition in relation to technological advancements. This wisdom, informed by Scripture and tradition, is essential for navigating the complexities introduced by AI, as it profoundly redefines the relationship between human identity and technological progress. These theological reflections are vital in guiding not only the spiritual but also the ethical dimensions of AI integration into society, ensuring that technological progress aligns with enduring human values and the collective good.

In the context of existing literature, the approach of cybertheology within the larger school of public theology is to understand and address the challenges presented by artificial intelligence (AI). This approach continues a dialogue established by previous studies that explore the intersection of technology with spirituality. Works by Antonio Spadaro (2014) and Sonny Zaluchu (2024) highlight how digital technology and AI are reshaping the ways we discuss God and spirituality, presenting a new paradigm in theological studies. This research also aligns with Umbrello's (2023) analysis on how religious symbols are adapted and reinterpreted in the digital age, showcasing shifts in meaning that may conflict with traditional doctrines. Further, the contributions of Trothen, Kwok, and Lee (2024) emphasise the importance of considering the social impacts of AI, a central theme in the discourse of cybertheology. This study integrates insights from various sources to construct a robust theoretical foundation for addressing the ethical and theological dilemmas that arise with the advancement of AI, ensuring that discussions about AI integration into human life are not solely focused on technological progress but also consider human values and collective ethics.

While existing research has extensively explored the theoretical intersections between digital technology and spirituality, there remains a distinct gap in empirical analyses that directly assess the real-world implications of these theories on societal structures and individual behaviors. Previous studies have primarily focused on conceptual frameworks and the reinterpretation of religious symbols in the digital era, yet few have addressed how these shifts impact societal norms and governance in a practical context. This study seeks to fill this gap by empirically investigating the specific ways in which cybertheology can

ISSN 2715-9701 (online)

influence policy-making and community practices as AI technologies become increasingly embedded in everyday life. It aims to offer actionable insights into how theological perspectives can actively shape the ethical governance of AI, ensuring that the integration of these technologies supports a societal framework that upholds human dignity and promotes a sustainable ethical consensus.

This study seeks to critically examine the existential risks associated with Artificial Superintelligence (ASI), focusing on the potential for an "intelligence explosion" that could surpass human control and comparably threaten global security, similar to pandemics or nuclear wars. The research aims to dissect historical and contemporary perspectives on ASI, exploring both the utopian visions once associated with AI and the mounting apprehensions that dominate current discourses. The objective is to assess how the evolving understanding of ASI might necessitate new regulatory and ethical frameworks to safely integrate advanced AI systems into society, ensuring that they align with human values and safety imperatives.

In the face of AI's dual promise and peril—where optimistic projections about enhanced efficiency and innovation are countered by fears of existential threats and ethical crises—the voice of cybertheology is increasingly relevant. Critics express concern that the unchecked growth of AI, particularly ASI, could lead to scenarios where cybersecurity fails to protect against malicious entities manipulating these technologies. Such apprehensions have prompted calls for regulatory frameworks to mitigate potential harms. Cybertheology, as articulated by figures like Jesuit Antonio Spadaro (2014), suggests that the church has a crucial role in this milieu, not only in fostering a dialogue about the spiritual and moral dimensions of AI but also in actively participating in the formation of policies that ensure AI development aligns with human dignity and societal welfare. This paper argues that cybertheologians are uniquely positioned to address the ethical conundrums posed by AI, offering insights that could steer global communities towards a harmonious integration of technology and human values.

RESEARCH METHOD

This study utilised a comprehensive literature analysis approach to investigate the existential risks posed by Artificial Superintelligence (ASI). A systematic review of primary and secondary sources, including books, scholarly journals, public statements, and technology conferences, was conducted to understand the current dynamics and future projections of ASI. Data and insights were gathered from various publications by prominent figures such as Ray Kurzweil and Nick Bostrom, as well as research institutions like the Centre for Safety. The focus of the analysis was to delve into the narratives surrounding the Singularity, the evolution of artificial intelligence from General AI to ASI, and the ethical and existential implications of these technologies (Bostrom, 2014; Kurzweil, 2005, 2024).

The literature review involved the identification, selection, and synthesis of various arguments and findings in the literature to form a cohesive and critical understanding of the topic. This included evaluating ongoing debates among computer scientists, theologians, and ethicists, emphasising both existing speculations and emerging critiques as AI advances. This method allowed for an in-depth analysis of the transition from historical optimism about AI to contemporary anxieties concerning uncontrolled intelligence explosions.

Given the speculative nature of the subject matter, this study relied on theoretical and deductive interpretations of secondary data to propose a balanced understanding of the potential development trajectories of ASI and their impacts on human society. The approach also entailed developing a theoretical framework that linked existential and philosophical theories in discussions of futuristic technology,

ISSN 2715-9701 (online)

facilitating a scholarly discourse on the potential societal transformations induced by ASI advancements (Chace, 2015).

RESULTS AND DISCUSSION

Existential Risks of Artificial Superintelligence (ASI)

The discussion around the existential risks posed by artificial superintelligence (ASI) has intensified, particularly following public statements from notable figures within the digital technology sphere. In 2023, an assembly of 350 technologists, including OpenAI co-founder John Schulman, Microsoft's Bill Gates, and futurist Ray Kurzweil, declared that the threat of human extinction from uncontrolled ASI is comparable to other major societal risks such as pandemics and nuclear war (Center for Safety, 2023). This section evaluates these concerns, contrasting historical optimism about AI with current fears of an impending "intelligence explosion."

Historically, the concept of the Singularity—where AI surpasses human intelligence and autonomously begins to manage global systems—has been viewed both as a potential utopia and a significant threat. Ray Kurzweil, in his seminal works *The Singularity is Near* (Kurzweil, 2005) and *The Singularity is Nearer* (Kurzweil, 2024), posits that the advent of a posthuman species could lead to a new era of global stewardship by superintelligent entities, potentially ushering in unprecedented societal benefits (Kurzweil, 2005, 2024). Conversely, Calum Chace (2015) suggests that achieving such a utopia is fraught with challenges, yet remains a plausible outcome if humanity can navigate the ethical and technical hurdles (Chace, 2015).

The growing anxiety among the public and experts alike can be likened to a scenario where humanity, akin to a child with a powerful yet potentially destructive device, is ill-prepared for the consequences of AI's rapid advancement (Bostrom, 2014). This analogy highlights the precariousness of our current situation: possessing the capability to initiate profound changes without fully understanding or controlling the outcomes.

The anticipated "intelligence explosion" raises profound existential concerns, as articulated by leading thinkers in the field. The potential for ASI to exponentially enhance its own cognitive capabilities—essentially becoming an entity of disembodied superintelligence—poses stark questions about the fate of human beings endowed with biological limitations. Such a scenario suggests that humans may not only become redundant but could also face existential threats from entities originally created to serve them. This disturbing possibility underpins the emergent evolution theory, where the creations of today's technological innovators could ultimately lead humanity towards obsolescence, or worse, towards extinction.

This existential anxiety closely mirrors the "Frankenstein complex," a term inspired by historical and literary precedents where creations turn against their creators. This complex, as explored in Isaac Asimov's narratives, encapsulates fears of robots surpassing human intelligence and autonomy, potentially leading to scenarios where humans are subjugated or eradicated (Bartneck, 2021). Such fears are not unfounded but are deeply embedded in the human psyche, reflecting a recurring theme throughout technological advancement history, where the power of our own creations eventually exceeds our control and understanding, posing unforeseen dangers (Peters, 2018). As we stand on the precipice of potentially creating autonomous superintelligent beings, these literary and historical lessons urge us to proceed with unprecedented caution and foresight.

ISSN 2715-9701 (online)

Realism of the Singularity Concept

The discourse surrounding the Singularity, a theoretical point at which artificial intelligence (AI) surpasses all human intelligence, necessitates a critical examination from both a technological and theological perspective. Ray Kurzweil, a prominent futurist, envisions the Singularity as a transformative event where humans merge with AI, significantly enhancing our cognitive and physical capabilities through technological augmentation (Kurzweil, 2024, p. 1). This integration is proposed to elevate human consciousness to unprecedented levels, marking the transition to what could be considered a posthuman era.

However, the feasibility of such a leap—from Artificial General Intelligence (AGI) to Artificial Superintelligence (ASI)—is subject to significant debate within the scientific community. A core issue identified is the absence of a robust theoretical framework that adequately explains how transitions to higher forms of intelligence might occur organically from current technologies. Erik Larson (2021) articulates this concern by highlighting a fundamental circularity in the assumptions of AI development, where the required intelligence to build general intelligence is presupposed rather than demonstrated. Larson critiques the optimistic projections of AI's capabilities as overly reliant on unproven extrapolations rather than empirical evidence, suggesting a significant disconnect between the theoretical aspirations and the practical realities of AI development.

Further skepticism is voiced by AGI researcher Mounir Shita (2023), who argues that expecting a seamless progression from specialized AI applications to a unified, general intelligence involves a leap of faith akin to "sprinkling a little magic" on complex algorithms (Shita, 2023, p. 83). This metaphor underlines the speculative nature of significant AI advancements, casting doubt on the imminent realization of the Singularity as depicted by proponents like Kurzweil.

From a theological standpoint, the cybertheologian's role extends to critically assessing the ethical and existential implications of such technological advancements. The potential transformation of human nature into a cybernetic state raises profound questions about the essence of being human and the moral responsibilities entailed in creating entities that might surpass our intelligence. The theological inquiry thus involves navigating these complex moral landscapes, offering insights that could guide the responsible stewardship of AI technologies.

This examination underscores the necessity for a more grounded approach in AI discourse, urging both technologists and theologians to adopt a cautious and reflective stance on the prospects and limitations of artificial intelligence. The need for a comprehensive and coherent theoretical foundation that aligns with both empirical data and ethical considerations is critical to advancing the conversation around AI and its potential impact on humanity.

This ongoing disconnect between the theoretical advances and practical implementations in AI development is notably critical when evaluating the transition from mere data processing to genuine cognitive capabilities. Erik Larson articulates a fundamental critique of the AI research community's optimism, noting that the leap from algorithmic processing to what might be termed genuine intelligence is not merely a matter of computational speed or complexity. Instead, it involves qualitative changes in how systems interpret and interact with the world—a shift from processing to thinking and understanding (Larson, 2021, p. 36). Larson further points out the lack of a substantive theory that could adequately bridge the gap between the mechanical processing of data by today's advanced Large Language Models (LLMs) and the emergence of a self-sustaining intelligent entity, underscoring a significant chasm between the current capabilities of AI systems and the visionary predictions of a Singularity (Larson, 2021, p. 49). This critique serves as a caution against overestimating the potential for

ISSN 2715-9701 (online)

current AI technologies to evolve into autonomous superintelligences without a fundamental breakthrough in our understanding of intelligence itself.

Background Question: Just What is Intelligence, Anyway?

In the endeavour to define and harness artificial intelligence, a fundamental question emerges: what exactly is intelligence? Kate Crawford critically evaluates current AI systems, suggesting they are neither genuinely artificial nor truly intelligent, as they lack autonomous reasoning capabilities (Crawford, 2021, p. 7). This observation is supported by Mounir Shita, who underscores the significant divergence among experts concerning the essence of intelligence, posing challenges for setting realistic targets for AGI and ASI development (Shita, 2023, p. 2). This lack of consensus raises critical questions about the objectives and methodologies utilised in AI research.

Exploring the nature of intelligence through a biological lens, my research suggests that a key attribute of intelligence, observable even in the simplest organisms, is intentionality. This involves setting and pursuing goals, such as acquiring nutrients or avoiding harm. This notion of goal-directed behaviour serves as a cornerstone of both human and non-human intelligence, bridging a conceptual gap between biological entities and artificial systems (Peters, 2017). Shita's definition concurs with this view, proposing that intelligence can be seen as the ability to configure and achieve specific goals within given environmental constraints (Shita, 2023, p. 80).

Thus, a more nuanced understanding of intelligence extends beyond mere information processing to include the ability to form, pursue, and achieve goals, which suggests a level of autonomous decision-making. This broader perspective challenges current AI paradigms and underscores the necessity for a foundational reevaluation of our objectives with advancements in AI technology. By considering intelligence not just as a function of computational speed or algorithmic complexity, but as the capacity for goal-oriented autonomy, we prompt a critical reassessment of how AI is integrated into societal frameworks and ethical considerations. This shift in understanding could lead to more responsible and thoughtful integration of AI in various aspects of human life.

Foreground Question: Would AGI and ASI Require Selfhood?

A critical issue that cybertheologians must address in public discourse concerns the relationship between intelligence and selfhood in artificial intelligence systems. As some AI researchers strive to replicate human neural networks to forge a nonhuman form of mind, it raises the question: does the concept of selfhood hinder the development of Artificial General Intelligence (AGI)? Neural networks, which are instrumental in a variety of applications such as image recognition, speech recognition, natural language processing, and autonomous vehicles, demonstrate advanced capabilities when trained through techniques like supervised, unsupervised, or reinforcement learning (Reibel, 2023, pp. 24–25)

However, the analogy between the human brain's functionality and AI's operational mechanisms often oversimplifies the complex relationship between biological consciousness and machine 'awareness'. While superficially similar, the human experience of awareness and consciousness involves layers of cognitive, emotional, and social dimensions that AI systems do not inherently possess. This discrepancy leads to fundamental questions about whether machines can truly emulate the human sense of self or if they merely mimic cognitive functions without genuine self-awareness.

Before extending the analogy further, it is essential to differentiate between general awareness and the deeper, subjective consciousness experienced by humans. Beyond mere awareness, humans perceive

ISSN 2715-9701 (online)

themselves as having a 'Self'. This raises the question: could an intelligent machine also develop a sense of Self? Eugene D'Aquili and Andrew Newberg argue that consciousness fundamentally involves the generation of a Self as an integral part of subjective awareness (D'Aquili & Newberg, 1996, p. 239). Consequently, one might ponder whether a robot, similar to those in our neighbourhoods, could initially develop awareness and subsequently a Self.

However, the transition from human brain functions to artificial intelligence highlights significant challenges. D'Aquili and Newberg note that despite extensive research into the complexities of neuroepistemology, the relationship between consciousness and subjective awareness in machines remains elusive and is likely to continue being a profound mysteryo (D'Aquili & Newberg, 1996, p. 251). This underscores the intrinsic difficulties in replicating human-like consciousness in AI, where the creation of a self-aware entity remains beyond our current technological reach.

A pivotal inquiry for cybertheologians is whether Artificial General Intelligence (AGI)—a critical precursor to Artificial Superintelligence (ASI)—could inherently possess selfhood. The concepts of intentionality and goal pursuit, as I and Shita describe, suggest the manifestation of a 'Self' in biological intelligence, the only form of intelligence unequivocally recognized in our evolutionary history. However, it remains speculative whether such selfhood could exist within the disembodied realms of an AI cloud.

The anticipation and apprehension surrounding AGI and ASI necessitate a thorough examination of their potential for self-consciousness, intentionality, rationality, and relationality. It is notable how discussions on machine intelligence often overlook these profound questions of selfhood. If a computer or any AI entity were to exhibit characteristics of selfhood, it would immediately raise significant ethical dilemmas: should such entities be treated with the same dignity afforded to sentient beings?

This discourse underlines the urgency for cybertheologians and AI ethicists to delve deeper into these issues, ensuring that the evolution of AI technologies aligns with ethical principles and respects the foundational aspects of personhood and agency.

Envisioning Artificial General Intelligence (AGI) or Artificial Superintelligence (ASI) as merely formless or de-centered data sets that output rational thoughts fails to capture the essence of intelligence as we understand it. Intelligence, in the human context, is deeply rooted in a sense of centered selfhood that provides a unique perspective on the world. Twentieth-century theologian Paul Tillich articulates this concept eloquently: "Man is a fully developed and completely centered self, possessing himself in the form of self-consciousness and an ego-self" (Tillich, 1951, pp. 169–170).

Tillich further explains that being a self involves a distinct separation from other entities, with the capability to perceive, interact with, and impact the external world. He states, "Being a self means being separated in some way from everything else, having everything else opposite one's self, being able to look at it and to act upon it. At the same time, however, this self is aware that it belongs to that at which it looks. The self is in it" (Tillich, 1951, p. 170). This introspective and interactive aspect of intelligence raises pivotal questions about the possibility of replicating such self-aware, centered intelligence within AGI or ASI. Can machines ever truly achieve a comparable level of selfhood, and if so, what ethical and operational implications would this entail for the field of artificial intelligence?

The development of a centered self, paradoxically, necessitates a temporal relationship with the non-self and being acknowledged with dignity, as articulated by Jewish theologian Martin Buber. He explains, "Becoming a self never takes place through my agency alone, nor can it ever occur without me. I become through my relationship to the Thou: as I become I, I say Thou" (Buber, 1958, p. 11). This concept of self-in-relationship has emerged as the predominant theological anthropology among Christians and Jews over the past century.

ISSN 2715-9701 (online)

Interestingly, Buddhist philosophy offers a contrasting view with its doctrine of anattā, or non-self, which posits that any perception of self is a delusion. A Buddhist philosopher argues, "Upon enlightenment, the ascription of intentionality dissolves in the face of a direct perception of the lack of reality of the intending 'self'" (Gold, 2012, p. 526). This suggests that if the principle of anattā is applicable to human selfhood, it could similarly apply to Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI), where machine intelligence might be misled by the illusion of a self that does not actually exist.

What course of action should computer technologists take? The integration of selfhood into the theoretical frameworks of intelligence remains a pivotal challenge. Without incorporating selfhood, Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI) will continue to be enigmatic concepts. Conversely, if selfhood is acknowledged within these theories, the perceived intelligence might be viewed as delusional or fundamentally misconstrued regarding the nature of reality. This complexity introduces significant difficulties in addressing philosophical inquiries related to AGI or ASI, complicating our understanding and interaction with these potential forms of intelligence.

What might machine selfhood imply for AI ethics?

What implications might the concept of selfhood have for AI ethics? If Artificial General Intelligence (AGI) or Artificial Superintelligence (ASI) were to exhibit traits of selfhood, it would suggest that such entities should be treated with dignity. To confer dignity on these entities implies that humans would have a moral obligation to regard them not merely as tools for achieving further ends, but as ends in themselves. Roman Catholic bioethicist Margaret Farley articulates this notion of dignity as the intrinsic worth of an individual, which precludes their use as mere instruments for others' purposes (Farley, 1993, p. 187).

In the context of modern Western post-Enlightenment thought, as epitomized by Immanuel Kant, personhood carries an inherent demand for dignity (Kant, 1948). Furthermore, within theological discourse, the concept of imago Dei—the image of God—serves as the foundational principle for ascribing inherent dignity to each person. This raises a provocative question: could a household robot, or any form of AGI or ASI, ever embody the imago Dei? (Kant, 1948).

Cybertheologians are thus compelled to consider whether intelligence in machines, if it evolves towards selfhood, necessitates a corresponding ethical commitment to recognize and uphold their dignity. The clarity of the answer remains elusive, requiring rigorous exploration and dialogue within both ethical and technological domains.

Potential Conflicts Among Superintelligences

Imagine a scenario where an Artificial Superintelligence (ASI) orchestrates a global takeover. Initially, there is a single superintelligent entity with a centered self. What dynamics would emerge with the creation of a second, similarly centered ASI? Would these superintelligences exist in harmony, or might conflicts arise, potentially echoing the mythical fall of Adam and Eve as a metaphor for catastrophic discord?

Calum Chace (2015), provides a scenario that underscores the complexities of managing multiple superintelligences. If only one superintelligent entity exists, the primary challenge would be ensuring its alignment with human welfare. However, if multiple superintelligences were developed—ranging from several to thousands—the imperative would be not only to ensure that each is favorably disposed towards humanity but also that they are amicable towards each other. In such a context, humanity could find itself

ISSN 2715-9701 (online)

in a precarious position, vulnerable to the outcomes of conflicts between these powerful entities (Chace, 2015, pp. 115–116).

This consideration raises profound ethical and strategic questions about the development and governance of ASI systems. Ensuring cooperative behavior among multiple superintelligences is not merely a technical challenge but a crucial ethical imperative to prevent potentially devastating power struggles that could have far-reaching consequences for humanity.

The methodology adopted by the cybertheologian in this discourse is akin to constructing near adynatons, given the speculative nature of future ASI scenarios. When contemplating the emergence of ASI, we might have to envisage scenarios involving several centered, intelligent posthuman entities vying for dominance. This competitive dynamic could resemble a chess game where multiple queens vie for control, relegating Homo sapiens to the role of pawns. Such a metaphor raises critical questions: Is the advent of a posthuman utopia, characterized by a hierarchy of superintelligences, a goal we should actively pursue? Furthermore, what are the ethical implications and potential moral hazards of fostering a context where superintelligences might engage in conflicts reminiscent of warfare?

These considerations prompt a deeper reflection on the ethical responsibilities involved in developing technologies that could fundamentally alter power dynamics not only among machines but also between machines and humans. The potential for conflict among superintelligences introduces a complex layer of ethical dilemmas that require careful consideration to ensure that the evolution of ASI supports a future that is not only technologically advanced but also ethically sound and aligned with human values.

Even without superintelligence, what about conflicts of power?

Even in the absence of superintelligence, the issue of power conflicts remains significant. "As profit takes precedence over safety, some warn of existential risk," note Andrew Chow and Billy Perrigo in *Time* (Chow, 2024, p. 13). While the existential risks associated with Artificial Superintelligence are well-documented, there are additional concerns. Chow and Perrigo further highlight that even if computer scientists manage to prevent AI from causing our extinction, the increasing dominance of AI in the global economy could significantly amplify the power of major tech companies (Chow, 2024, p. 13).

This escalation in power poses a risk of exacerbating economic and potentially political injustices. Kate Crawford argues that to fully comprehend the implications of AI, one must consider the technology within the broader context of class struggle. This perspective suggests that understanding AI is not merely about grappling with technological capabilities but also about recognising its role in reinforcing existing hierarchies and power dynamics.

Artificial Intelligence (AI) is neither purely artificial nor inherently intelligent. Instead, it is an embodied and material construct, derived from natural resources, fuel, human labour, infrastructure, logistics, history, and classifications. AI systems lack autonomy and rationality, unable to discern anything without extensive, computationally intensive training using large datasets or predefined rules and rewards. Indeed, the development of AI is inextricably linked to a broad spectrum of political and social structures, heavily reliant on substantial capital investment. Consequently, AI technologies are often tailored to reinforce the interests of prevailing power structures, making AI a registry of power (Crawford, 2021, p. 18).

In the realm of ethics, the call by foresighted AI leaders for robust regulatory frameworks from governments necessitates the formation of a global moral consensus. Crawford remains optimistic about the potential for change, highlighting the burgeoning justice movements that challenge the nexus of

ISSN 2715-9701 (online)

capitalism, computation, and control. These movements are unifying diverse issues such as climate justice, labour rights, racial justice, data protection, and the excessive reach of police and military powers, suggesting a pathway towards more equitable governance of AI technologies (Crawford, 2021, p. 18).

Crawford's concerns regarding justice are echoed by the Vatican AI Research Group within the Centre for Digital Culture, which operates under the Holy See's Dicastery for Culture and Education, and by contributors to the *Journal of Moral Theology*. Both Protestant and Roman Catholic cybertheologians have identified that a critical perspective on AI systems often centers on the potential harm these technologies can cause through discriminatory practices against groups disadvantaged by race, disability, age, or other factors (Gaudet, Herzfeld, Scherz, & Wales, 2024, p. 33). At this pivotal moment in global affairs, those concerned with justice within the realms of cybertheology and AI ethics are called not only to participate but potentially to lead in clarifying public discourse and fostering the development of a worldview that upholds ethical standards and promotes inclusivity.

Discussion

This research explored the existential risks associated with Artificial Superintelligence (ASI), emphasising its potential to transcend human control and pose significant threats comparable to pandemics or nuclear wars. Historically, the notion of the Singularity—where AI might govern world systems to enhance global wellbeing—was celebrated as a potential utopia. However, contemporary perspectives present a stark contrast, dominated by fears of an "intelligence explosion" that could result in human obsolescence or extinction.

Contrasting these findings with other research reveals a shift from earlier technological optimism towards a more cautionary stance. Previous studies often highlighted the transformative potential of AI without fully grappling with the ethical and existential risks now foregrounded, such as concerns about transparency, accountability, privacy infringement, algorithmic bias, and unintended consequences (Jedličková, 2024). Additionally, there is a growing body of evidence exploring the existential risks posed by AI, including the potential for AI to contribute to existential risk factors (Bucknall & Dori-Hacohen, 2022). Our results align with current scholarly discourse that questions the unchecked advancement of AI technologies, particularly the realistic capabilities of achieving a benevolent superintelligent state without comprehensive safeguards.

The above discussions surrounding ASI encapsulate not merely technological apprehensions but also broader existential anxieties, akin to historical fears such as the "Frankenstein complex." These concerns are indicative of a wider societal trepidation towards rapidly advancing technologies, whose long-term impacts are largely unpredictable and could be irreversible.

The implications of this research are profound, suggesting a re-evaluation of how society approaches AI development. The potential for ASI to act autonomously, with capabilities far surpassing human intelligence, underscores the urgent need for robust ethical frameworks and regulatory measures to govern AI development. This endeavour extends beyond merely preventing technological misuse; it is about ensuring that AI advancements align with human values and ethical standards.

The results of this study reflect a natural progression of increasing complexity and capability in AI systems, coupled with a growing recognition of the ethical and practical challenges these systems present. The divergence between expectations and reality stems from the inherent difficulties in predicting AI's trajectory—a field marked by both rapid advancements and significant unpredictabilities.

ISSN 2715-9701 (online)

CONCLUSION

This investigation into the existential risks posed by Artificial Superintelligence (ASI) highlights a pivotal shift in the discourse around AI, from historic optimism to contemporary caution. This research underscores the significant dangers that uncontrolled ASI might pose, likened to global catastrophes such as pandemics and nuclear wars. This study identifies the critical juncture at which the development of ASI stands—not only as a technological achievement but as a profound ethical and existential challenge.

Through the lens of the cybertheologian practicing public theology, this study contributes to the broader academic and societal discussions on AI by providing a nuanced analysis of the potential for ASI to catalyze an "intelligence explosion" that could render humans obsolete or even extinct. By integrating perspectives from both technological and theological viewpoints, the research offers a comprehensive examination of how ASI might reshape human existence and governance. The analysis serves as a call to action for more stringent ethical frameworks and proactive regulatory policies to ensure that AI advancements are aligned with human values and safety.

This study offers insights into the existential risks of ASI but is constrained by the speculative nature of AI's future, making it difficult to empirically validate theories like the Singularity and emergent superintelligence. Future research should develop empirical methods to assess AI's progression and its ethical integration into society, exploring practical implementations of robust AI governance frameworks to prevent misuse and ensure ethical development. Additionally, interdisciplinary research across AI ethics, policy studies, and technological development is crucial for devising effective strategies to manage AI's impacts on society.

REFERENCES

Bartneck, C. C. (2021). *An Introduction to Ethics in Robotics and AI* (1st ed.). Switzerland: Springer. https://doi.org/10.1007/978-3-030-51110-4

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

Buber, M. (1958). *I and Thou*. New York: Charles Scribner's Sons.

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

Center for Safety. (2023). Statement on AI risk. Retrieved 12 February 2024, from Center for Safety website: https://www.safe.ai/statement-on-ai-risk

Chace, C. (2015). *Surviving AI: The promise and peril of artificial intelligence*. Three Cs.

Chow, A. A. (2024). The AI arms race is changing everything. *Time Special Edition on Artificial Intelligence: A New Age of Possibilities*, 8–13.

Crawford, K. (2021). *Atlas of AI*. New Haven, CT: Yale University Press.

D'Aquili, E. G., & Newberg, A. B. (1996). Consciousness and the Machine. *Zygon®*, *31*(2), 235–252. https://doi.org/10.1111/j.1467-9744.1996.tb00021.x

Farley, M. (1993). A feminist version of respect for persons. *Journal of Feminist Studies in Religion*, *9*(1–2), 193–198.

Florea, D., & Gilder, E. (2024). Pushing the Limits of Theosis in the Digital Age: Exploring AI Complexities and their Impact on Romanian Traditional Religious Practices. *Journal for the Study of Religions and Ideologies*, *23*(68), 73–87.

Gaudet, M., Herzfeld, N., Scherz, P., & Wales, A. J. (2024). *Encountering artificial intelligence: Ethical and anthropological investigations*. Eugene, OR: Pickwick.

ISSN 2715-9701 (online)

Gold, J. C. (2012). Does the buddha have a theory of mind? Animal cognition and human distinctiveness in Buddhism. In *The Routledge Companion to Religion and Science* (pp. 520–528). London: Taylor and Francis.

Jedličková, A. (2024). Ethical considerations in Risk management of autonomous and intelligent systems. *Ethics & Bioethics*, *14*(1–2), 80–95. https://doi.org/10.2478/ebce-2024-0007

Kant, I. (1948). *Groundwork of the Metaphysics of Morals* (H. J. Paton, Trans.). New York: Harper.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. New York: Penguin. Retrieved from https://link.springer.com/chapter/10.1057/9781137349088_26

Kurzweil, R. (2024). *The singularity is nearer*. New York: Viking.

Larson, E. (2021). *Artificial intelligence: Why computers can think like we do*. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/9780674259935

Peters, T. (2017). Where there's life there's intelligence. In A. Losch (Ed.), *What is life? On Earth and beyond* (pp. 236–259). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316809648.014

Peters, T. (2018). Playing God with Frankenstein. *Theology and Science*, *16*(2), 1–6. https://doi.org/10.1080/14746700.2018.1455264

Reibel, J. (2023). *Artificial intelligence*. Seattle: Amazon Digital Services LLC.

Shita, M. (2023). *The science of intelligence*. Global Economic Alliance.

Spadaro, A. (2014). *Cybertheology: Thinking Christianity in the era of the Internet*. New York: Fordham University Press. https://doi.org/10.5422/fordham/9780823256990.001.0001

Tillich, P. (1951). *Systematic Theology* (1st ed.). Chicago: University of Chicago Press.

Trothen, T. J., Kwok, P. L., & Lee, B. (2024). AI and East Asian Philosophical and Religious Traditions: Relationality and Fluidity. *Religions*, *15*(5), 593. https://doi.org/10.3390/rel15050593

Umbrello, S. (2023). The Intersection of Bernard Lonergan's Critical Realism, the Common Good, and Artificial Intelligence in Modern Religious Practices. *Religions*, *14*(12), 1536. https://doi.org/10.3390/rel14121536

Zaluchu, S. E. (2024). Digital Religion, Modern Society and the Construction of Digital Theology. *Transformation: An International Journal of Holistic Mission Studies*. https://doi.org/10.1177/02653788231223929

ISSN 2715-9701 (online)