Threat Posed by Artificial Intelligence to Humanity in Future

S. Sai Naaresh
I MBA, School of Management
Dwaraka Doss Govardhan Doss Vaishnav College, Chennai, Tamil Nadu

Krishanu G Aadityan
I MBA, School of Management
Dwaraka Doss Govardhan Doss Vaishnav College, Chennai, Tamil Nadu

Abstract
Artificial Intelligence (AI) emerges as a transformative force set to reshape industries and societies, yet it also presents significant threats to humanity. This study explores the multifaceted risks associated with AI, analyzing ethical, economic, and governance dimensions. Ethical concerns arise from algorithmic bias, privacy breaches, and the potential misuse of autonomous weapons. Economically, AI-driven automation poses a threat of widespread job displacement, demanding proactive measures to address societal inequality. Additionally, AI's security vulnerabilities jeopardize cyber security and national safety, necessitating robust protective measures. Moreover, the prospect of AI surpassing human intelligence raises existential questions, prompting the exploration of strategies for its safe development. Tackling these challenges requires coordinated international efforts, including regulatory reforms and ethical frameworks. By examining AI threats comprehensively, this research aims to guide policymakers, technologists, and society toward responsible AI deployment for humanity's well-being in the AI-driven era.
Keywords: AI-Driven Era, Artificial Intelligence, Ethics, Existential Threat, Governance, Security

Introduction

Artificial Intelligence (AI) stands at the forefront of technological advancement, poised to reshape industries and societies in profound ways. As AI proliferates across various domains, from autonomous vehicles to healthcare systems, its transformative potential is undeniable. However, amidst the promises of progress lie significant ethical, economic, and governance challenges that warrant careful consideration.

Ethically, AI raises concerns about algorithmic bias, privacy infringements, and the potential erosion of human values and rights. Economically, the specter of widespread job displacement looms large as AI-driven automation reshapes traditional employment models. Moreover, AI's security vulnerabilities pose significant risks to cybersecurity and national security, demanding robust strategies for safeguarding against malicious exploitation.

Beyond immediate concerns, the prospect of AI surpassing human intelligence raises existential questions about the future of humanity. Addressing these challenges necessitates a coordinated international

effort, encompassing regulatory reforms, ethical frameworks, and collaborative governance mechanisms.

This research endeavors to comprehensively examine the multifaceted dimensions of the risks posed by AI, seeking to inform policymakers, technologists, and society at large about the imperative of responsible AI development and deployment. By navigating the complexities of AI's impact from diverse perspectives, we aim to safeguard the well-being of humanity in the AI-driven era.

Review of Literature

Andrzej Nowak, Paul Lukowicz, and Pawel Horodecki (2018): In their research, they delve into the transformative impact of recent advancements in artificial intelligence (AI) on humanity's trajectory. They acknowledge the historical significance of Homo sapiens' cognitive capacity, which has long distinguished humans as the sole species capable of higher cognitive functions. However, the researchers highlight the paradigm shift brought about by AI's rapid progress, reaching a stage where it not only rivals but also surpasses human performance on various cognition-related tasks.

The researchers illustrate AI's prowess through notable achievements, including victories in complex games like Go and chess against human champions. Beyond gaming, AI demonstrates remarkable efficiency in practical domains such as speech and object recognition, autonomous driving, intelligent tutoring, decision support systems, and data analysis. These advancements prompt a fundamental question: has humanity's historical monopoly on intelligence been challenged?

Their research underscores the urgent need to address this question and grapple with the ethical, societal, and existential implications of AI's ascent. As AI continues to integrate with human cognitive abilities, critical considerations emerge regarding ethics, governance, and the future trajectory of society. The researchers emphasize the importance of developing robust ethical frameworks and regulatory mechanisms to guide responsible AI development and deployment. Ultimately, their work highlights the imperative of understanding and navigating the profound implications of AI's rise for the broader landscape of humanity.

Veysel Bozkurt and Dogan Gursoy (2023): In their study, they investigate the nuanced perceptions of artificial intelligence (AI) among a sample of 1443 participants. Their research aims to discern whether individuals view AI primarily as a threat or an opportunity, while also examining how demographic factors, religious commitment, trust in science, beliefs in conspiracy theories, and previous experiences with AI shape these perceptions.

The study reveals a complex landscape of attitudes towards AI. While a majority of respondents (63%) believe that AI will simplify life and increase efficiency (62%), indicating an optimistic outlook, a significant portion express concerns about its potential negative consequences. Specifically, over half of the participants (52%) worry that AI will lead to increased unemployment, and nearly half (47%) fear it will exacerbate social inequalities. Alarmingly, around 21% of respondents express the belief that AI could ultimately pose a threat to humanity.

Furthermore, the research identifies several factors that significantly influence individuals' perceptions of AI. Respondents' age, gender, occupation, religious commitment, beliefs in conspiracy theories, and previous experiences with AI play pivotal roles in shaping their views of AI as both an opportunity and a threat.

Overall, the findings suggest a paradoxical perception of AI as a double-edged sword – viewed simultaneously as a tool for advancement and a potential source of peril. This duality underscores a notable confusion among respondents regarding the implications of AI. Bozkurt and Gursoy's study sheds light on the multifaceted nature of public perceptions towards AI, highlighting the need for nuanced understanding and informed discourse in navigating the challenges and opportunities presented by this technology.

Kevin Warwick (2018): In this chapter authored by Kevin Warwick, the debate surrounding the potential threat posed by artificial intelligence (AI) to humanity is examined. While some individuals assert that AI will never be a threat under any circumstances, Warwick explores the arguments put forth by proponents of this viewpoint, with specific attention given to the perspectives of Bundy, Searle, and Penrose.

The chapter critically assesses the rationale behind such beliefs, acknowledging the concerns raised by these scholars. However, an opposing argument is presented, emphasizing the logical and pragmatic reasons that indicate AI, as currently understood and utilized, does indeed harbor the potential to pose a significant threat to humanity.

Warwick's analysis underscores the importance of considering the safety implications of AI and draws attention to critical lessons for managing and mitigating potential risks. By examining the arguments for and against AI as a threat, the chapter provides valuable insights into how society can navigate the challenges posed by increasingly advanced AI technologies.

Throughout the chapter, Warwick adopts a broad approach, treating terms such as computers, machines, and robots as largely interchangeable, without delving into detailed distinctions between them. This approach allows for a comprehensive examination of the overarching concerns related to AI's impact on humanity and the imperative for proactive management and regulation in this rapidly evolving field.

Andrea Lavazza and Murilo Vilaca (2024): They present a thought-provoking exploration in their study regarding the concept of human extinction and its implications. While human extinction is universally viewed as undesirable, the authors acknowledge that some scholars have proposed it as a potential solution to the challenges facing Earth, arguing that it could alleviate moral evil and suffering caused by humans. However, Lavazza and Vilaca assert that humans possess inherent value as sentient, conscious, and rational beings, and thus advocate for preserving humanity from extinction.

Despite this stance, the authors recognize the existence of severe threats such as climate change and incurable viruses that could potentially push humanity to the brink of extinction. In the event of such a catastrophe, the authors suggest envisioning a successor to humanity capable of preserving and transmitting its intrinsic value. One proposed avenue is the creation of humanoid robots endowed with artificial intelligence (AI) powered by machine learning, replicating key human characteristics through imitation.

However, a significant challenge arises in determining the characteristics necessary for the success of these successors. Lavazza and Vilaca propose that an algorithm could be entrusted with this selection process, offering a methodical approach to decision-making at both the societal and individual levels. They argue that algorithmic selection may offer advantages over traditional decision-making methods, providing a rational and impartial means of identifying traits conducive to thriving.

In this way, the authors suggest that contemplating the possibility of human extinction can serve as a catalyst for identifying innovative solutions applicable to contemporary challenges. By engaging in this speculative exercise, Lavazza and Vilaca encourage critical reflection on the fundamental nature of humanity and the potential pathways for ensuring its continuity and flourishing in the face of existential threats.

Izabela Oleksiewicz (2022): In her article, she addresses the evolving landscape shaped by advancing technologies, particularly the accelerating implementation of artificial intelligence (AI) systems. Concurrently, she highlights the growing significance of "soft skills" sought by employers in prospective employees, emphasizing qualities such as empathy, communication, ingenuity, intelligence, and humanism. Oleksiewicz underscores the critical role of these attributes

in fostering effective teamwork and shaping business perspectives amidst the increasing utilization of automatons and robots in various tasks.

The article aims to delve into the research problem concerning social robots and the potential legal status of artificial intelligence in the future. It examines the distinctions between artificial intelligence and artificial consciousness, seeking to demystify these concepts. Additionally, Oleksiewicz discusses the societal challenges posed by AI and the pressing need for rapid legal adjustments to accommodate its advancements

The analysis extends to the current legal status of artificial intelligence within the European Union (EU), considering the transformative impact of cyberspace and the internet on human life. Oleksiewicz acknowledges the benefits brought by these technological innovations but also highlights the associated risks, which are inherent to human development. She suggests that autonomous systems leveraging AI might offer solutions to these challenges, but underscores the importance of addressing cybersecurity concerns.

Ultimately, Oleksiewicz argues that the security of cyberspace hinges on various factors, some of which may be beyond our control, while others may be influenced by human actions and decisions. Through her analysis, she prompts reflection on the complexities surrounding AI, its legal implications, and the imperative of responsible governance to navigate the evolving technological landscape.

Ha Hwang and Min-Hye Park (2020): In their study, they explore the ongoing changes in society driven by Artificial Intelligence (AI) and the efforts to mitigate potential risks associated with its deployment. They highlight the effectiveness of AI while acknowledging the need for responsible usage and the emergence of discussions surrounding the ethical implications of AI, leading to the creation of the AI Charter of Ethics (AICE).

The study focuses on how Korean society is addressing potential threats posed by AI by examining various AICEs in the Republic of Korea. Initially, the researchers identify seven distinct AI threats, which they categorize into three groups: AI's value judgment, malicious use of AI, and human alienation. They then analyze Korea's seven AICEs, extracting fourteen topics categorized into three overarching themes: protection of social values, AI control, and fostering digital citizenship.

Through their analysis, Hwang and Park assess the alignment between the identified AI threats and the responses outlined in Korea's AICEs. They find that while Korea has made strides in addressing certain threats, such as the development of lethal AI weapons, there are gaps in responses to other critical concerns, particularly regarding AI's potential to usurp human occupations (jobs). Moreover, they note that the realization of provisions related to lethal AI weapons may face challenges due to escalating competition among military powers.

Overall, the study sheds light on the evolving discourse surrounding AI ethics and the efforts to establish frameworks for responsible AI governance in Korea. By evaluating the alignment between perceived AI threats and societal responses, Hwang and Park provide valuable insights into areas where further attention and action may be necessary to ensure the responsible and beneficial integration of AI into society.

Ikedinachi Ayodele Wogu, Sanjay Misra, Patrick Assibong, Adewole Adewumi, Robertas Damasevicius, and Rytis Maskeliunas (2018): In their study, they address the growing concern among scholars regarding the potential consequences of Artificial Intelligent Machines (AIMs) acquiring the capacity for intentionality. They highlight the division among scholars, with some expressing apprehension about the existential threats posed by AIMs, while others, notably Katja Grace and Kevin Drum, have sounded alarm bells about the impending job apocalypse and other dangers.

The authors draw on the Marxian Alienation Theory to analyze the degree of extinction and existential threat facing humanity as AIMs continue to advance. They also explore the politicization

of predictions regarding the timeline for when AIMs might take over human jobs, employing ex-post facto research methodology and Derrida's reconstructive and deconstructive analytical method.

The paper underscores the urgent need for policymakers and scholars to address the impending era of mass joblessness, which is identified as one of the greatest challenges facing progressive governments and thinkers. The authors advocate for the swift identification of pathways to distribute the gains of robot labor in a manner that mitigates the threat to humanity. They emphasize the importance of devising policies and strategies that ensure AIM operations are not detrimental to human well-being.

In conclusion, Wogu, Misra, Assibong, Adewumi, Damasevicius, and Maskeliunas call for concerted efforts to address the ethical, social, and economic implications of AIM advancement. They stress the necessity of proactive measures to safeguard against the potential negative consequences while harnessing the benefits of AI technology for the betterment of society.

Research Objectives
Primary Objective
- To analyze and identify the specific threats posed by Artificial Intelligence (AI) to humanity in the future, encompassing ethical, economic, and societal dimensions.
- To propose actionable strategies and recommendations for mitigating the identified threats and ensuring responsible AI development and deployment to safeguard the well-being of humanity.

Secondary Objective
- To examine the historical and current trends in AI development and deployment, highlighting key advancements and potential risk factors that may amplify threats to humanity.
- To investigate the ethical, economic, and governance implications of AI technologies, including algorithmic bias, job displacement, privacy infringements, and existential risks, to gain a comprehensive understanding of the multifaceted challenges posed by AI.

The Evolution of AI: From Classical to Super Intelligence
The evolution of Artificial Intelligence (AI) has been marked by significant milestones, each representing a leap forward in computational capabilities and cognitive functions. Classical AI, rooted in rule-based systems and symbolic reasoning, laid the groundwork for subsequent advancements. The emergence of Artificial Narrow Intelligence (ANI) marked a significant shift, as AI systems became capable of performing specific tasks with precision, albeit within narrow domains. This specialization paved the way for more sophisticated AI architectures, leading to the development of Artificial General Intelligence (AGI).

AGI represents a critical milestone in AI research, as it denotes the ability of AI systems to perform tasks across multiple domains with human-like intelligence. Unlike ANI, which is constrained to specific tasks, AGI possesses the capacity for abstract reasoning, problem-solving, and learning from experience. The pursuit of AGI has been a longstanding goal in AI research, driven by the desire to create machines capable of emulating human-level intelligence.

The ultimate goal of AI research is the attainment of Artificial Super Intelligence (ASI), where AI systems surpass human intelligence across all domains. ASI represents a paradigm shift in AI capabilities, as it encompasses not only cognitive abilities but also creativity, intuition, and emotional intelligence. Achieving ASI has profound implications for society, as it promises to revolutionize industries, accelerate scientific discovery, and reshape the fabric of human existence.

Recent advancements in AI, such as Google's LLaMda2 and the development of Chat GPT-3 and GPT-4, signify significant progress towards ASI. These systems demonstrate unprecedented

levels of language understanding, natural language generation, and problem-solving abilities, bringing ASI within reach in the foreseeable future. However, the rapid pace of AI development also raises ethical, societal, and existential questions that must be addressed as we approach the era of superintelligent machines.

Implications of ASI: Accelerated Innovation and Potential Disruption

The imminent arrival of Artificial Super Intelligence (ASI) heralds a new era of accelerated innovation and potential disruption across various sectors of society. ASI's unparalleled cognitive abilities promise to revolutionize industries, catalyze scientific breakthroughs, and fundamentally transform the way we live and work. From healthcare and transportation to finance and entertainment, ASI has the potential to drive unprecedented levels of productivity, efficiency, and creativity.

In the realm of medicine, ASI holds the promise of revolutionizing healthcare delivery, diagnosis, and treatment. With its ability to process vast amounts of medical data, ASI can accelerate drug discovery, personalize treatment regimens, and predict disease outbreaks with unprecedented accuracy. Similarly, in the energy sector, ASI-powered algorithms can optimize energy production, distribution, and consumption, leading to more efficient use of resources and reduced environmental impact.

However, alongside its promises of innovation and progress, ASI also presents significant challenges and potential disruptions. One of the most pressing concerns is the impact of AI-driven automation on employment and labor markets. As ASI becomes increasingly capable of performing tasks traditionally carried out by humans, there is a risk of widespread job displacement and economic dislocation. Industries reliant on human labor, such as manufacturing and transportation, may face significant disruptions, leading to social upheaval and income inequality.

Moreover, the rapid proliferation of ASI raises ethical and societal concerns regarding its deployment and impact on human society. Questions surrounding AI ethics, bias, privacy, and accountability become increasingly complex in the era of superintelligent machines. Ensuring that ASI is developed and deployed responsibly, with due consideration for its societal implications, requires concerted efforts from policymakers, technologists, and society at large.

While ASI holds tremendous potential to drive innovation and progress, its deployment must be accompanied by careful consideration of its societal impact and ethical implications. By addressing these challenges proactively, we can harness the transformative power of ASI to create a future that benefits all of humanity.

Diverging Paths: Dystopian Nightmares and Utopian Visions

As Artificial Super Intelligence (ASI) looms closer, diverging narratives emerge regarding its societal impact, reflecting contrasting visions of the future. Pessimistic outlooks, articulated by figures like Elon Musk, Bill Joy, and Stephen Hawking, warn of a dystopian future where humanity becomes obsolete in a computer-controlled world. In this bleak scenario, ASI's unmatched cognitive abilities lead to the subjugation of human autonomy and the erosion of human values and rights. The prospect of a world dominated by superintelligent machines, devoid of human agency and purpose, raises profound existential questions about the future of humanity.

Conversely, optimistic voices envision a utopian society guided by the transformative potential of ASI. Drawing inspiration from the Kardashev scale, which measures a civilization's level of technological advancement, proponents of this vision imagine a future where ASI enables unprecedented progress and prosperity for humanity. In this utopian narrative, ASI serves as a catalyst for solving humanity's most pressing challenges, from poverty and disease to environmental degradation and resource scarcity. Through collaboration with superintelligent machines, humans unlock new frontiers of knowledge, creativity, and exploration, leading to a golden age of human

civilization. As society grapples with these diverging paths, it confronts fundamental questions about the ethical, societal, and existential implications of ASI. Balancing the potential benefits and risks of ASI requires careful consideration of its ethical and governance dimensions. Ensuring that ASI is developed and deployed in a responsible and equitable manner requires proactive engagement from policymakers, technologists, and society at large. By fostering dialogue and collaboration across diverse stakeholders, we can navigate the uncertainties of the AI-driven future and chart a course towards a future that safeguards the well-being and dignity of humanity.

Navigating the Future: Ethical Considerations and Collaborative Endeavors

As society confronts the transformative potential of Artificial Super Intelligence (ASI), navigating the future requires careful consideration of its ethical, societal, and existential implications. Ethical considerations surrounding AI governance, accountability, transparency, and bias become increasingly complex in the era of superintelligent machines. Balancing the benefits and risks of ASI demands proactive engagement from policymakers, technologists, and society at large to ensure responsible development and deployment.

Collaborative endeavors are essential in addressing the multifaceted challenges posed by ASI. Initiatives like the San Francisco Superior Court's ruling on breach of contract underscore the importance of legal frameworks in regulating AI technologies and ensuring accountability for their actions. By fostering collaboration between governments, industry stakeholders, and civil society organizations, we can develop ethical guidelines and governance mechanisms that promote the responsible and equitable use of ASI.

AI should be treated like gun culture, where everyone has access to it but with strict regulations and safeguards in place to prevent misuse. OpenAI's decision to license GPT-3 to Microsoft highlights the importance of responsible technology transfer and collaboration in advancing AI research and innovation. By promoting transparency, accountability, and inclusivity in AI development and deployment, we can harness the transformative potential of ASI to create a future that benefits all of humanity.

Navigating the future of AI requires a collective effort to address ethical, societal, and governance challenges. By fostering dialogue, collaboration, and responsible stewardship of AI technologies, we can ensure that ASI serves as a force for good, advancing human progress and well-being in the AI-driven era.

Exploring the Fermi Paradox and Kardashev Scale in Understanding the Perils and Prospects of AI Development

Fermi Paradox

The Fermi paradox poses a compelling question: if intelligent life is common in the universe, why haven't we detected any signs of it? Named after physicist Enrico Fermi, who famously asked "Where is everybody?" during a conversation about extraterrestrial civilizations, the paradox underscores the apparent contradiction between the high probability of extraterrestrial life and the lack of observable evidence for it.

In the context of the topic "Threat posed by AI to Humanity," the Fermi paradox prompts reflection on the potential consequences of technological advancement and the implications for the long-term survival of intelligent civilizations. As humanity continues to develop increasingly sophisticated technologies, including Artificial Intelligence (AI), we may inadvertently become participants in our own version of the Fermi paradox.

One interpretation of the Fermi paradox suggests that intelligent civilizations may face existential threats, such as those posed by advanced technologies like AI, that ultimately lead to their demise

or isolation. This notion aligns with concerns raised by prominent figures like Elon Musk, Stephen Hawking, and others, who warn of the potential risks associated with the development of superintelligent AI. If unchecked, AI could pose significant threats to human civilization, ranging from economic disruption and job displacement to more existential risks such as loss of control over AI systems or the emergence of super intelligent AI surpassing human capabilities.

Moreover, the Fermi paradox highlights the importance of responsible stewardship of AI technologies to mitigate the potential risks they pose. By addressing ethical, societal, and governance challenges associated with AI development and deployment, we can strive to prevent unintended consequences and ensure that AI serves as a force for positive transformation rather than existential threat.

The Fermi paradox underscores the need for careful consideration of the potential consequences of technological advancement, particularly in the context of AI development. By exploring the relationship between the Fermi paradox and the threats posed by AI to humanity, we can gain valuable insights into the challenges and opportunities associated with the responsible use of AI technologies in shaping the future of our civilization.

Kardashev Scale

The Kardashev Scale is a theoretical framework proposed by Russian astrophysicist Nikolai Kardashev in 1964 to categorize and measure the level of technological advancement of civilizations based on their energy consumption. The scale consists of three main types of civilizations, each characterized by their ability to harness and utilize energy on increasingly larger scales.

Type I civilizations, also known as planetary civilizations, have achieved mastery over the resources of their home planet, including renewable energy sources such as solar, wind, and hydroelectric power. They are capable of harnessing and controlling the entire energy output of their planet, enabling them to sustain a high level of technological development and societal complexity.

Type II civilizations, known as stellar civilizations, have advanced to the point where they can harness the energy output of an entire star, typically through technologies like Dyson spheres or stellar engines. This level of energy mastery allows them to achieve even greater levels of technological sophistication and to colonize entire star systems.

Type III civilizations, referred to as galactic civilizations, have attained the ability to harness the energy output of an entire galaxy. This level of energy mastery is orders of magnitude beyond that of Type II civilizations and would require the ability to manipulate vast cosmic structures and phenomena.

In the context of the topic "Threat posed by AI to Humanity," the Kardashev Scale offers a framework for understanding the potential implications of advanced technologies, including artificial intelligence (AI), on the trajectory of human civilization. As humanity progresses along the Kardashev Scale, our ability to harness and utilize energy becomes increasingly intertwined with our technological capabilities and societal development.

For example, the development of advanced AI technologies could play a pivotal role in advancing humanity towards Type I or even Type II status on the Kardashev Scale by enabling more efficient energy management, resource allocation, and technological innovation. However, it also raises concerns about the potential risks and challenges associated with the proliferation of AI, particularly as we approach Type I status.

Moreover, the Kardashev Scale prompts reflection on the long-term implications of AI development for the future trajectory of human civilization. As we strive to harness the power of AI to address pressing global challenges, such as climate change, economic inequality, and healthcare, we must also consider the ethical, societal, and existential risks associated with the unchecked proliferation of AI technologies.

The Kardashev Scale provides a useful framework for contextualizing the relationship between technological advancement, energy consumption, and the potential threats posed by AI to humanity. By considering the implications of AI development within the broader context of our progress along the Kardashev Scale, we can gain valuable insights into the challenges and opportunities associated with shaping the future of our civilization.

Results and Discussion

Projecting the results of research on the threat posed by AI to humanity entails considering various scenarios and their potential implications for society. Here's a brief projection and discussion of the possible outcomes:

Scenario 1: Responsible AI Development and Governance

In this scenario, policymakers, technologists, and society collaborate to develop and deploy AI technologies responsibly.

Ethical guidelines, regulatory frameworks, and governance mechanisms are implemented to mitigate risks and ensure the beneficial use of AI.

The result is the continued advancement of AI technologies that enhance productivity, improve quality of life, and address societal challenges while minimizing negative impacts.

Scenario 2: Unchecked AI Proliferation and Misuse

Alternatively, if AI development proceeds unchecked without adequate safeguards, it could lead to unintended consequences and societal disruption.

Job displacement, economic inequality, privacy violations, and algorithmic bias become widespread, exacerbating social tensions and widening disparities.

The potential for AI to surpass human intelligence (ASI) raises existential risks, including loss of control over AI systems and threats to human autonomy and dignity.

Scenario 3: Dual-Use AI and Geopolitical Competition

A third scenario involves the dual-use nature of AI technologies, where advancements intended for beneficial purposes are also weaponized for military or malicious ends.

Geopolitical competition intensifies as nations vie for AI supremacy, leading to an arms race in AI-driven technologies.

The result is increased cyber warfare, autonomous weapons proliferation, and heightened global instability, posing significant threats to international security and stability.

Implications

The projected results underscore the importance of proactive measures to ensure responsible AI development and governance.

Collaboration between governments, industry stakeholders, and civil society is essential to establish ethical guidelines, regulatory frameworks, and governance mechanisms that prioritize human well-being and safety.

Addressing the ethical, societal, and existential risks associated with AI requires a multidisciplinary approach, encompassing technology, ethics, law, economics, and governance.

Failure to address these challenges could lead to dire consequences, including societal upheaval, economic disruption, loss of human autonomy, and even existential threats to humanity.

By considering the potential outcomes and implications of AI development, stakeholders can make informed decisions and take proactive steps to shape a future where AI serves as a force for positive transformation rather than a threat to humanity.

Research Gaps

The research on the threat posed by AI to humanity has made significant strides in identifying and analyzing various risks and challenges associated with AI development and deployment. However, there remain several notable research gaps that warrant further investigation:

Long-term Societal Impacts: While much attention has been given to the immediate ethical, economic, and governance challenges posed by AI, there is a need for more research on the long-term societal impacts of AI on human civilization. This includes understanding how AI may reshape social structures, cultural norms, and human relationships over time.

Existential Risks: Despite growing recognition of the potential existential risks associated with advanced AI, there is still a lack of consensus on how to effectively mitigate these risks. Further research is needed to explore strategies for safeguarding humanity against catastrophic outcomes, such as the emergence of Artificial Super Intelligence (ASI) surpassing human control.

Governance and Regulation: The governance and regulation of AI technologies remain underdeveloped and fragmented, with significant gaps in international coordination and enforcement. More research is needed to explore alternative governance models and regulatory frameworks that can effectively address the ethical, legal, and societal implications of AI.

Public Perception and Engagement: Understanding public perceptions, attitudes, and concerns about AI is crucial for shaping effective policy responses and fostering societal acceptance of AI technologies. Further research is needed to examine public perceptions of AI, as well as strategies for promoting AI literacy and public engagement.

Conclusion

While research on the threat posed by AI to humanity has made significant progress, there are still important gaps that need to be addressed. By focusing on these research gaps, scholars can contribute valuable insights to inform policy decisions, guide technological development, and ensure that AI serves the best interests of humanity. Ultimately, addressing these gaps will be essential for navigating the complexities of the AI-driven future and shaping a world where AI benefits all of humanity.

References

1. https://ieeexplore.ieee.org/abstract/document/8558761
2. https://www.tandfonline.com/doi/abs/10.1080/10447318.2023.2297114
3. https://www.taylorfrancis.com/chapters/edit/10.1201/9781351251389-28/smart-machines-threat-humanity-kevin-warwick
4. https://link.springer.com/article/10.1007/s13347-024-00706-2
5. https://www.ceeol.com/search/article-detail?id=1111714
6. https://openurl.ebsco.com/EPDB%3Agcd%3A13%3A6018349/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A142995659&crl=c
7. https://link.springer.com/chapter/10.1007/978-3-319-95171-3_18
8. Nowak, Andrzej, et al. "Assessing Artificial Intelligence for Humanity: Will AI be the Our Biggest Ever Advance? or the Biggest Threat." Institute of Electrical and Electronic Engineers, 2018.
9. Bozkurt, Veysel, and Dogan Gursoy. "The Artificial Intelligence Paradox: Opportunity or Threat for Humanity?" tandfonline, 2023.
10. Warwick, Kevin. "Smart Machines ARE a Threat to Humanity." Chapman and Hall/CRC, 2018.

11. Lavazza, Andrea, and Murilo Vilaca. "Human Extinction and AI: What We Can Learn from the Ultimate Threat." Springer, 2024.

12. Oleksiewicz, Izabela. "Artificial intelligence versus human – a threat or a necessity of evolution?" Wydawnictwa Uniwersytetu Warszawskiego, Issue No: 3, 55-69, 2022.

13. Hwang, Ha, and Min-Hye Park. "The Threat of AI and Our Response: The AI Charter of Ethics in South Korea." Asian Journal of Innovation & Policy, 2020, Vol 9, Issue 1, p56, 2020.

14. Wogu, Ikedinachi Ayodele, et al. "A Critical Review of the Politics of Artificial Intelligent Machines, Alienation and the Existential Risk Threat to America's Labour Force." Lecture Notes in Computer Science ((LNTCS,volume 10963)), 2018.