Review

Not peer-reviewed version

Navigating Artificial General Intelligence (AGI): Societal Implications, Ethical Considerations, and Governance Strategies

Dileesh Chandra Bikkasani [*]

Posted Date: 21 October 2024

doi: 10.20944/preprints202407.1573.v3

Keywords: Artificial General Intelligence (AGI); AGI Ethics; AGI Governance; Security; Societal Impact

Review

Navigating Artificial General Intelligence (AGI): Societal Implications, Ethical Considerations, and Governance Strategies

Dileesh Chandra Bikkasani

University of Bridgeport, USA: dbikkasa@my.bridgeport.edu

Abstract: Artificial General Intelligence (AGI) represents a pivotal advancement in AI with far-reaching implications across technological, ethical, and societal domains. This paper addresses the following: (1) an in-depth assessment of AGI's transformative potential across different sectors and its multifaceted implications, including significant financial impacts like workforce disruption, income inequality, productivity gains, and potential systemic risks; (2) an examination of critical ethical considerations, including transparency and accountability, complex ethical dilemmas and societal impact; (3) a detailed analysis of privacy, legal and policy implications, particularly in intellectual property and liability, and (4) a proposed governance framework to ensure responsible AGI development and deployment. Additionally, the paper explores and addresses AGI's political implications, including national security and potential misuse. By analyzing and considering computer science, philosophy, economics, and policy perspectives, we offer a multidisciplinary view of AGI's challenges and opportunities, advocating for proactive measures to align AGI development with human values and societal interests.

Keywords: artificial general intelligence (AGI); AGI ethics; AGI governance; security; societal impact

1. Introduction

"Let us define an ultra-intelligent machine as one that could surpass human capabilities in all domains. Since "designing a machine" is one of those domains, it becomes a cycle of self-improvement called the Intelligence Explosion. The human who oversaw all this would be left far behind. Making AGI the final invention we will ever have made" (Good, 1966). AGI refers to highly autonomous systems that can surpass human performance in most economically valuable tasks. This paper argues that while the development of AGI holds transformative potential, it presents unprecedented risks that require robust governance frameworks to ensure alignment with human values. Unlike narrow AI, which is tailored for specific functions, AGI would possess human-like intelligence across a broad spectrum of cognitive abilities. This includes reasoning, problem-solving, planning, learning, natural language understanding, and adapting to new situations. While current AI excels in limited domains, AGI aims to match or exceed human-level performance across virtually any cognitive task. Achieving AGI involves creating systems capable of flexibly understanding, learning, and applying knowledge across different domains, ultimately transforming industries and society.

The development of AGI is a topic of significant debate within the computing and AI communities. Contrary to some perspectives suggesting AGI's inevitability, this paper contends that the uncertainties surrounding AGI's feasibility, timeline, and societal impact underscore the need for critical examination and proactive governance. Disagreements persist about whether AGI is achievable in the near term, with scholars like Yoshua Bengio and Stuart Russell asserting that current AI technologies remain far from achieving true general intelligence; there is also significant disagreement regarding the feasibility and timeline (Mueller, 2024). Additionally, various definitions of AGI exist, ranging from broad conceptualizations of human-like cognition to more restrictive criteria involving specific functional benchmarks. This divergence reflects a broader lack of consensus on AGI's precise nature and requirements.

Humans dominate the earth mainly due to our cognitive capabilities, such as language, reasoning, social interactions, energy, and tool usage. The development and expansion of the human brain has been a gradual process spanning millions of years (Defelipe, 2011). AI has evolved at an unprecedented rate compared to the human brain due to technological advancements and increased computational power. The AI development landscape took a significant turn when Geoffrey Hinton introduced deep belief nets, paving the way for deep learning and developing many algorithms, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Generative Adversarial Networks (GAN) (Shao, Zhao, Yuan, Ding, & Wang, 2022). A critical checkpoint in the path to AGI was the recent breakthrough in the field of Natural Language Processing (NLP), with the Large Language Models (LLMs) utilizing the transformer architecture and attention mechanism (Vaswani et al., 2017), a neural network, specifically utilizing probabilistic models like those seen in generative AI, was able to predict the next word in a sentence after being trained on massive corpora of text. **This process relies on statistical patterns in language, enabling the network to generate coherent text sequences based on prior knowledge**.

While Large Language Models (LLMs) like GPT-4 have demonstrated impressive capabilities, they also exhibit significant limitations, such as 'hallucinations' or the generation of incorrect or nonsensical information. These issues raise questions about the path from LLMs to AGI. According to Gary Marcus and Ernest Davis, the current generation of LLMs lacks true understanding and generalization abilities, which are essential for AGI. Addressing these limitations is crucial for assessing the plausibility of AGI's imminent arrival. Furthermore, critiques of AGI development, as discussed by Ben Goertzel, highlight significant theoretical and practical hurdles, such as the complexities of generalization, ethical considerations, and computational requirements. Engaging with these critiques provides a more nuanced perspective on AGI development challenges and outlines potential strategies to address them.

Initially, AI researchers focused on "narrow AI," systems designed for specific tasks, such as games and pattern recognition, due to the complexity of achieving general intelligence. However, with the emergence of LLMs capable of generating human-like text and reasoning, concerns have grown about the societal impacts of AGI. As AI systems improve in language processing and reasoning, particularly their ability to engage in human-like conversations, examining their societal implications and the ethical challenges posed by their increasing influence across multiple sectors becomes essential.

The journey towards AGI requires a multidisciplinary approach, engaging academia, government, industry experts, and civil society to navigate the vast landscape of intelligent systems. There is also an existential risk associated with the development of AGI, including the possibility of an Artificial Superintelligence (ASI) that could pose an existential threat to humanity if not managed responsibly (Bostrom, 2014) (Russell, 2022). On the one hand, AGI can change how we live by enhancing our lives. On the other hand, it raises ethical and existential concerns, such as the potential for job displacement, privacy issues, and the risk of creating systems that could surpass human control (Bostrom, 2014). However, it's equally important to recognize the counterarguments presented by other scholars, such as Rodney Brooks, who argues that fears of AGI and ASI are exaggerated because current AI systems are highly specialized and lack the generalizable cognitive abilities needed for true AGI. Brooks emphasizes that the trajectory of AI research has been focused on task-specific achievements rather than the broad cognitive competencies required for AGI (Brooks, 2018). Similarly, leading figures like Gary Marcus and Judea Pearl have criticized the notion that current AI developments are on a direct path to AGI, emphasizing the need for fundamentally different approaches to achieve true general intelligence.

This paper analyzes the societal and ethical challenges posed by AGI and argues for the necessity of establishing governance frameworks to mitigate existential risks. It emphasizes the significant challenges and concerns associated with AGI, including its potential to disrupt various aspects of society, the economy, and daily life. Establishing a robust governance framework ensures that AGI development and deployment align with human values and interests, mitigating risks and maximizing benefits. Such a framework should include ethical guidelines, transparency, accountability, and regulatory measures to safeguard against potential misuse and promote the responsible evolution of AGI technologies.

The governance of AGI presents a complex challenge, requiring revisiting the current regulatory frameworks and innovative frameworks to oversee development and deployment. Transparency, accountability, and ethical considerations are essential to ensuring that AGI serves our best interests without compromising privacy and security (Floridi, 2023). Achieving these goals remains a complex and ongoing challenge. Current AI systems have already raised significant societal concerns, such as privacy breaches, biased decision-making, and lack of accountability, which have proven resistant to easy solutions. As a result, new proposals for ethical guidelines, oversight boards, and regulatory agencies are emerging to address these issues and steer AGI development in a more responsible direction.

As AGI transitions from a theoretical possibility to an emerging reality, the central thesis of this paper is that proactive governance and ethical oversight are imperative to ensure AGI's safe and beneficial integration into society. By encouraging a dialogue that engages stakeholders across different industries and prioritizes human values, we can harness the power of AGI while mitigating its risks, ensuring a future where humans are complemented by the system rather than a source of uncertainty and instability.

2. Current State of AGI

The path towards AGI has seen significant progress in recent years due to breakthroughs in machine learning, increased computational power, and the vast amounts of data collected, as illustrated in Figure 1, which shows AI's performance surpassing human baselines on various benchmarks. However, it's important to interpret these comparisons carefully. The term "human baseline" in Figure 1 refers to the average performance level achieved by a representative sample of humans on these tasks. Although these benchmarks demonstrate impressive advancements, factors such as differences in task difficulty and the diversity of human abilities can influence the baseline and don't necessarily equate to general intelligence. Experts predict a technological 'singularity' by 2045 (Brynjolfsson, Rock, & Syverson, 2017a), but the path is complex and multifaceted.

A significant milestone in the pursuit of AGI was the AlphaGo program developed by DeepMind, which combined deep neural networks with techniques like Monte Carlo tree search and reinforcement learning to master the decision-making processes for the complex game of Go. In 2016, AlphaGo defeated Lee Sedol, one of the top Go players in the world, in a historic 4-1 victory (Silver et al., 2017). This demonstrates that through self-learning and playing against itself, AI can improve at a game known for its vast complexity, intuitive elements, and combinatorial challenges, which were traditionally thought to be mastered only by humans.

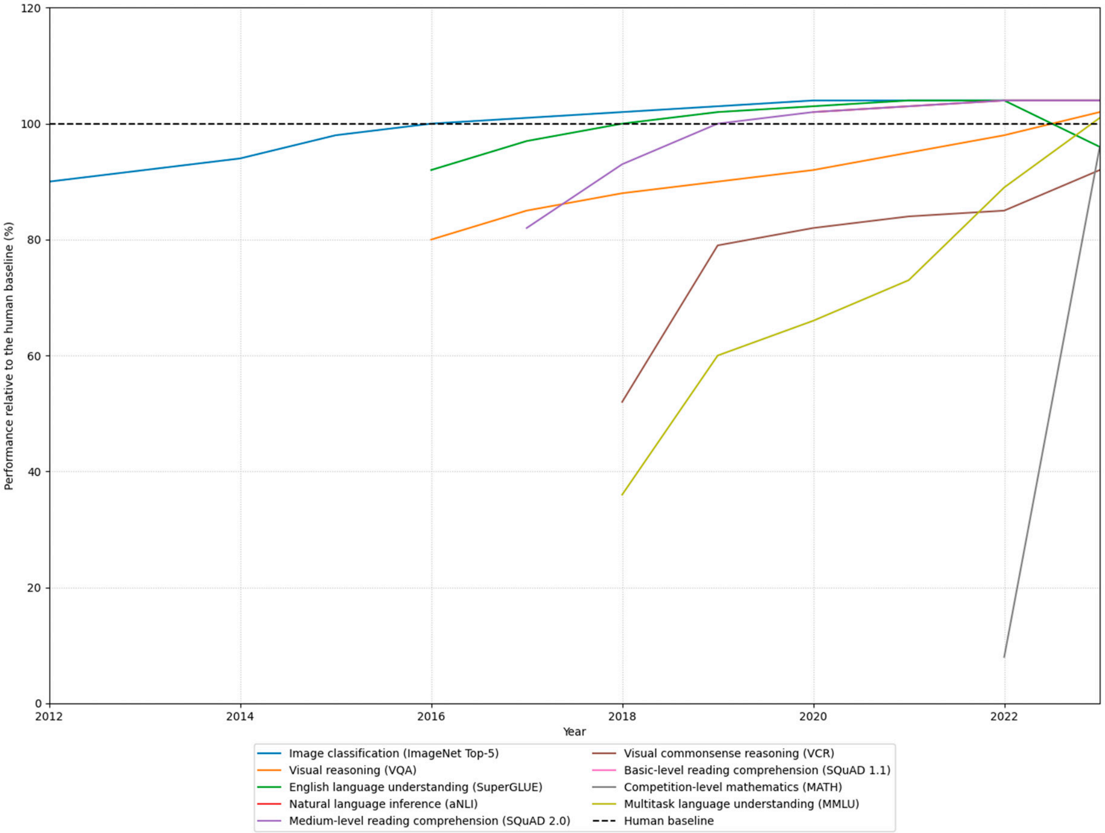

Figure 1. Timeline showing AI's performance surpassing human performance on several benchmarks, including image classification, visual reasoning, and English understanding tasks. The term "human baseline" in the figure represents the average performance achieved by a representative sample of humans on these tasks on these tasks. Source: AI index, 2024.

Machine learning algorithms have primarily driven recent advancements in AGI. The current approach to AGI research spans different methodologies. One notable approach is whole brain emulation (WBE), where researchers in Europe are working on the Human Brain Project, which combines symbolic reasoning with deep learning capabilities to enhance AGI (Marcus, 2018). This approach bridges the gap between symbolic AI, known for logical reasoning, and the statistical strengths of deep learning models (Battaglia et al., 2018). Researchers from Google's DeepMind and Harvard took a significant step forward by utilizing deep reinforcement learning (DRL) within a biomechanical simulation of a rat. They employed advanced DRL techniques, such as Proximal Policy Optimization (PPO), enabling the virtual agent not only to learn complex motor behaviors but also to adapt these behaviors based on feedback from its simulated environment. Unlike earlier models, which were limited to basic, pre-programmed actions, this rat model was designed with high anatomical and physiological precision, incorporating detailed data to mimic real rat movements and motor control accurately. This innovation advanced the understanding of how AI can simulate intricate biological processes, offering deeper insights into motor function and behavioral modeling that were previously unattainable with more simplistic simulations (Aldarondo et al., 2024).

With recent technological advancements, major tech companies like Google, Microsoft, OpenAI, and Anthropic are investing heavily in developing LLMs and AI systems that are multimodal and capable of perceiving, comprehending, and interacting with the world more like humans. Recent milestones include models such as OpenAI's GPT-3 (2020) and GPT-4 (2024), which show remarkable capabilities in natural language understanding and generation across different domains. Google's AlphaCode 2 demonstrated the ability of an AI system to compete with humans in the competitive programming space. With the integration of LLMs and training on 30+ million lines of code, this model could be placed in the 85th percentile (*AlphaCode 2 technical report*). Another company, Anthropic, explored constitutional AI by instilling beneficial values into AI systems to increase intelligence across multiple domains (Bai et al., 2022).

Another focus for researchers is continual learning, which addresses the challenges of AGI and its ability to learn and retain knowledge over extended periods without forgetting previously learned information. Research in this area explores mechanisms such as elastic weight consolidation and synaptic intelligence, which allow models to learn new tasks while retaining knowledge from past experiences (Chaudhry et al., 2019).

In model architectures, developing transformer-based models customized for AGI tasks, such as integrating memory-augmented networks with transformers, enhances the ability to perform complex reasoning and decision-making (Rae, Potapenko, Jayakumar, & Lillicrap, 2019). These models aim to emulate human reasoning processes by scaling cognitive capabilities and processing and manipulating information more effectively.

Furthermore, the generative capabilities of LLMs and their integration with other modalities, such as vision and robotics, have led to the development of new types of AI systems that can interact with the world more dynamically than traditional AI. Researchers are exploring hybrid models that combine rule-based systems and neural networks with memory to create cognitive architectures capable of reasoning and learning like humans. Another promising approach is transfer learning, which allows AI systems to transfer knowledge from one domain to another. For instance, Microsoft, Google, MIT, and Oxford researchers developed DenseAV, an AI algorithm that learns the meaning of words and sounds by “watching” videos. In this context, “watching” refers to the AI’s ability to analyze video content by simultaneously processing visual and auditory signals. The algorithm utilizes advanced techniques in computer vision to identify objects, actions, and contexts depicted in the video while simultaneously interpreting associated spoken or written language. This dual input enables the AI to form associations between words and their meanings as they are presented in dynamic, real-world contexts, potentially advancing our understanding of how language and visual learning interact. Ultimately, this could lead to a more human-like learning experience for AI systems (Hamilton, Zisserman, Hershey, & Freeman, 2024).

In summary, the field of AGI is witnessing significant progress on multiple fronts, from WBE to advances in continual learning and transformer-based architectures. These developments are rapidly bridging the gap between AI and AGI systems. These advancements necessitate addressing the ethical considerations for their development and deployment in society, including the challenges surrounding governance, value alignment, and reliability, which must be resolved before taking the monumental step toward achieving AGI.

3. Implications for the Economy

Automation already plays a vital role in our daily lives, and the widespread adoption of AI technologies will likely have far-reaching implications for the economy. A critical question is whether AI, particularly when it reaches AGI, will complement or replace the human workforce. The answer to this question depends on the type of work being performed and the capabilities of AGI systems. While automation can lead to job displacement, there is also the potential for the “reinstatement effect,” where new opportunities are created as certain tasks become automated (Acemoglu & Restrepo, 2019). Although advancements in AI promise increased efficiency and output, they also raise significant concerns about job displacement, especially in routine and repetitive roles.

Historically, technological advancements have led to new job opportunities and industries. As AI and automation technologies evolve, they may also give rise to new job roles and skill requirements. For instance, developing and maintaining AI systems will require a skilled workforce in data science, machine learning, and software engineering. According to the World Economic Forum, 97 million new job roles may be created due to the adoption of such technologies (World Economic Forum, 2020).

However, one of the primary economic impacts of achieving AGI is the potential for job cuts, particularly in industries like manufacturing, data entry, customer service, and accounting, where many routine tasks can be automated. A study by McKinsey indicated that automation, primarily driven by advancements in AI technologies, could replace up to 800 million jobs by 2030 (Manyika et al., 2017). This projection encompasses various forms of automation, highlighting that while narrow AI has already begun to transform the job market, the advent of AGI could accelerate these changes. Industries such as manufacturing, transportation, and specific administrative roles may experience significant job disruptions as AI systems become more capable of performing traditional tasks done by human workers. For instance, self-driving vehicles and automated logistics systems could displace millions of truck drivers and delivery workers (Autor, 2015).

As the development and control of such technologies lie in the hands of higher-skilled workers, it might lead to income inequality. A report by the International Monetary Fund states, "If AI significantly complements higher-income workers, it may lead to a disproportionate increase in their labor income." (Cazzaniga et al., 2024) which could destabilize economies. Addressing these challenges would require policy measures, including progressive taxation, universal basic income, and social safety nets. Promoting inclusive growth through investments in education and healthcare can ensure that the benefits from AGI can be broadly shared across society (OECD, 2019).

Human capital is a crucial aspect of any economy. A typical timeline to develop a human worker, including education, is more than two decades (Bostrom, 2014), depending on the expertise required for specific industries. This process requires significant investment in education and skill development. Unlike an AI, whose training time depends on the number of resources available, training an LLM, like the GPT-3 model, would take approximately 355 years on a single "*Graphic Processor Unit*" (GPU) (Baji, 2017). In contrast, utilizing massive clusters of GPUs and parallel processing can reduce the training time for such LLMs to around 34 days (Narayanan et al., 2021). This accelerated training process enables the rapid development of AI systems capable of performing complex language tasks. However, the training is a one-time process, and the skills could transfer across different domains, making it much cheaper to deploy new AI agents.

Another concern about AGI handling the financial markets is the inherent "*systemic risk.*" Systemic risk in finance refers to the risk of failure of the entire economic system (Schwarcz, 2008), which arises from the interconnected nature of securities, where the failure of one system can cause a cascading effect on the whole system (Havlin & Kenett, 2015). An unconstrained AGI system tasked with maximizing profits without proper constraints could cause more significant damage than the 2010 flash crash, where a high-frequency trading algorithm rapidly sold S&P 500 E-mini futures contracts, causing stock market indices to drop up to 9% intraday (Staffs of CFTC & SEC, 2010). The full consequences of an unconstrained profit-maximizing AGI system remain unknown.

Given these potential impacts, policymakers face a challenging environment in which to foster innovation while mitigating its economic risks. Some possible policy considerations could be implementing robust safety and ethics regulations, developing AGI-focused antitrust measures to prevent monopoly over markets, and creating retraining programs for displaced workers. As AGI development progresses, addressing these challenges proactively through thoughtful policy-making and inclusive dialogue is crucial.

4. Implications for Energy and Climate

The development of AGI faces significant challenges in terms of energy consumption and sustainability. While journalists and environmental activists have been drawing attention to the ecological impacts of AI advancements, companies and governments often overlook these concerns, prioritizing technological progress and economic benefits. The computational power required to sustain AI models doubles every 100 days (Zhu et al., 2023). Increasing the capacity of a model by tenfold can result in a 10,000-fold rise in power demand. As AI systems become more advanced, their computational demands for training and running the system also increase, refer to Figure 2.

Initiatives such as the Global Alliance on Artificial Intelligence for Industry and Manufacturing (AIM-Global) by the United Nations Industrial Development Organization (UNIDO) highlight the importance of aligning AI advancements with global sustainability goals, particularly in mitigating the environmental impacts associated with AI and AGI technologies (UNIDO, 2023).

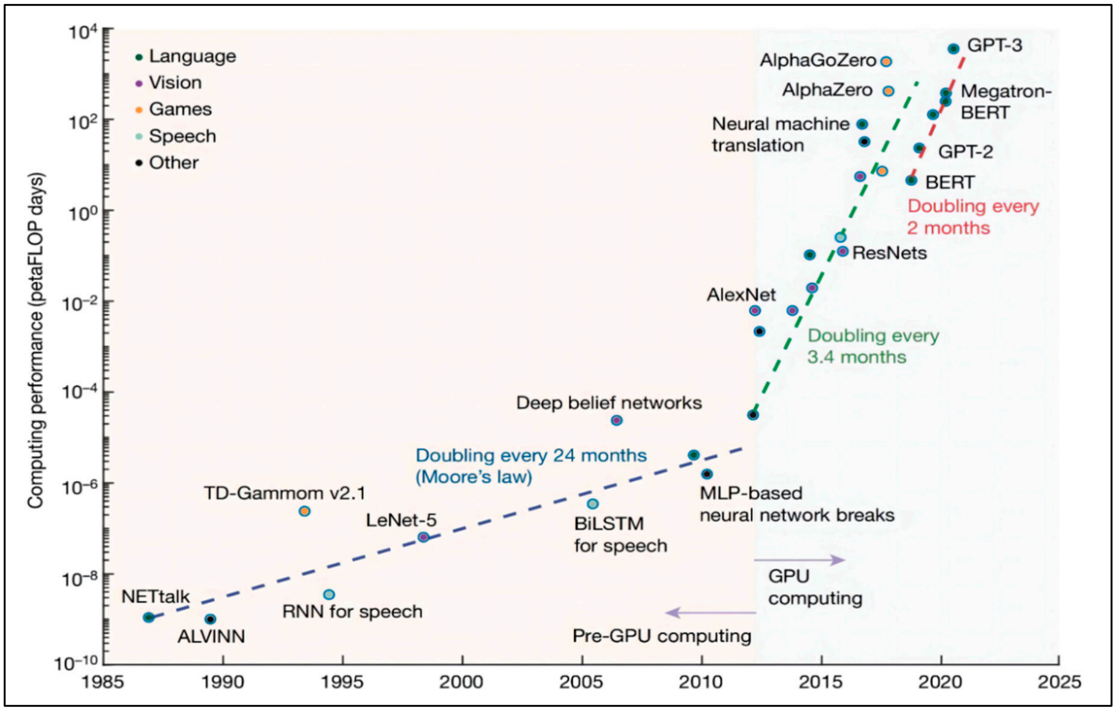

Figure 2. Over the past decade, growth in computing power demands substantially outpacing macro trends (Mehonic & Kenyon, 2022).

The two phases of energy consumption for AI systems are training and inference, with training consuming around 20% and inference consuming 80% of the resources. The energy demand for Natural Language Processing (NLP) tasks is exceptionally high, as the models have to be trained on vast datasets. Researchers estimated that training GPT-3 would have consumed around 1300MWh (Patterson et al., 2021). In comparison, GPT-4 is estimated to have consumed 51,772- 62,318 MWh of energy, which is roughly equivalent to the monthly output of a nuclear power plant (EIA, 2024).

Additionally, the AI models have a significant carbon footprint, with some transformer neural networks emitting over 626,000 lbs. of CO2 (Kurshan, 2023). Techniques such as power-capping the GPUs would reduce energy usage by 15% and only a marginal increase in computational time (McDonald et al., 2022). Another promising avenue is the development of AI-specific energy-efficient hardware designed for workloads. Companies like Nvidia, Google, and Intel are investing in AI chips and tensor processing units (TPUs) that deliver high performance while consuming less power than CPUs and GPUs (Sze, Chen, Yang, & Emer, 2020). Distributed and federated learning are also being considered to distribute the computing load across multiple devices or edge nodes, reducing the energy demands on centralized data centers (Konečný, McMahan, Ramage, & Richtárik, 2016).

If current language models require such immense amounts of energy and computational power for training and inference, developing AGI would necessitate far more resources, leading to even more significant environmental impacts on society. Narrow AI systems, like those used in deep learning, typically require significant computational power to train on large datasets for specific tasks. However, these systems are specialized and not designed to transfer their knowledge to other domains.

In contrast, AGI would need to process and integrate vast amounts of diverse information, reason across different contexts, and make decisions in real-time—capabilities that would demand exponentially more computational resources. As the AGI systems increase in scale and complexity, the energy requirement and carbon footprint would escalate exponentially, potentially straining existing energy infrastructure and exacerbating climate change concerns. Addressing sustainability, energy efficiency, and their challenges will be crucial to mitigate its societal and ecological consequences (Ammanath, 2024).

Shifting towards renewable energy sources and energy-efficient computing infrastructure is essential to minimize the associated environmental impact. Leveraging renewable resources like solar, wind, and hydroelectric power is critical to reducing the strain on infrastructure and mitigating the carbon footprint. However, the energy demands associated with AGI are projected to be exponentially greater than current AI models. While renewable energy is essential across industries, the scale and complexity of AGI would intensify the need for sustainable energy solutions. In addition to transitioning to renewables, AGI development requires further innovation in energy-efficient computing infrastructure to manage its unprecedented power demands. This ensures that AGI's environmental impact is minimized, aligning with efforts to combat climate change.

5. Ethical Implications

Until the advent of machine learning, machines were only relied on to execute a programmed set of instructions. However, with the development of AI and ML systems, there has been a shift from human-centric decision-making to algorithmic systems capable of making autonomous choices. This shift involves transferring certain decision-making responsibilities, such as data analysis, predictions, and recommendations from human experts to machine learning models. Such a transition raises questions about the ethics involved in their decision-making processes. For instance, driverless cars make decisions based on sensor information and the data used to train the algorithms, such as miles driven, driving patterns, and weather conditions. Discussing the ethical implications of decisions made by AGI systems is essential.

5.1. Transparency and Accountability

The development and research leading to AGI systems must prioritize transparency, but concerns surrounding intellectual property, trade secrets, and proprietary algorithms often complicate this goal. While transparency is critical for identifying algorithmic biases and ensuring public trust in AGI's ethical development and use, companies frequently guard the inner workings of their systems under intellectual property laws (Donovan, Caplan, Matthews, & Hanson, 2018). This creates a tension between the need for openness and the protection of competitive advantages. Despite these challenges, balancing transparency and protecting proprietary information is essential. Companies should explore mechanisms such as third-party audits or regulatory oversight, which can allow for necessary scrutiny without compromising sensitive information. Establishing a common governance framework that upholds ethical standards while addressing legal and competitive concerns will ensure accountability in AGI development and reduce the risks associated with opaque decision-making processes. A multidisciplinary approach to transparency in AI, especially AGI, is essential. The literature often refers to explainability, which includes interpretability and user trust in these systems (Ribeiro, Singh, & Guestrin, 2016). The assumption made in recent studies is that transparency must consider how ordinary users understand explanations and assess their relationship with AI products and services. The development of explainable AI is driven by evidence suggesting that many AI applications still need to be used due to a lack of user trust. Users could better comprehend and trust these intelligent agents and predictive models by building more explainable systems.

Furthermore, transparency in AGI also entails addressing various challenges and tensions that arise when AI systems interact with markets and society. As Wachter et al. (2017) suggest, defining transparency requires a broader understanding beyond mere algorithmic explainability. This includes legal ownership aspects, potential transparency abuses, user literacy, and the complex nature of data ecosystems. Striking a balance between transparency and privacy protection is crucial, as excessive openness can lead to misuse, while insufficient transparency can obscure issues such as bias and discrimination (Wachter, Mittelstadt, & Floridi, 2017). A comprehensive approach to transparency incorporating insights from social sciences and humanities alongside technical AI research will ensure that AGI systems align with societal values and ethical standards.

5.2. Ethical Dilemmas

The ethical dilemmas in AGI stem from its broader and more autonomous decision-making capabilities, which could impact various sectors simultaneously. Unlike current AI systems, which are usually limited to specific tasks, AGI's potential to operate across multiple domains introduces ethical challenges of greater complexity. For instance, AGI systems may be tasked with optimizing healthcare resource allocation, financial market decisions, or even national defense strategies, where the consequences of their decisions could have far-reaching societal implications. One of the critical issues is the need for more transparency in AGI systems, as their decision-making processes may be challenging to interpret or audit. There have been instances where opacity in AI systems has led to complex legal and liability issues, such as in cases involving fatalities caused by self-driving vehicles (Griggs & Wakabayashi, 2018). This lack of transparency in AGI could be even more problematic, as decisions may be made autonomously across multiple domains without clear accountability or traceability. The ethical dilemmas that AGI introduces extend beyond the challenges seen in current AI systems. AGI could make autonomous decisions affecting global economic systems, healthcare access, or even national security without human oversight or intervention. As a result, establishing robust governance frameworks is crucial to address incidents where AGI decisions could lead to unintended or harmful consequences (Bird et al., 2020). These frameworks must consider the need for transparency, accountability, and oversight across various sectors, such as healthcare, finance, and public policy, where AGI decisions could profoundly impact human lives.

5.3. Social Responsibility

There should be a strong emphasis on social responsibility when developing and deploying AGI systems, going beyond mere accountability. As (Floridi & Cowls, 2022) argued, "Accountability calls for an ethical sense of who is responsible for how AI works." AGI poses significant social, ethical, and technical challenges, and its development requires consideration of how these systems impact various communities and stakeholders. To address these concerns, a comprehensive and inclusive approach that considers global social responsibility is critical (Saveliev & Zhurenkov, 2021).

The concept of social responsibility, as defined in the broader literature, extends beyond individual stakeholder actions and encompasses the collective obligation to ensure technological advancements benefit society at large. This includes aligning AGI development with the well-being of diverse groups and communities, which may vary across cultures and contexts. As such, defining "well-being" should involve input from a wide range of stakeholders, including policymakers, ethicists, industry leaders, and community representatives, to create an inclusive and culturally sensitive approach to AGI deployment (Floridi & Cowls, 2022).

Moreover, the discussion surrounding automation and its potential impact on employment is particularly relevant to AGI systems. As automation technologies continue to evolve, there is a growing concern that AGI could exacerbate job displacement across various sectors. Unlike traditional automation, which typically replaces repetitive and predictable tasks, AGI systems can perform complex decision-making tasks, potentially displacing finance, healthcare, and law professionals (Brynjolfsson & McAfee, 2014a). Therefore, it is crucial to engage in proactive discussions about the ethical implications of AGI in the workforce, considering the economic impact and the social ramifications on individuals and communities that rely on these jobs for their livelihoods.

In light of these concerns, fostering a dialogue about the future of work in an AGI-driven landscape is essential. This involves exploring innovative approaches to workforce development, including reskilling and upskilling programs that equip individuals with the skills needed to thrive in an increasingly automated economy (Bessen, 2019). Additionally, policymakers and industry leaders must collaborate to create safety nets and support systems for those affected by job displacement, ensuring that the benefits of AGI are distributed equitably across society. Ultimately, a commitment to social responsibility in AGI development necessitates focusing on technological advancement and the well-being of all stakeholders impacted by these transformative changes.

5.4. Bias and Fairness

AI systems, including AGI, can inadvertently perpetuate and even amplify existing biases present in the training data. This issue raises significant ethical concerns about fairness and discrimination. Research has shown that biased algorithms can lead to discriminatory practices in various sectors, such as hiring, law enforcement, and lending (Julia Angwin, 2016). Furthermore, the ethical concerns are exacerbated by the monological nature of many rule-based approaches to AI ethics. These approaches often assume that ethical decisions can be derived from predefined principles, such as the Categorical Imperative or Utilitarianism (Cox, 2015), without dialogue or consideration of diverse perspectives (Abney, 2012). However, such a rigid framework can fail to account for fairness's nuanced and context-dependent nature, leading to ethical blind spots. This underscores the importance of integrating dialogical reasoning and a broader understanding of social contexts into AGI systems to mitigate the risk of bias and promote more equitable outcomes (Goertzel, Pitt, & Novamente, 2012).

6. Privacy and Security Implications

The path to AGI poses many security and privacy implications that must be considered and addressed. Key concerns include the potential for abuse, lack of transparency, surveillance, consent issues, threats to human dignity, and cybersecurity risks.

6.1. Potential for Abuse

AGI could breach individuals' privacy by collecting and analyzing personal data from various sources, including social media, internet searches, and surveillance cameras. Advanced AGI systems might exploit the vulnerabilities in devices or systems to access sensitive information. Companies and organizations could use this data to construct comprehensive profiles detailing of individuals' behaviors, preferences, and vulnerabilities, potentially exploiting this information for commercial or political advantages.

Moreover, AGI-generated content could be indistinguishable from human-generated content, affecting information and disrupting public opinion, leading to confusion and chaos, making it particularly vulnerable to abuse. Autonomous weapons systems using facial recognition further exacerbate these security risks (Brundage et al., 2018).

6.2. Lack of Transparency and Accountability

Many AI systems operate as "black boxes," making their decision-making processes unclear and opaque. This lack of transparency in model interpretation makes it hard to hold the underlying systems accountable for any breach of privacy. It may be unclear why the system made certain decisions or who is responsible for the outcomes (Xi, 2020). The implications of this opacity are particularly concerning in the context of AGI, where decisions may significantly impact individuals' lives. For instance, when an AGI system is involved in critical areas such as healthcare or law, the inability to understand the rationale behind its decisions can lead to unjust outcomes or reinforce existing biases.

The complexity of AGI systems can further exacerbate accountability challenges. As these systems become more advanced and autonomous, attributing responsibility for their actions becomes increasingly difficult. This raises ethical questions regarding who should be held accountable when an AGI system causes harm or violates privacy, whether the developers. These organizations deploy the technology or the AGI itself. Addressing this issue requires a concerted effort to establish clear guidelines and frameworks for accountability in AGI systems, ensuring that transparency is prioritized in their design and implementation. By fostering an environment of openness, stakeholders can work toward building trust and ensuring that AGI systems are developed in a manner that respects individual rights and societal values.

6.3. Surveillance and Civil Liberties

Governments could use the capabilities of AGI to conduct mass surveillance and invade the privacy of individuals. In China, social management is done through systems like the Social Credit System, which contains a set of mechanisms to punish or reward citizens based on their behavior, moral actions, and political conduct based on extensive surveillance (Creemers, 2018). This amount of control tends to push governments toward autocracy and erosion of fundamental human rights. Such practices would have far-reaching consequences on people's lives, including their ability to interact with society, secure jobs, obtain loans, and travel freely.

Furthermore, the normalization of surveillance through AGI technologies could create a pervasive culture of monitoring that undermines civil liberties. Individuals may feel pressured to conform to societal norms and expectations, knowing their actions are continuously observed. This impacts personal freedoms and has chilling effects on free expression and the ability to challenge governmental authority. The potential for AGI to analyze vast amounts of data in real-time could lead to preemptive actions against perceived threats, further curtailing civil liberties. As AGI systems become increasingly integrated into societal governance, it is imperative to establish legal frameworks and ethical guidelines that safeguard individual rights and prevent the misuse of surveillance technologies, ensuring that privacy is upheld in the face of advancing capabilities.

6.4. Difficulty of Consent

The widespread collection and use of personal data by companies often need proper consent. Despite data protection laws, unauthorized data collection incidents, like those involving Cambridge Analytica and YouTube, are common (Andreotta, Kirkham, & Rizzi, 2022). Companies frequently employ shady techniques like burying consent within their lengthy terms and conditions document or leveraging dark patterns to nudge users to trick them into sharing their information. The complexity of AI systems will exacerbate this issue, making it difficult for individuals to contemplate the implications of their consent and data collection. Individuals often have difficulty opting out of such systems since their data is treated as a commodity that can be exploited for commercial gain. This raises concerns about autonomy, privacy, and the potential for discrimination, where AI can acquire personal data for potentially harmful outcomes and misuse.

6.5. Human Dignity

The rise in technologies like deepfakes poses a significant threat to human dignity. Deepfakes create realistic synthetic media, including images and videos that depict individuals saying or doing things they never did. This violates the individual's autonomy over their likeness and will not only raise severe reputational harm and emotional distress but also a sense of loss of control. Moreover, the potential misuse of such technologies for non-consensual pornography or other forms of harassment is a grave danger to the public. The case "Clarkson v OpenAI" highlighted the malicious use of generative AI where the model Dall-E was used to train on public images of non-consenting individuals and used for creating pornography. This involved not only individuals but also kids. The content was then used to extort money by threatening to propagate over social media, which led to intense psychological harm (Moreno, 2024).

The malicious use of such technologies affects not only individuals but also celebrities and government officials, including actresses, country presidents, and high-profile individuals. Incidents caused by AI have been increasing rapidly, see Figure 3. Safeguarding human dignity in the era of deepfakes requires a robust ethical framework and accountability mechanism to prevent such violations and provide recourse and counselling to those affected by such technologies (Anderson & Rainie, 2023).

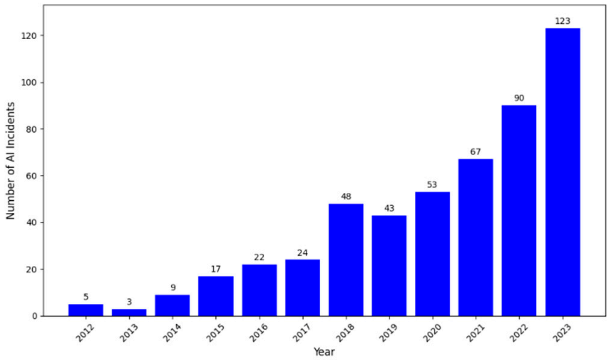

Figure 3. Since 2013, the number of AI-related incidents has surged more than twenty times. One striking example involves AI-generated deepfakes, including sexually explicit content featuring Taylor Swift, which spread widely online. Source: AI Incident Database (AIID), 2023.

6.6. Cybersecurity Risks

AGI presents profound cybersecurity implications. It could be leveraged to develop sophisticated cybersecurity attacks that are highly adaptive, which makes them harder to detect than current attacks. Given their advanced capabilities, these systems could rapidly scan for vulnerabilities and adapt to security measures, simultaneously launching coordinated attacks across multiple systems. AGI-powered attacks pose a significant threat to critical infrastructure like transportation networks, power grids, and communication systems and potentially cause widespread disruptions and failures (Raimondo et al., 2022). Robust security measures designed for AGI systems must be in place to defend against these advanced threats.

Cybersecurity risks from AGI on critical infrastructure are particularly severe regarding public safety, national security, and economic stability. The potential compromise of critical infrastructure due to a cyber-attack could be devastating. This was evident in the case of "*notPetya*" where a cyber-attack on Maersk resulted in malicious software disrupting a fifth of the global shipping capacity, causing $10 billion in damage (Greenberg, 2018). In another instance, the ransomware software "*wannacry,*" a self-propagating malware that encrypts victims' data, causing a worldwide catastrophe impacting hospitals and other critical institutions (Chen & Bridges, 2017). The estimated cost of cybercrime worldwide would skyrocket with the leverage of technologies like generative AI and AGI.

7. Legal and Policy Implications

The implications of AGI development in legal space are significant and span different domains, including intellectual property, liability, privacy, and ethical governance.

7.1. Intellectual Property and Patenting

The emergence of AGI systems capable of generating novel content, ideas, and innovations poses a significant challenge around current Intellectual Property (IP) and patenting regimes. These legal frameworks are designed with humans in mind, and it is necessary to reevaluate how these laws are applied to non-human entities.

The use of AI in creating works challenges the current laws since they are designed to protect the creative work of individuals while making some free for the public. With AI's current capabilities, which can be used to write poems, compose music, draw paintings, and create movie scripts, it becomes perplexing in the legal world to determine if it should be copyrighted (Lee, Hilty, & Liu, 2021).

Researchers have found that some leading LLMs can produce copyrighted content, ranging from passages from The New York Times articles to movie scenes. The central legal question is whether the generation of such content by AI systems violates copyright law.

7.2. Liability and Accountability

With each stride towards AGI, accountability and liability issues become increasingly complex. When an AGI system causes potential harm, who will be held accountable for its wrongdoings: the developers, the company, the users, or the system itself (Scherer, 2015)? The uncertainty about such issues proves a need for more clarity and challenges towards liability. Any AI system and its work should be treated as a product; hence, they must assume the same liability standards as a product (Cabral, 2020). Another issue is that the current liability laws must cover more about personality rights. Therefore, the bias of a system and any damages caused by an incorrect assessment are not covered by product liability laws (Boch, Hohma, & Trauth, 2022).

7.3. Ethical governance and Incentives

A key issue with the development of AGI is its potential to be used for malicious purposes and the significant risk of unintended consequences. We must ensure that the AGI systems align with human values and interests. Industry leaders and policymakers must work together to establish ethical guidelines and incentives to promote responsible development and the use of AGI. One key aspect is mandating ethical reviews and transparency requirements for AGI projects; this could involve third-party audits to assess the ethical implications, potential abuses, and societal impacts. Companies should disclose their ethical principles, decision-making processes, and risk mitigation strategies. Financial incentives like tax breaks or research grants could encourage companies to prioritize ethical considerations, safety, and security when developing AGI (Dafoe, 2018).

However, defining and operationalizing concepts such as fairness, transparency, and accountability remains challenging. These are complex and often context-dependent values, and building consensus around their precise definitions requires continuous effort from industry, governments, and civil society. Achieving this will demand extensive dialogue and collaboration among various stakeholders and adaptability in ethical guidelines to accommodate emerging challenges in AGI development (Graham, 2022). While these concepts are difficult to standardize universally, guidelines must provide a flexible framework to ensure accountability in AGI systems, even as understanding these ethical principles evolves.

Governments must also play a critical role in shaping AGI development through procurement policies. They should set precise requirements for ethical and accountable development and provide incentives and public contracts for companies focusing on s The public sector, through its purchasing power, can encourage the ethical development of AGI by giving preference to solutions that actively address societal challenges and reduce potential risks solving societal problems and aligning with public interests (Dafoe, 2018). Through its purchasing power, the public sector can incentivize ethical AGI development by prioritizing solutions that demonstrate a commitment to addressing societal challenges and minimizing risks.

8. Technological Singularity

Technological singularity refers to a pivotal moment where the capabilities of AGI surpass those of human intelligence by orders of magnitude, potentially leading to unprecedented societal implications. A critical aspect of the singularity hypothesis is the notion of recursive improvement of itself (Dilmegani, 2023). Once AGI reaches a point where it can enhance its capabilities, it initiates an intelligence explosion.

AGI is the catalyst for a singularity because, once achieved, it could recursively improve itself, leading to an intelligence explosion and a rapid expansion of technological progress in a runaway cycle. Each successive iteration of AI would emerge more rapidly and demonstrate greater cognitive prowess than its forerunner (Eden, Steinhart, Pearce, & Moor, 2013). This could result in the creation of artificial superintelligence, an entity whose capabilities would surpass those of any creature on earth. Such a system might autonomously innovate in all aspects of science and technology that humans cannot comprehend or control (Issac, Sangeetha, & Silpa, 2020).

The singularity hypothesis proposes that the emergence of a superintelligence could dramatically transform economies, societies, and human conditions. The rapid advancement of technologies might lead to a scenario where the pace of innovation accelerates so quickly that humans could lose their status as the most capable entities, profoundly impacting our identities, values, and future perspectives. Although the singularity remains a theoretical concept, some experts, such as Brynjolfsson (2017), suggest that AGI could emerge within this century (Brynjolfsson, Rock, & Syverson, 2017b). However, the timeline and feasibility of AGI are subjects of ongoing debate. Chalmers (2010) offers a nuanced philosophical examination of the brain-emulation argument, addressing and countering objections from earlier critics like Dreyfus (2007), who argue that the challenges of achieving such advanced intelligence are significant and may never be overcome (Dreyfus, 2007).

Regardless of its feasibility, the possibility of a singularity demands a proactive approach to the development of AGI, mitigating the existential risks that could arise from such systems. Technological singularity raises profound ethical and existential questions about humanity and its future. If AGI does indeed lead to an intelligence explosion, it questions the role of humans in a world dominated by super-intelligent systems.

The impacts of technological singularity extend beyond philosophical and ethical considerations. This suggested that rapid acceleration could disrupt economies, labor markets, and social structures on an unprecedented scale. The advent of a superintelligence could lead to a period of "brilliant technologies" that would render many human skills obsolete, exacerbating societal tensions (Brynjolfsson & McAfee, 2014b). Economies might face severe disruptions as the industries become increasingly automated, potentially leading to significant job displacements. Social structures can be strained as the gap between technologically augmented and non-augmented individuals widens.

The concept of singularity presents both exciting possibilities and daunting challenges. As we approach the potential development of AGI, it is crucial to engage these profound questions.

9. Proposed Governance Framework

The rapid acceleration towards AGI necessitates a new shift in paradigm for global governance and policy framework. While narrow AI systems have already significantly impacted society, such as through social media platforms that have influenced public discourse and destabilized governments, AGI holds the potential for even more profound and far-reaching changes. This proposal advocates for a comprehensive governance framework designed to navigate the complexities surrounding AGI development and usage while safeguarding human values and interests. By addressing ethical concerns, international collaboration, and regulatory oversight, it aims to prioritize human values, AGI design transparency, accountability, and trustworthiness (Dignum, 2019).

The proposed framework comprises several vital components. First, An AGI oversight board, comprising experts from fields such as AI ethics, law, and sociology, will be empowered through legislative frameworks and international agreements to oversee AGI research and development. This board will work in tandem with existing regulatory bodies to ensure adherence to ethical standards and provide guidance on cross-sectoral issues. A national AGI regulatory agency will enforce AGI-related laws and regulations, coordinating with sector-specific regulators in industries such as healthcare, transportation, and nuclear energy to ensure comprehensive and consistent oversight. This agency will monitor compliance, investigate violations, and impose penalties where necessary. Additionally, establishing national ethical guidelines for AGI will be developed by drawing insights and expertise from diverse stakeholders across multiple industries and disciplines. A global data protocol will define and standardize how data is collected, stored, used, and shared within AGI systems. This protocol will address critical issues such as data ownership, transparency, and bias mitigation. By implementing uniform guidelines, the protocol will ensure AGI systems comply with globally accepted privacy, security, and fairness standards, minimizing risks of misuse or data breaches; see Figure 4. for reference. For example, regulatory bodies like the European Union's

General Data Protection Regulation (GDPR) have set clear frameworks for data protection, and similar principles can be applied globally to AGI. International cooperation is crucial for developing these shared governance standards, much like how the GDPR has become a precedent for cross-border collaboration on data privacy and security (Voigt & Von Dem Bussche, 2017).

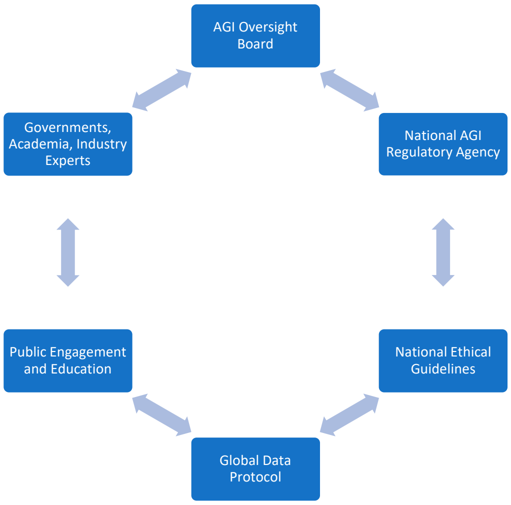
AGI Oversight Board
National AGI Regulatory Agency
National Ethical Guidelines
Global Data Protocol
Public Engagement and Education
Governments, Academia, Industry Experts

Figure 4. Governance Framework for AGI.

Public engagement and education initiatives are also needed to empower citizens with knowledge and information about AGI, fostering trust and informed policy decisions aligning with societal values. Implementing this governance framework will rely on collaboration between governments, academia, and industry experts. Continuous review and iteration will be required to ensure the framework remains responsive to the ever-changing landscape of AGI development, ultimately maintaining its relevance and effectiveness. This approach differs from the current governance frameworks, which focus primarily on Narrow AI, by offering a comprehensive, proactive, and globally coordinated strategy specifically tailored to the unique challenges posed by AGI.

The framework offers a more comprehensive and proactive approach than narrow AI governance models, such as those regulating AI in autonomous vehicles, facial recognition, or algorithmic decision-making systems. For example, autonomous vehicle AI is regulated through transportation safety standards, while privacy laws like the GDPR govern facial recognition technology. The inclusion of an AGI oversight board composed of multidisciplinary experts ensures that diverse perspectives are considered, promoting a balanced approach to ethical and societal concerns (Schuett et al., 2023). This board's role is crucial in maintaining oversight and ensuring adherence to ethical standards, significantly improving over current narrow AI governance mechanisms that often lack comprehensive oversight. Establishing a national AGI regulatory agency provides a centralized body responsible for enforcing AGI-related laws and regulations across various industries, contrasting with the fragmented regulatory landscape for narrow AI, where different sectors may have disparate regulations. A dedicated agency ensures consistent and comprehensive enforcement, reducing regulatory gaps and improving compliance.

Additionally, developing national ethical guidelines for AGI, informed by various stakeholders, ensures robust ethical considerations reflective of societal values. This inclusivity is vital for gaining public trust and legitimacy, which is often a challenge with current narrow AI frameworks that may not adequately address public concerns. Introducing a global data protocol standardizes data practices across AGI systems, addressing critical issues such as data ownership, transparency, and bias. This global coordination is essential for ensuring that AGI development does not exacerbate existing inequalities or create new ethical dilemmas. The comparison with the GDPR highlights the importance of international cooperation and standard-setting in data protection, providing a model for applying similar principles to AGI governance. By addressing the unique challenges of AGI through centralized oversight, robust ethical guidelines, and standardized data practices, the framework ensures that AGI development aligns with human values, promoting transparency, accountability, and trustworthiness. This holistic strategy is essential for managing the profound impacts AGI is expected to have on society.

10. Conclusions

The development of AGI will be transformative with profound implications across technological, ethical, philosophical, legal, and governance domains. This paper has explored these implications comprehensively, delving into key themes such as societal impact, ethical considerations, and governance framework needed to navigate the complexities of AGI responsibly.

From a technical standpoint, AGI promises advancements in automation, decision-making, and problem-solving capabilities. However, these advancements come with significant ethical and societal challenges, and discussing these has highlighted concerns regarding transparency, accountability, and potential misuse. Philosophically, the advent of AGI requires reflection on the nature of human consciousness, moral agency, and human identity. The debate over whether AGI systems can have consciousness poses fundamental questions about the essence of intelligence and its implications for human values. Legal and policy considerations highlight the need for updated intellectual property, liability, and governance frameworks to address the unique problems that AGI might bring. Moreover, technological singularity presents both futuristic possibilities and profound existential risks, which could cause societal disruptions and economic inequalities. In response to these challenges, proposed governance frameworks call for international collaboration, ethical guidelines, and public engagement to foster trust and ensure AGI development aligns with societal values and interests.

References

1. Abney, K. (2012). Robotics, ethical theory, and metaethics: A guide for the perplexed. *Robot ethics: The ethical and social implications of robotics*, 35-52.
2. Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: how technology displaces and reinstates labor. *J Econ Perspect, 33*(2), 3–30. doi:10.1257/jep.33.2.3
3. Aldarondo, D., Merel, J., Marshall, J. D., Hasenclever, L., Klibaite, U., Gellis, A., . . . Ölveczky, B. P. (2024). A virtual rodent predicts the structure of neural activity across behaviors. *Nature*. doi:10.1038/s41586-024-07633-4
4. AlphaCode 2 technical report.
5. Ammanath, B. (2024). How to manage AI's energy demand — today, tomorrow and in the future. Retrieved from https://www.weforum.org/agenda/2024/04/how-to-manage-ais-energy-demand-today-tomorrow-and-in-the-future/
6. Anderson, J., & Rainie, L. (2023). Themes: the most harmful or menacing changes in digital life that are likely by 2035. In As AI spreads, experts predict the best and worst changes in digital life by 2035: they have deep concerns about people's and society's overall well-being (pp. 114–158). Washington: Pew Research Center.
7. Andreotta, A. J., Kirkham, N., & Rizzi, M. (2022). AI, big data, and the future of consent. *AI Soc, 37*(4), 1715–1728. doi:10.1007/s00146-021-01262-5
8. Autor, D. (2015). Why are there still so many jobs? The history and future of workplace automation. *J Econ Perspect, 29*(3), 3–30. doi:10.1257/jep.29.3.3
9. Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., . . . Kaplan, J. (2022). Constitutional AI: harmlessness from AI feedback. arXiv preprint arXiv:2212.08073.
10. Baji, T. (2017). GPU: the biggest key processor for AI and parallel processing. In *Photomask Japan 2017: XXIV symposium on photomask and next-generation lithography mask technology* (Vol. 10454, pp. 24–29). Washington: SPIE.
11. Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., . . . Faulkner, R. (2018). Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261.
12. Bessen, J. (2019). Automation and jobs: when technology boosts employment*. *Economic Policy, 34*, 589-626. doi:10.1093/epolic/eiaa001

13. Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E., & Winfield, A. (2020). *The ethics of artificial intelligence: issues and initiatives*. Brussels: European Parliamentary Research Service.
14. Boch, A., Hohma, E., & Trauth, R. (2022). *Towards an accountability framework for AI: ethical and legal considerations*. Germany: Institute for Ethics in AI, Technical University of Munich.
15. Bostrom, N. (2014). Superintelligence: paths, dangers, strategies.
16. Brooks, R. A. (2018). Intelligence without reason. In *The artificial life route to artificial intelligence* (pp. 25-81): Routledge.
17. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., . . . Amodei, D. (2018). The malicious use of artificial intelligence: forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228.
18. Brynjolfsson, E., & McAfee, A. (2014a). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. New York, NY, US: W W Norton & Co.
19. Brynjolfsson, E., & McAfee, A. (2014b). The second machine age: Work, progress, and prosperity in a time of brilliant technologies: WW Norton & Company.
20. Brynjolfsson, E., Rock, D., & Syverson, C. (2017a). *Artificial intelligence and the modern productivity paradox: a clash of expectations and statistics*. Cambridge: National Bureau of Economic Research.
21. Brynjolfsson, E., Rock, D., & Syverson, C. (2017b). Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. *Kauffman: Large Research Projects (Topic)*.
22. Cabral, T. S. (2020). Forgetful AI: AI and the right to erasure under the GDPR. *Eur Data Prot Law Rev, 6*, 378. doi:10.21552/edpl/2020/3/8
23. Cazzaniga, M., Jaumotte, M. F., Li, L., Melina, M. G., Panton, A. J., Pizzinelli, C., . . . Tavares, M. M. M. (2024). *Gen-AI: artificial intelligence and the future of work*. Washington: IMF.
24. Chaudhry, A., Rohrbach, M., Elhoseiny, M., Ajanthan, T., Dokania, P. K., Torr, P. H., & Ranzato, M. A. (2019). On tiny episodic memories in continual learning. arXiv preprint arXiv:1902.10486.
25. Chen, Q., & Bridges, R. A. (2017). Automated behavioral analysis of malware: a case study of wannacry ransomware. In *2017 16th IEEE international conference on machine learning and applications (ICMLA)* (pp. 454–460). Cancun: IEEE.
26. Cox, J. G. (2015). Reframing ethical theory, pedagogy, and legislation to bias open source AGI towards friendliness and wisdom. *Journal of Ethics and Emerging Technologies, 25*(2), 39-54.
27. Creemers, R. (2018). China's social credit system: an evolving practice of control. *SSRN Electron J.* doi:10.2139/ssrn.3175792
28. Dafoe, A. (2018). *AI governance: a research agenda*. Oxford: Governance of AI Program, Future of Humanity Institute, University of Oxford.
29. Defelipe, J. (2011). The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Front Neuroanat, 5*, 29. doi:10.3389/fnana.2011.00029
30. Dignum, V. (2019). Responsible artificial intelligence: how to develop and use AI in a responsible way. Cham: Springer.
31. Dilmegani, C. (2023). When will singularity happen? 1700 expert opinions of AGI: AIMultiple Research.
32. Donovan, J. M., Caplan, R., Matthews, J. N., & Hanson, L. (2018). *Algorithmic accountability: a primer*. New York: Data & Society.
33. Dreyfus, H. L. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philos Psychol, 20*(2), 247–268. doi:10.1080/09515080701239510
34. Eden, A. H., Steinhart, E., Pearce, D., & Moor, J. H. (2013). Singularity hypotheses: an overview. In A. H. Eden, J. H. Moor, J. H. Søraker, & E. Steinhart (Eds.), *Singularity Hypotheses: A scientific and philosophical assessment* (pp. 1–12). Berlin, Heidelberg: Springer.
35. EIA. (2024). How much electricity does a power plant generate? Retrieved from https://www.eia.gov/tools/faqs/faq.php?id=104&amp%3Bt=3
36. Floridi, L. (2023). The ethics of artificial intelligence: Principles, challenges, and opportunities.
37. Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. In S. Carta (Ed.), *Machine learning and the city: Applications in architecture and urban design* (pp. 535–545). Hoboken: Wiley.
38. Goertzel, B., Pitt, J., & Novamente, L. (2012). Nine ways to bias open-source AGI toward friendliness. *Nine, 22*(1).
39. Good, I. J. (1966). Speculations concerning the first ultraintelligent machine. *Adv Comput, 6*, 31–88. doi:10.1016/S0065-2458(08)60418-0
40. Graham, R. (2022). Discourse analysis of academic debate of ethics for AGI. *AI Soc, 37*(4), 1519–1532. doi:10.1007/s00146-021-01228-7
41. Greenberg, A. (2018). The untold story of NotPetya, the most devastating cyberattack in history. Wired. *August*. Retrieved from https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/
42. Griggs, T., & Wakabayashi, D. (2018). How a self-driving Uber killed a pedestrian in Arizona. The New York Times. Retrieved from https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html

43. Hamilton, M., Zisserman, A., Hershey, J. R., & Freeman, W. T. (2024). Separating the" chirp" from the" chat": self-supervised visual grounding of sound and language. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13117–13127): IEEE.
44. Havlin, S., & Kenett, D. Y. (2015). *Cascading failures in interdependent economic networks.* Paper presented at the Proceedings of the International Conference on Social Modeling and Simulation, plus Econophysics Colloquium 2014.
45. Issac, R., Sangeetha, S., & Silpa, S. (2020). Technological singularity in artificial intelligence.
46. Julia Angwin, J. L., Surya Mattu and Lauren Kirchner, ProPublica. (2016). Machine Bias. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
47. Konečný, J., McMahan, H. B., Ramage, D., & Richtárik, P. (2016). Federated optimization: distributed machine learning for on-device intelligence. arXiv preprint arXiv:1610.02527.
48. Kurshan, E. (2023). Systematic AI approach for AGI: addressing alignment, energy, and AGI grand challenges. arXiv preprint arXiv:2310.15274.
49. Lee, J. A., Hilty, R., & Liu, K. C. (2021). *Artificial intelligence and intellectual property*. Oxford: Oxford University Press.
50. Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst, M. (2017). *A future that works: automation, employment, and productivity*. New York: McKinsey & Company.
51. Marcus, G. (2018). Deep learning: a critical appraisal. arXiv preprint arXiv:1801.00631.
52. McDonald, J., Li, B., Frey, N., Tiwari, D., Gadepally, V., & Samsi, S. (2022). Great power, great responsibility: recommendations for reducing energy for training language models. arXiv preprint arXiv:2205.09646.
53. Mehonic, A., & Kenyon, A. J. (2022). Brain-inspired computing needs a master plan. *Nature, 604*(7905), 255–260. doi:10.1038/s41586-021-04362-w
54. Moreno, F. R. (2024). Generative AI and deepfakes: a human rights approach to tackling harmful content. *Int Rev Law Comput Technol*, 1–30. doi:10.1080/13600869.2024.2324540
55. Mueller, M. (2024). The Myth of AGI. *Internet Governance Project*.
56. Narayanan, D., Shoeybi, M., Casper, J., LeGresley, P., Patwary, M., Korthikanti, V., . . . Zaharia, M. (2021). Efficient large-scale language model training on GPU clusters using megatron-LM. In *SC21: international conference for high performance computing, networking, storage and analysis* (pp. 1–14). New York: Association for Computing Machinery.
57. OECD. (2019). An OECD learning framework 2030. In G. Bast, E. G. Carayannis, & D. F. J. Campbell (Eds.), *The Future of Education and Labor* (pp. 23–35). Cham: Springer International Publishing.
58. Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. M., Rothchild, D., . . . Dean, J. (2021). Carbon emissions and large neural network training. arXiv preprint arXiv:2104.10350.
59. Rae, J. W., Potapenko, A., Jayakumar, S. M., & Lillicrap, T. P. (2019). Compressive transformers for long-range sequence modelling. arXiv preprint arXiv:1911.05507.
60. Raimondo, S. G. M., Carnahan, L., Mahn, A., Scholl, M., Bowman, W., Chua, J., . . . Nistir, B. (2022). *Prioritizing cybersecurity risk for enterprise risk management*. Gaithersburg: National Institute of Standards and Technology.
61. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *" Why should i trust you?" Explaining the predictions of any classifier.* Paper presented at the Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.
62. Russell, S. (2022). Human-Compatible Artificial Intelligence. In.
63. Saveliev, A., & Zhurenkov, D. (2021). Artificial intelligence and social responsibility: the case of the artificial intelligence strategies in the United States, Russia, and China. *Kybernetes, 50*(3), 656–675. doi:10.1108/K-01-2020-0060
64. Scherer, M. U. (2015). Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. *Harv J Law Technol, 29*, 353. doi:10.2139/ssrn.2609777
65. Schuett, J., Dreksler, N., Anderljung, M., McCaffary, D., Heim, L., Bluemke, E., & Garfinkel, B. (2023). Towards best practices in AGI safety and governance. *Surv. Expert Opin.*
66. Schwarcz, S. L. (2008). Systemic risk. *Geo. Lj, 97*, 193.
67. Shao, Z., Zhao, R., Yuan, S., Ding, M., & Wang, Y. (2022). Tracing the evolution of AI in the past decade and forecasting the emerging trends. *Expert Syst Appl, 209*, 118221. doi:10.1016/j.eswa.2022.118221
68. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., . . . Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature, 550*(7676), 354–359. doi:10.1038/nature24270
69. Staffs of CFTC, & SEC. (2010). Findings regarding the market events of MAY 6, 2010. Retrieved from https://www.sec.gov/news/studies/2010/marketevents-report.pdf
70. Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2020). *Efficient processing of deep neural networks*. Cham: Springer.

71. UNIDO. (2023). UNIDO launches global alliance on ai for industry and manufacturing (AIM-Global) at world AI conference 2023. Retrieved from https://www.unido.org/news/unido-launches-global-alliance-ai-industry-and-manufacturing-aim-global-world-ai-conference-2023
72. Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. In *NIPS'17: proceedings of the 31st international conference on neural information processing systems* (pp. 6000–6010). California: Curran Associates Inc.
73. Voigt, P., & Von Dem Bussche, A. (2017). *The EU general data protection regulation (GDPR): a practical guide* (Vol. 10). Cham: Springer International Publishing.
74. Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law, 7*(2), 76-99. doi:10.1093/idpl/ipx005
75. World Economic Forum. (2020). The future of jobs report 2020. Retrieved from https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf
76. Xi, B. (2020). Adversarial machine learning for cybersecurity and computer vision: current developments and challenges. *Wires Comput Stat, 12*(5), e1511. doi:10.1002/wics.1511
77. Zhu, S., Yu, T., Xu, T., Chen, H., Dustdar, S., Gigan, S., . . . Pan, Y. (2023). Intelligent computing: the latest advances, challenges, and future. *Intell Comput, 2*, 0006. doi:10.34133/icomputing.0006

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.