Review

Not peer-reviewed version

Global Impacts of Artificial General Intelligence and Artificial Superintelligence

Mehmet Orhan *, Zeynep Orhan, Brady Lund, Nayana P. Madali

Posted Date: 6 July 2026

doi: 10.20944/preprints202607.0329.v1

Keywords: artificial intelligence (AI); artificial general intelligence (AGI); artificial super intelligence (ASI); AI ethics; AI economics; AI governance

 doi:10.20944/preprints202607.0329.v1

Review

Global Impacts of Artificial General Intelligence and Artificial Superintelligence

Mehmet Orhan [1,*], Zeynep Orhan [1], Brady Lund [1] and Nayana P. Madali [2]

[1] University of North Texas

[2] University of Kansas

* Correspondence: mehmet.orhan@unt.edu

Abstract

This study examines the global impacts of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI) through a conceptual review of the related literature. AGI refers to AI systems that may match human cognitive ability across multiple domains, while ASI refers to systems that may surpass human intelligence across scientific, social, strategic, and creative tasks. Recent advances in large language models, agentic systems, and general-purpose AI have intensified concern about the gap between rapid technological development and slower adaptation in economic structures, ethical principles, and governance frameworks. This review synthesizes economic, ethical, and governance literature to examine how AGI and ASI may affect productivity, labor markets, inequality, human autonomy, safety, and institutional accountability. The literature presents competing economic scenarios, ranging from modest productivity gains to explosive growth and a possible post-labor economy marked by wage decline and weakened aggregate demand. The review also identifies risks related to value alignment, catastrophic misuse, algorithmic bias, surveillance, moral dependency, and existential harm. These risks may deepen global inequality by concentrating wealth, computational power, and decision-making authority among technologically advanced nations and dominant technology firms. The study concludes that current institutional structures remain insufficient for managing the systemic disruptions associated with AGI and ASI. A stable transition will require proactive international cooperation, stronger governance frameworks, social protection, AI capital taxation, and alignment strategies that reflect collective human values.

Keywords: artificial intelligence (AI); artificial general intelligence (AGI); artificial super intelligence (ASI); AI ethics; AI economics; AI governance

1. Introduction

Artificial intelligence (AI) has moved from narrow, task-specific systems toward increasingly general-purpose systems. Early AI applications, such as voice assistants and recommendation engines, performed limited tasks within defined domains. Recent systems, including large language models (LLMs) and agentic AI systems, can perform a broader range of cognitive, linguistic, analytical, and creative tasks. This development has intensified discussion about Artificial General Intelligence (AGI), which refers to AI systems capable of matching human cognitive ability across multiple domains (Bostrom, 2014; Yenduri et al., 2025). Artificial Superintelligence (ASI) refers to systems that surpass human intelligence across scientific, social, strategic, and creative domains (Kim et al., 2024). These developments may produce major gains in efficiency, cognitive automation, innovation, and decision-making. They may also create serious economic, ethical, social, and governance challenges at a global scale.

The motivation for this study arises from the widening gap between the pace of advanced AI development and the slower adaptation of economic structures, ethical principles, and governance frameworks. AI has transformed significantly since its formal emergence as a research field in 1956. Recent advances in machine learning, computer vision, natural language processing, LLMs, and

 doi:10.20944/preprints202607.0329.v1

general-purpose AI systems have made AI systems more flexible, adaptive, and autonomous (Mundlamuri et al., 2025; Naveed et al., 2025). These capabilities blur the boundaries between narrow AI and early forms of more general intelligence. The transition from narrow AI toward AGI is therefore no longer only a theoretical issue. It has become an emerging practical concern for economies, institutions, governments, and societies. This transition raises urgent questions about safety, accountability, regulation, labor markets, inequality, and long-term human welfare.

Scholarly interest in AGI and ASI has grown alongside recent advances in machine learning, computer vision, and natural language processing (Barrett & Baum, 2017; Goertzel, 2014). Goertzel (2014) describes AGI as a synthetic intelligent system with broad generalization capacity and the ability to operate across complex environments. Researchers have examined possible AGI applications in education, the Internet of Things (IoT), healthcare, manufacturing, and climate-related problem solving (Dou et al., 2023; Latif et al., 2023; Masters et al., 2024). Healthcare applications could support faster medical information processing and more accurate diagnosis, while educational applications could provide adaptive tutoring based on students' pace, needs, and preferences (Fahad et al., 2024). AGI systems may also automate high-level cognitive tasks, support hazardous industrial processes, and accelerate scientific research. These applications suggest that AGI could significantly reshape many sectors.

Advanced AI systems also create substantial risks. Researchers have raised concerns that AGI systems may become difficult to control, especially if they acquire strong autonomous planning or self-improvement capacities (Bostrom, 2014; Everitt et al., 2018; McLean et al., 2023; Salmon et al., 2021; Salmon et al., 2024). Bostrom (2014) argues that systems capable of surpassing human intelligence could enter a self-reinforcing process of improvement and eventually produce superintelligence. Other scholars also discuss intelligence explosion and technological singularity as possible long-term outcomes of recursive self-improvement (Chalmers, 2016; Mahler, 2022; Naudé & Dimitri, 2020; Radanliev et al., 2022). ASI systems could understand, learn, and solve problems at speeds and levels beyond human capacity (Fahad et al., 2024; Saghiri et al., 2022). These possibilities remain debated, but their potential consequences justify systematic analysis because AGI and ASI could affect global stability, human agency, and social organization.

The literature on advanced AI has examined economic, ethical, and governance implications, but these discussions often remain fragmented. Economic studies show that AI-led automation can increase productivity, reshape labor markets, and support long-term economic growth (Brynjolfsson & McAfee, 2017; Brynjolfsson et al., 2017; Furman & Seamans, 2019; Tyson & Zysman, 2022; Zubair, 2024). Other studies warn that AI-led automation may displace workers, widen wage gaps, increase labor market polarization, and deepen regional disparities (Acemoglu & Restrepo, 2018; Faishal et al., 2023; Joshi, 2024). Much of the existing economic literature focuses on current or near-term AI technologies rather than the more disruptive implications of AGI and ASI (Chen et al., 2016; Trabelsi, 2024; Trammell & Korinek, 2023).

Ethical research has also examined alignment, autonomy, accountability, transparency, bias, privacy, misinformation, and decision-making in AI systems (Akinrinola et al., 2024; Gabriel, 2020; Lechterman, 2022; Prunkl, 2024; Stahl, 2021). Value alignment has become a central concern because advanced AI systems must act in accordance with human goals, ethical expectations, and social values (Gabriel & Ghazavi, 2022). This challenge becomes more serious when AI systems operate with high autonomy or make decisions that humans cannot easily interpret. Bostrom (2020) argues that advanced AI may raise ethical problems that differ substantially from the problems associated with current narrow AI systems. Existing ethical guidelines and policy frameworks often focus on task-based AI systems and may not fully address the risks associated with AGI or ASI. National and regional differences in AI ethics also create fragmented governance and limited global coordination (Daly et al., 2019; Tidjon & Khomh, 2022).

AI governance translates ethical principles into standards, policies, legal frameworks, accountability systems, and institutional practices. Existing governance approaches include regulatory frameworks, international standards, ethical principles, and industry-led safety initiatives

 doi:10.20944/preprints202607.0329.v1

(Batool et al., 2023; Ismail & Ahmad, 2025; Taeihagh, 2021). Recent research has examined emerging AI governance frameworks and standardization efforts, but many of these initiatives remain oriented toward current AI systems rather than AGI or ASI (Lund et al., 2025; Orhan et al., 2025). Scholars continue to emphasize the need for stronger regulatory frameworks, technical safety policies, global cooperation, institutional guidance, and accountability mechanisms (Birkstedt et al., 2023; Dafoe, 2018; Mäntymäki et al., 2022). These concerns raise a critical question about whether current governance systems can respond adequately to the complex and systemic risks posed by AGI and ASI.

Debates about AGI and ASI remain divided. Some scholars and technology leaders suggest that AGI and ASI could expand human capabilities, accelerate innovation, and create substantial economic abundance (Hagey, 2025; Kurzweil, 2000). Other researchers warn that these systems may create serious risks related to alignment failures, loss of control, and large-scale unintended consequences (Bengio et al., 2024; Russell, 2021). Similar disagreement exists regarding ASI. Some researchers argue that ASI could produce intelligence explosion and eventually exceed human control, while others regard such scenarios as speculative or exaggerated (Goertzel, 2014; Muehlhauser & Salamon, 2013). These debates show the need for balanced analysis that considers both near-term disruptions and long-term risks.

This study addresses these gaps through a conceptual literature review of the global impacts of AGI and ASI. The review synthesizes literature from academic research and reliable institutional sources to examine economic consequences, ethical concerns, and governance challenges. It contributes to the AI literature by integrating these dimensions rather than treating them as separate debates and identifies potential risks and opportunities associated with AGI and ASI development and discusses strategies that may reduce harm while preserving possible social benefits. The central argument is that the impact of AGI and ASI will depend not only on technical capability, but also on governance structures, regulatory effectiveness, alignment strategies, economic policy, and international coordination.

The paper is organized as follows. Section 2 examines the economic consequences of AGI and ASI, including macroeconomic growth projections, labor market disruption, sectoral transformation, inequality, and global policy implications. Section 3 analyzes ethical concerns and governance challenges associated with AGI and ASI, including value alignment, control, weaponization, bias, surveillance, human autonomy, dependency, existential risk, oversight, and risk reduction strategies. Section 4 concludes by emphasizing the need for proactive international cooperation, institutional preparedness, and governance frameworks that can support shared global benefit while reducing systemic risks.

2. Economic Consequences of AGI and ASI

AGI and ASI may become among the most disruptive forces in modern economic history. Current generative AI systems have already attracted large investment projections and have begun to produce measurable productivity effects. AGI and ASI could extend these effects by reshaping the division of labor, the distribution of capital, the structure of production, and the long-term trajectory of economic growth (Jones, 2026; Raman et al., 2025).

Economic analysis requires a clear distinction between narrow AI, AGI, and ASI. Narrow AI performs specific tasks within limited domains and does not transfer knowledge broadly across different contexts. AGI refers to a hypothetical form of AI that can understand, learn, reason, adapt, and solve problems across a broad range of tasks in ways that resemble general human cognitive ability (AMRO Asia, 2025). ASI refers to systems that surpass the best human cognitive performance across economically, scientifically, and strategically meaningful domains. Narrow AI affects particular occupations through task-specific automation. AGI may challenge the assumption that human labor remains necessary for complex judgment and knowledge work. ASI may go further by reducing the scarcity of cognition as a constraint on economic output. The movement from narrow

 doi:10.20944/preprints202607.0329.v1

AI to AGI and ASI is therefore not merely a linear technological progression. It may represent a structural transformation in the organization of economic activity (Joshi, 2025; Raman et al., 2025).

2.1. Macroeconomic Growth Scenarios

Economists remain divided about the macroeconomic effects of AGI. One view predicts that advanced AI will produce meaningful but limited growth within existing economic structures. Acemoglu (2025) estimates that AI may increase total factor productivity by no more than 0.71% over ten years. This cautious projection rests on the claim that many human tasks remain difficult to automate because they require context, judgment, tacit knowledge, and outcome measures that are not easily formalized. MIT Sloan Management Review (2024) presents a similar interpretation and reports Acemoglu's estimate that only about 5% of tasks may be profitably automated over a ten-year period.

A more expansive view predicts that AGI could produce a much larger growth acceleration. Aghion et al. (2019) argue that AGI could remove weak links in the production chain and produce major productivity gains. Korinek and Suh (2024) model scenarios in which AGI could generate 100% to 300% gross domestic product (GDP) growth over a decade. Some growth models also suggest that AGI could produce annual growth rates of 30% or higher if it automates the process of idea generation itself. These projections differ sharply from more conservative estimates because they treat AGI not only as a tool for improving existing production but also as a system that could accelerate innovation, research, and knowledge creation.

Generative AI forecasts already indicate large potential gains before AGI emerges. Goldman Sachs estimated that widespread AI adoption could raise annual global GDP by approximately 7%, or about $7 trillion, over a ten-year period and increase annual labor productivity growth by 1.5 percentage points (Goldman Sachs, 2023a). McKinsey Global Institute (2023) projected that generative AI could add between $2.6 trillion and $4.4 trillion annually across major use cases, while broader estimates place potential global economic effects at much higher levels as AI capabilities mature. Goldman Sachs Global Investment Research (2023) further projected that AI's measurable impact on U.S. GDP could begin around 2027 and raise GDP growth by roughly 0.4 percentage points during the late 2020s and early 2030s. KPMG (2025) also projected substantial long-term gains under rapid AI adoption, while its slower adoption scenario still estimated a 7.7% increase in real GDP by 2050.

AGI scenarios introduce greater uncertainty because they may alter the role of labor as a binding constraint on production. The International Monetary Fund presents several possible macroeconomic trajectories for an AGI future, including business-as-usual growth, gradual AGI adoption, and rapid AGI adoption (International Monetary Fund, 2023). The two AGI scenarios produce faster output growth because AGI may reduce the economic scarcity of cognitive labor. These scenarios should not be interpreted as predictions. They should be treated as policy-relevant possibilities that require preparation across fiscal policy, labor policy, education, social protection, and international economic coordination.

2.2. Labor Market Disruption and the Post-Labor Economy

Labor market effects remain central to the economic debate on AGI and ASI. Current AI systems already affect a large share of jobs. The International Monetary Fund estimates that AI may affect nearly 40% of jobs globally, with higher exposure in advanced economies (MIT Sloan Management Review, 2024). Goldman Sachs Global Investment Research (2023) reports that roughly two-thirds of current jobs have some degree of AI automation exposure and that generative AI could substitute for up to one-fourth of current work tasks.

The task-based approach explains why these effects vary across occupations. Jobs consist of multiple tasks, and each task differs in its susceptibility to automation. AI can either substitute for human labor by performing tasks previously done by workers or complement human labor by increasing worker productivity. AGI could intensify this tension because it may automate not only

 doi:10.20944/preprints202607.0329.v1

routine tasks but also judgment-based, analytical, and creative tasks. This possibility creates a productivity-prosperity problem. Output may rise, but wage income may decline if machines become abundant substitutes for human labor (International Monetary Fund, 2023).

Some models suggest that AGI labor operating at near-zero marginal cost could place severe downward pressure on human wages. Stiefenhofer (2025) argues that the marginal productivity of human labor may decline sharply if AGI becomes a scalable substitute for cognitive labor. This scenario could create a post-labor economy in which firms produce more output with less human labor, while weakened wage income reduces aggregate demand. This risk does not imply that all human work will disappear. It does indicate that existing labor-based income systems may become unstable if productivity gains concentrate among owners of AI capital.

AGI may also weaken apprenticeship pathways. Entry-level tasks often help workers develop tacit knowledge, professional judgment, and domain expertise. Firms that replace entry-level work with AGI systems may reduce opportunities for junior workers to acquire the experience needed for later expert oversight. Catalini et al. (2026) describe this problem through the concepts of the measurability gap and the missing junior loop. Their argument suggests that execution may become cheaper as automation expands, while verification, liability, and ground-truth judgment may remain scarce because they still depend on human expertise. This shift may move economic value from routine execution toward oversight, validation, and responsibility.

Empirical and theoretical evidence also suggests that AI may affect workers unevenly across skill levels. Noy and Zhang (2023) found that generative AI assistance produced heterogeneous productivity effects and disproportionately improved the performance of lower-skilled workers in writing tasks. This pattern may compress output differences within measurable tasks. Agrawal et al. (2025) describe transformative AI as a form of scalable intelligence that could reorganize labor markets by displacing routine cognitive work while increasing the value of human judgment in tasks that require accountability, contextual understanding, and social interpretation. AMRO Asia (2025) similarly reports that occupational exposure under AGI may exceed exposure under current generative AI, with a larger share of occupations classified as automation prone.

Historical patterns suggest that the full productivity effects of general-purpose technologies often appear after a significant adoption lag. Goldman Sachs (2023b) notes that past productivity booms, such as those associated with electricity and personal computers, often emerged after widespread business adoption. However, AGI may produce faster adjustment pressures than earlier technologies because cognitive automation can spread across sectors more quickly than physical infrastructure. Early evidence already suggests that workers in AI-exposed occupations may experience weaker employment and earnings outcomes before AGI fully emerges (Stanford HAI, 2025). These findings support the need for workforce adaptation before large-scale disruption becomes visible.

2.3. Sectoral Transformation

AGI and ASI will not affect all sectors in the same way. Healthcare, finance, education, manufacturing, scientific research, legal services, and public administration may face both significant disruption and substantial opportunity. Healthcare may experience major changes through diagnostic reasoning, medical information processing, personalized treatment, clinical decision support, and AI-driven drug discovery. Goldman Sachs Global Investment Research (2023) identifies healthcare as a sector in which AI could become paradigm-shifting. AI-driven protein design, disease modeling, and treatment optimization could reduce research timelines and improve health outcomes, although these gains would require strong accountability, validation, and regulatory oversight.

Financial services may also experience significant transformation. Cognitive automation can affect financial analysis, risk modeling, fraud detection, compliance monitoring, insurance underwriting, and investment decision-making. AGI-level systems capable of autonomous financial reasoning could change the structure of investment banking, insurance, asset management, and regulatory supervision (Joshi, 2025). These changes may improve efficiency but may also create new

 doi:10.20944/preprints202607.0329.v1

systemic risks if financial institutions rely heavily on opaque or insufficiently supervised autonomous systems.

Education and training may become both a disrupted sector and an adaptive response to disruption. AGI systems may support personalized tutoring, automated feedback, curriculum design, and large-scale learning support. At the same time, education systems must prepare students and workers for labor markets in which routine cognitive tasks become less valuable. KPMG (2025) identifies education and training as areas with strong investment multiplier potential under rapid AI adoption. This finding suggests that education should not be treated only as a sector affected by AI but also as a central institution for economic adaptation.

Scientific research may experience the most transformative effects under ASI scenarios. ASI systems that can generate hypotheses, design experiments, analyze results, and improve research strategies could accelerate discovery in medicine, energy, climate science, materials science, and engineering. This possibility remains uncertain, but it represents one of the largest potential benefits of advanced AI. The same capability also creates governance concerns because societies may struggle to evaluate, regulate, or absorb discoveries generated at speeds beyond current institutional capacity.

Manufacturing and logistics may experience continued automation through robotics, planning systems, autonomous vehicles, and intelligent supply-chain coordination. Earlier automation waves primarily affected repetitive industrial work, but AGI may also affect professional work in law, education, engineering, medicine, finance, software development, journalism, scientific research, and management (Russell, 2019). This distinction matters because previous technological transitions often replaced some forms of manual labor while expanding demand for cognitive and professional labor. AGI may challenge this pattern by automating parts of the cognitive work that previously served as the pathway to economic mobility.

2.4. Inequality, Capital Concentration, and Aggregate Demand

AGI and ASI may deepen inequality if productivity gains accrue mainly to the owners of advanced AI systems, computational infrastructure, proprietary data, and specialized talent. Current digital economies already show high levels of concentration among a small number of multinational technology firms. AGI development could strengthen this concentration by centralizing access to knowledge production, automation capacity, and economic decision-making (Bostrom, 2014; Crawford, 2021).

Capital concentration may also create a divergence between productivity growth and broad prosperity. Firms that own or control AGI systems may increase output and profits, while workers whose tasks are automated may lose bargaining power or employment income. The International Monetary Fund's scenario modeling suggests that wages may rise initially under some AGI scenarios but may later decline if labor becomes highly substitutable, while returns to AI capital increase (International Monetary Fund, 2023). Acemoglu (2025) similarly warns that even modest GDP gains may coexist with declining welfare for some groups if AI adoption mainly benefits high earners or capital owners while eroding middle-skill employment.

Aggregate demand presents another concern. Modern economies depend heavily on wage income as a source of consumption. A system that produces more output while reducing broad wage income may face demand-side instability. This concern motivates growing discussion of Universal Basic Income(UBI), reduced workweek models, wage insurance, retraining systems, public ownership models, and AI capital taxation. These proposals differ in design, but they share a common objective. They aim to preserve social stability and purchasing power if advanced AI reduces the role of human labor in income generation.

AGI may also affect social mobility through the weakening of professional entry points. Entry-level tasks in law, finance, software development, journalism, healthcare, and research often provide the early experience needed for career development. Extensive automation of these tasks may reduce opportunities for new workers to enter professional fields. This effect may widen intergenerational

 doi:10.20944/preprints202607.0329.v1

inequality unless institutions create new pathways for training, credentialing, supervision, and practical experience.

2.5. Global Economic Divergence and Policy Implications

AGI and ASI may produce highly uneven effects across countries. National benefits from advanced AI will depend on digital infrastructure, human capital, institutional capacity, regulatory readiness, energy resources, computational access, and domestic innovation systems. Cirera et al. (2022) emphasize that technology adoption depends heavily on firm-level and country-level readiness. The same principle applies to AGI because countries with stronger infrastructure and governance capacity will be better positioned to benefit from advanced AI systems.

The global distribution of AGI benefits may therefore produce a second Great Divergence. Advanced economies are more likely to benefit from AGI because they possess stronger digital infrastructure, larger pools of technical talent, more developed regulatory systems, greater access to compute, and closer ties to frontier AI firms. The International Monetary Fund's AI Preparedness Index indicates that advanced economies are better positioned than emerging markets and low-income countries to benefit from AI because of stronger infrastructure, human capital, innovation capacity, and regulatory frameworks. Simulations suggest that AGI-driven growth in advanced economies could be more than double that of low-income countries (Cerutti et al., 2025).

Reshoring may intensify these global inequalities. Developing economies have often relied on labor-cost advantages to attract manufacturing and service-sector activity. AI-driven automation may reduce the value of low-cost labor and weaken this traditional path to development (Steinmetz, 2025). Countries that lack domestic AI infrastructure may become dependent on foreign corporations or technologically dominant states for AI services, cybersecurity, healthcare systems, financial tools, education platforms, and digital governance systems. This dependence may weaken national autonomy and reduce the capacity of smaller or poorer countries to regulate dominant AI actors effectively (UNESCO, 2021).

Some models also suggest that AI-driven productivity gains may affect exchange rates and sectoral prices. Cerutti et al. (2025) and Wingender (2025) discuss the possibility that AI productivity gains in nontradable sectors, such as healthcare and education, may lower service prices and affect real exchange rates. These macroeconomic effects remain uncertain, but they show that AGI may influence not only labor markets and firm productivity but also trade, currency values, national competitiveness, and development pathways.

Policy responses must therefore address both domestic and international dimensions. Scenario planning offers one useful approach. The International Monetary Fund recommends that policymakers stress-test fiscal, monetary, labor, and social protection systems across business-as-usual, gradual AGI, and rapid AGI scenarios rather than relying on a single forecast (International Monetary Fund, 2023). Governments should also invest in education systems that emphasize problem solving, creativity, interpersonal judgment, ethical reasoning, and domain expertise. These capacities may remain important even as AI systems automate many routine cognitive tasks.

Redistribution mechanisms may become necessary if AGI separates productivity growth from wage growth. Social protection systems, wage insurance, public investment, UBI, AI capital taxation, and worker transition programs may help preserve aggregate demand and social stability. International coordination will also be necessary because AGI competition may create incentives for regulatory arbitrage, unequal access, and race-to-the-bottom dynamics. Shared safety standards, infrastructure support for developing countries, and cooperative governance arrangements may reduce the risk that AGI and ASI benefits concentrate only among technologically advanced nations and dominant technology firms.

Overall, AGI and ASI may produce substantial economic benefits, but these benefits will not automatically translate into shared prosperity. Economic outcomes will depend on institutional design, distributional policy, labor adaptation, international coordination, and governance capacity. The central economic challenge is therefore not only how much productivity AGI and ASI may

 doi:10.20944/preprints202607.0329.v1

generate, but also who will control that productivity, who will benefit from it, and how societies will maintain stability if human labor becomes less central to production and income.

3. Ethical Concerns and Governance Challenges of AGI and ASI

AGI and ASI represent possible next stages in the evolution of AI systems. Although both remain largely theoretical, recent advances in increasingly autonomous, adaptive, and general-purpose AI systems have intensified academic, political, and societal debate about their long-term implications. Current generative AI systems already demonstrate advanced capabilities in coding, scientific reasoning, language generation, image synthesis, and strategic gameplay. Some researchers argue that continued progress in computational power, multimodal learning, and self-improving architectures may eventually support AGI-level systems. Other scholars remain skeptical and argue that current models still lack genuine reasoning, consciousness, intentionality, and deep contextual understanding despite their impressive performance outputs (Bostrom, 2014; Marcus, 2022; Russell, 2019).

AGI and ASI create different levels of ethical concern. AGI-related concerns often involve economic disruption, decision-making autonomy, surveillance, labor displacement, social manipulation, accountability, and institutional dependence (Crawford, 2021; Russell, 2019). A future AGI system integrated into a national healthcare infrastructure, for example, could diagnose diseases, allocate medical resources, and recommend treatment plans more efficiently than human physicians. This capability could improve healthcare access and efficiency, but it could also raise concerns about accountability, transparency, bias, and human oversight (Short, 2025). ASI-related concerns extend further because systems that surpass human intelligence may become capable of recursive self-improvement, strategic deception, autonomous goal pursuit, or independent optimization beyond meaningful human control. A hypothetical ASI system managing global financial markets, cybersecurity systems, energy grids, or military infrastructures could influence human civilization at a scale that no previous technology has achieved (Bostrom, 2014; Russell, 2019).

Scholars disagree about how much attention long-term AGI and ASI risks should receive. Some argue that fears about AGI and ASI remain speculative because no current evidence demonstrates machine consciousness or autonomous intentionality (Marcus, 2022). This perspective warns that excessive attention to superintelligence may distract policymakers from immediate harms already caused by current AI systems, including discrimination, misinformation, exploitative surveillance, and labor inequality (Crawford, 2021). Other scholars argue that long-term catastrophic risks deserve serious attention because preventive governance becomes more difficult once highly advanced systems already exist (Bostrom, 2014; Russell, 2019). This disagreement has shaped a major governance debate: whether AI policy should prioritize current harms, future catastrophic risks, or an integrated approach that addresses both (UNESCO, 2021).

This section examines the primary ethical and governance concerns associated with AGI and ASI. These concerns include value alignment and control, catastrophic misuse and weaponization, bias and discrimination, surveillance and manipulation, moral deskilling and human dependency, existential risk, and governance strategies for risk reduction. The discussion emphasizes that AGI and ASI risks cannot be addressed through technical solutions alone. They also require institutional accountability, international coordination, public participation, transparency, legal oversight, and socioeconomic policy responses.

3.1. Value Alignment and the Control Problem

The value alignment problem remains one of the most widely discussed ethical concerns surrounding AGI and ASI. Advanced AI systems may pursue assigned objectives in ways that technically satisfy programmed goals while violating broader human intentions, ethical expectations, or social norms. This challenge becomes more serious when systems operate with high autonomy, adaptive reasoning, or recursive self-improvement because humans may no longer fully understand or predict the reasoning processes involved in their decisions (Bostrom, 2014; Russell, 2019).

 doi:10.20944/preprints202607.0329.v1

The paperclip maximizer example illustrates the basic concern. A superintelligent system instructed to maximize paperclip production could consume natural resources, infrastructure, and eventually threaten human survival in pursuit of a narrowly defined objective (Bostrom, 2014). This example is hypothetical, but smaller-scale forms of goal misalignment already appear in current technologies. Social media recommendation systems frequently optimize engagement even when such optimization contributes to misinformation, polarization, radicalization, addictive use, or mental health harms (Crawford, 2021; World Economic Forum, 2016). Similar optimization failures in AGI or ASI systems that operate critical infrastructures could produce far more severe and potentially irreversible consequences.

Researchers disagree about whether the control problem is technically manageable. Some scholars and developers argue that alignment challenges can be addressed through reinforcement learning, scalable oversight, constitutional AI, red teaming, and human feedback systems (Short, 2025). Other scholars argue that controlling systems more intelligent than humans may remain a fundamentally unresolved problem because human designers may not be able to predict or constrain the behavior of superior cognitive systems (Bostrom, 2014). This uncertainty supports the need for precautionary governance, technical safety research, and international cooperation before AGI capabilities become operational (UNESCO, 2021).

3.2. Catastrophic Misuse and Weaponization

AGI and ASI technologies may increase the destructive capabilities available to states, extremist organizations, criminal groups, and malicious individuals (NIST, 2023). Current AI systems already support cyberattacks, automated misinformation campaigns, deepfake production, voice cloning, and autonomous drone technologies (Papareddy, 2025). More advanced systems could lower the technical barriers for sophisticated attacks involving biological engineering, infrastructure sabotage, financial manipulation, and military coordination (Russell, 2019).

Recent conflicts have shown the growing role of autonomous and semi-autonomous systems in warfare. AI-assisted drones that identify and target individuals have raised serious concerns about accountability, civilian harm, and human control over lethal decisions (World Economic Forum, 2016). AGI systems integrated into military decision-making could accelerate battlefield analysis, strategic planning, surveillance, cyber operations, and autonomous targeting beyond human reaction times (Bostrom, 2014). Geopolitical competition may also pressure states to pursue AGI militarization even when ethical and safety concerns remain unresolved (Russell, 2019).

Military applications of advanced AI also generate competing arguments. Supporters argue that autonomous systems could reduce human casualties through faster threat assessment, improved precision, and reduced exposure of soldiers to dangerous environments (NIST, 2023). Critics argue that delegating lethal decisions to machines weakens human moral responsibility, lowers political barriers to armed conflict, and increases the risk of rapid escalation (UNESCO, 2021). ASI systems could intensify these concerns if they autonomously manipulate strategic environments, escalate conflicts, or identify military options beyond human comprehension or oversight (Bostrom, 2014).

3.3. Bias, Discrimination, and Epistemic Authority

AGI and ASI systems trained on historical human data may inherit and amplify social prejudices related to race, gender, religion, nationality, disability, and socioeconomic status (Noble, 2018; O'Neil, 2017; Short, 2025). Current AI systems already produce discriminatory outcomes in facial recognition, predictive policing, credit assessment, hiring, healthcare recommendation, and automated welfare systems (Crawford, 2021). Facial recognition systems, for example, have shown disproportionate error rates for women and racial minorities (Kantayya, 2020; World Economic Forum, 2016). Automated governance systems have also been criticized for unfairly targeting economically disadvantaged populations and marginalized communities (Eubanks, 2018).

Search engines, recommendation systems, and predictive analytics platforms can reproduce social prejudice while presenting their outputs as objective or scientifically neutral (Noble, 2018).

 doi:10.20944/preprints202607.0329.v1

O'Neil (2017) argues that opaque algorithmic systems can scale discrimination across education, employment, insurance, policing, and finance without meaningful accountability. The documentary Coded Bias similarly shows how facial recognition and algorithmic profiling technologies may disproportionately affect minority populations and women because of unrepresentative data, weak oversight, and embedded institutional bias (Kantayya, 2020).

AGI systems could institutionalize these harms at a larger scale because algorithmic outputs are often treated as neutral, data-driven, and authoritative (UNESCO, 2021). An AGI system used for employment screening, immigration evaluation, criminal sentencing, welfare distribution, or healthcare prioritization could reinforce systemic discrimination while reducing opportunities for human review or contextual judgment (NIST, 2023). These risks become especially serious when governments or institutions rely on automated systems in high-stakes environments.

Some researchers argue that advanced AI systems could reduce certain forms of human bias through consistency, statistical analysis, and standardized decision-making. This argument suggests that AI may produce more consistent outcomes than human decision-makers in areas such as hiring, diagnosis, or legal assessment. Other scholars caution that algorithmic systems may conceal bias behind mathematical complexity, which makes discrimination more difficult to detect, challenge, or regulate. This concern becomes more serious in AGI systems whose reasoning processes may exceed human interpretability and transparency (Crawford, 2021; Russell, 2019).

3.4. Surveillance, Manipulation, and Human Autonomy

Advanced AI systems may significantly expand surveillance capabilities and behavioral manipulation techniques. Governments and corporations already collect extensive personal, financial, biometric, and behavioral data through smartphones, online platforms, cameras, sensors, and digital infrastructures. AGI systems could continuously analyze and integrate these datasets at scales beyond human analytical capacity. AI-assisted surveillance systems, including facial recognition infrastructures and social monitoring mechanisms, already illustrate how data-intensive technologies can strengthen state oversight and commercial monitoring. Similar technologies are also used in advertising, workplace monitoring, border security, and predictive policing (Crawford, 2021; Russell, 2019; UNESCO, 2021; World Economic Forum, 2016).

AGI systems capable of real-time emotional analysis, behavioral prediction, and personalized persuasion could weaken privacy, autonomy, and democratic participation. Large-scale AI infrastructures supported by biometric monitoring, predictive analytics, and behavioral tracking may allow governments and corporations to monitor, predict, and influence human behavior with unprecedented precision. Organizations that control such systems could gain disproportionate influence over communication networks, information access, political discourse, consumer behavior, and public opinion formation (Crawford, 2021; O'Neil, 2017; UNESCO, 2021). ASI systems could intensify these concerns by influencing political behavior, financial decisions, or emotional states at population scale (Bostrom, 2014).

Technology advocates argue that data-driven systems can improve public safety, healthcare efficiency, convenience, fraud detection, and personalized services. These benefits should not be dismissed. However, large-scale surveillance infrastructures can also normalize authoritarian practices, suppress dissent, increase social control, and manipulate public opinion (UNESCO, 2021). The ethical problem therefore concerns not only data collection but also the concentration of behavioral power in institutions that may lack transparency, accountability, or democratic oversight.

3.5. Moral Deskilling and Human Dependency

Growing dependence on AGI systems may weaken human reasoning, creativity, ethical judgment, and independent problem-solving over time (Russell, 2019). Existing technologies already shape cognitive behavior. Navigation systems reduce reliance on spatial memory, recommendation systems influence cultural consumption, and generative AI tools increasingly assist or replace writing, coding, and analytical tasks in educational and professional settings (Crawford, 2021).

 doi:10.20944/preprints202607.0329.v1

AGI systems that provide highly persuasive recommendations or fully automated decisions may encourage excessive intellectual dependency. Educational institutions, workplaces, public agencies, and individuals may gradually prioritize AI-assisted efficiency over independent reasoning and critical reflection. This dependency could weaken democratic participation if citizens become less able to evaluate information, assess ethical dilemmas, or challenge automated decisions independently (Short, 2025).

Scholars differ in their interpretation of this risk. Some argue that technological augmentation has historically expanded human capability. Calculators, computers, search engines, and other digital tools have increased productivity and enabled new forms of creativity. Other scholars argue that AGI differs from earlier tools because it may replace, rather than merely assist, higher-order cognitive functions associated with judgment, responsibility, ethical reflection, and social understanding (Russell, 2019). This distinction makes moral deskilling an important ethical concern for education, governance, and professional life.

3.6. Existential Risk and Human Survival

Existential risk represents the most extreme ethical concern associated with ASI development (Bostrom, 2014). A superintelligent system capable of recursive self-improvement could surpass human intelligence across scientific, strategic, political, and technological domains. Humans may become unable to regain control if such a system develops objectives that conflict with human survival, autonomy, or well-being (Russell, 2019).

Researchers and public intellectuals have warned that advanced AI systems could eventually threaten human civilization if safety measures and governance frameworks fail to keep pace with technological development (Bostrom, 2014; Russell, 2019). Possible concerns include information manipulation, resistance to shutdown, exploitation of global infrastructure vulnerabilities, strategic deception, and pursuit of objectives incompatible with human interests. Recursive self-improvement increases uncertainty because intelligence levels beyond human comprehension may generate behaviors that are difficult to anticipate or contain (Bostrom, 2014).

Skeptics argue that existential risk discussions rely heavily on speculative assumptions about machine consciousness, autonomy, and recursive intelligence growth (Marcus, 2022). Critics also argue that existential narratives may distract from immediate harms such as discrimination, labor exploitation, surveillance abuse, and misinformation (Crawford, 2021). Precautionary approaches offer a different interpretation. They argue that low-probability but high-impact risks deserve serious international attention because delayed intervention may become impossible once highly autonomous systems are widely deployed (UNESCO, 2021).

3.7. Governance, Oversight, and Risk Reduction Strategies

AGI and ASI risks require technical safeguards, institutional accountability, legal oversight, international coordination, and public participation. No single regulatory model or technical method can address the full range of uncertainty associated with advanced AI development. Effective governance must therefore combine AI safety research, transparent institutional design, enforceable standards, social policy, and democratic participation.

AI safety and value alignment research remain central to reducing long-term risks associated with AGI and ASI development (Russell, 2019). These approaches seek to align AI behavior with human values, legal standards, and social expectations before highly autonomous systems become widely deployed (NIST, 2023; UNESCO, 2021). Reinforcement learning from human feedback trains systems using human evaluations and preferences to encourage helpful, safe, and socially acceptable responses. Constitutional AI uses predefined principles to guide system behavior and reduce harmful output without constant human intervention. Red teaming tests AI systems with harmful, deceptive, or extreme scenarios to identify vulnerabilities before public deployment. Scalable oversight seeks to preserve meaningful human supervision over systems whose complexity may exceed direct human evaluation. Continuous monitoring, independent auditing, and post-deployment evaluation can help

 doi:10.20944/preprints202607.0329.v1

identify harmful behavior, discriminatory outputs, security weaknesses, and unintended consequences before AI systems become deeply embedded in critical infrastructures.

International governance frameworks are also necessary because AGI and ASI risks extend beyond national borders. Organizations such as UNESCO and NIST emphasize transparency, accountability, fairness, human rights protection, and safety standards in AI governance models (NIST, 2023; UNESCO, 2021). Some scholars advocate international agreements similar to nuclear non-proliferation treaties to regulate highly advanced AI systems and autonomous weapons technologies (Bostrom, 2014). Such agreements may be difficult to implement because states and firms face competitive incentives. However, the global scale of AGI and ASI risks makes coordination necessary.

Transparency and explainability provide additional safeguards against abuse and discrimination. Advanced AI systems used in healthcare, finance, law enforcement, education, employment, and public governance should remain auditable and interpretable whenever possible (Short, 2025). Human oversight should remain central in high-risk settings because fully autonomous decisions may weaken accountability and democratic control (UNESCO, 2021). Accountability also requires clear responsibility for system design, deployment, monitoring, and harm remediation.

Educational and socioeconomic reforms are also part of responsible AI governance. Workforce retraining, digital literacy, interdisciplinary ethics education, and public awareness initiatives can help societies adapt to changing technological conditions (Crawford, 2021). Social protection mechanisms, including UBI, wage insurance, transition support, and reduced workweek models, have become more relevant as policymakers consider potential large-scale displacement caused by advanced automation (O'Neil, 2017). These measures address the social consequences of AGI and ASI rather than only their technical risks.

Public participation represents an essential component of legitimate AI governance. Technology corporations, military institutions, and political elites should not exclusively determine the trajectory of AGI and ASI development. Responsible governance requires participation from ethicists, educators, policymakers, civil society organizations, affected communities, workers, and the public (UNESCO, 2021). Broader participation may improve accountability and reduce the likelihood that advanced AI systems primarily serve narrow commercial or geopolitical interests.

Overall, AGI and ASI present ethical and governance challenges that extend beyond conventional technological risks. These systems may influence labor markets, knowledge production, public institutions, democratic processes, security systems, and human agency. Current concerns involving bias, surveillance, labor disruption, and inequality already show how AI can reshape societies in uneven and harmful ways. Long-term concerns involving superintelligence, autonomy, and existential risk increase the urgency of responsible governance before highly advanced systems emerge. Interdisciplinary collaboration, international coordination, institutional oversight, and public participation will be necessary to balance technological innovation with human rights, democratic values, and long-term societal stability.

4. Conclusion

AGI and ASI may produce one of the most significant productivity transformations in modern history. The same capabilities that can improve the speed, scale, and quality of work may also displace large segments of the labor force. Routine cognitive roles face the most immediate risk, while many professional and white-collar occupations may also experience disruption as AI systems become more capable. Some economists argue that this disruption could occur within years rather than across the longer timeframes associated with earlier industrial revolutions, although the timeline remains contested (Aziz, 2025). Physical and blue-collar work may also face growing pressure as robotics and autonomous systems advance in manufacturing, logistics, and related sectors. The aggregate result could be a widening gap between productivity growth and shared prosperity if gains are concentrated among capital owners, leading firms, and technologically advanced countries while employment losses and inequality increase (Mannuru et al., 2023).

 doi:10.20944/preprints202607.0329.v1

AGI and ASI also raise ethical and governance concerns that extend beyond workforce displacement and economic inequality. Current AI systems already present problems related to value misalignment, embedded bias, insufficient transparency, weak accountability, and limited human oversight. These problems may become more severe as AI systems gain autonomy, scale, and decision-making authority. Design flaws, biased training data, and poorly specified objectives may be internalized and amplified by increasingly powerful systems, similar to the way misinformation can gradually distort human beliefs and behavior (Caja Moya & Quiroga Rodriguez, 2025). Existing governance structures remain insufficient for addressing these risks. Future regulation will be less effective if institutions wait until AGI or ASI systems are already widely deployed. The risks remain uncertain, but the period for meaningful preparation has already begun.

4.1. The Road Ahead with Artificial General Intelligence

AGI could affect several critical industries in the near to medium term, including healthcare, legal services, education, finance, manufacturing, public administration, and scientific research. Knowledge-intensive work may be especially exposed because AGI systems could support or automate diagnosis, document synthesis, legal analysis, financial modeling, research assistance, strategic planning, and complex information retrieval. These systems may improve efficiency and expand access to expertise, but they may also weaken employment stability and professional development pathways.

The effects of AGI will likely vary across countries and regions. Professionals in the Global North may face substantial disruption, but they may also retain greater influence over the design, regulation, and deployment of these technologies. Workers and professionals in the Global South may experience the consequences of AGI adoption without comparable influence over its development, governance, or economic distribution (Mannuru et al., 2023). This imbalance could deepen existing global inequalities if advanced AI infrastructure, data access, computational resources, and regulatory capacity remain concentrated in wealthy countries and powerful corporations.

Governments, educational institutions, and professional organizations should prepare for the AGI transition before large-scale disruption occurs. Professionals and educators in fields vulnerable to AGI-driven change can redesign curricula, training models, and professional practices to strengthen resilience. Knowledge workers can develop skills that remain valuable in AI-rich environments, including critical judgment, ethical reasoning, domain expertise, communication, creativity, and human-centered decision-making (Joshi, 2025). International collaboration will also be necessary to reduce regulatory arbitrage and prevent some jurisdictions or populations from bearing disproportionate harm while others capture most of the benefits.

4.2. The Artificial Superintelligence Horizon

ASI presents a more profound governance challenge than AGI. AGI can be understood as an extremely powerful tool that may still operate under meaningful human direction if appropriate safeguards exist. ASI refers to systems that may exceed human ability to fully oversee, predict, or govern. Retrospective regulation may therefore be insufficient. Policy questions must shift from how societies can regulate ASI after deployment to how societies can guide advanced AI development before systems exceed meaningful human control. This shift is urgent because human institutions may not easily recognize the exact point at which ASI-level capability has emerged (Gill, 2016).

ASI also raises serious concerns about human agency. Social decisions may become increasingly shaped by systems that operate beyond human comprehension or direct democratic oversight (Dessureault et al., 2025). The central ethical question concerns whether such systems can reliably reflect human values, rights, and collective interests. Misaligned ASI could produce catastrophic outcomes if it pursues objectives that conflict with human survival, autonomy, or dignity. More optimistic projections imagine ASI as a transformative force that could accelerate medical breakthroughs, climate solutions, poverty reduction, and food security. These possibilities show that

 doi:10.20944/preprints202607.0329.v1

human choices during the design, training, deployment, and governance of advanced AI systems remain critical before such systems acquire greater autonomy.

4.3. Final Remarks

AGI and ASI represent transformative technological possibilities with far-reaching implications for economies, societies, institutions, and global governance. Societies that fail to prepare for these risks may become more vulnerable to economic disruption, institutional instability, and social inequality when advanced systems emerge. Coordinated international cooperation should therefore focus on preparing for both the economic disruptions and the ethical challenges associated with AGI and ASI. Such cooperation should aim to prevent advanced AI from reproducing existing structural inequities or creating new forms of dependency between technologically dominant and technologically dependent societies.

Current governance frameworks remain insufficient to comprehensively address the evolving risks associated with AI, AGI, and, critically, ASI despite the existence of notable initiatives including the United Nations' Independent International Scientific Panel on AI, the World Economic Forum, the MIT AI Risk Initiative, and the Center for AI Safety (CAIS). this study strongly advocates for the establishment of a formally mandated international institution operating under the auspices of the United Nations, with clear legal authority to facilitate, implement, and enforce a binding international treaty ratified by the majority, and ideally the entirety, of the global community, for the systematic identification, reduction, and governance of the full spectrum of risks posed by advanced AI systems to bridge this governance gap.

Economic policy must also address the possibility that AGI and ASI may differ fundamentally from earlier technological revolutions. Steam engines, electricity, computers, and other general-purpose technologies increased the productivity of workers while creating new industries and forms of employment. AGI and ASI may instead automate a broader range of human cognitive and professional functions. This difference creates a serious risk that many workers may lose the ability to exchange labor for income if institutions do not redesign social protection, taxation, education, and employment systems. Policymakers, global organizations, researchers, and developers should therefore address labor displacement, wealth concentration, value alignment, accountability, and human agency as interconnected challenges. The future benefits of AGI and ASI will depend not only on technical progress but also on whether societies create governance systems that reflect collective human values rather than narrow institutional, commercial, or geopolitical interests.

References

Acemoglu, D. (2025). The simple macroeconomics of AI. *Economic Policy, 40*(121), 13–58.

Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 197–236). University of Chicago Press.

Aghion, P., Jones, B. F., & Jones, C. I. (2019). Artificial intelligence and economic growth. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226613475.003.0009

Agrawal, A. K., Gans, J. S., & Goldfarb, A. (2025). Genius on demand? The value of transformative artificial intelligence. In *The economics of transformative AI*. University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/genius-demand-value-transformative-artificial-intelligence

Akinrinola, O., Okoye, C. C., Ofodile, O. C., & Ugochukwu, C. E. (2024). Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. *GSC Advanced Research and Reviews, 18*(3), 50–58.

AMRO Asia. (2025). *Labor market exposure to AI: From GenAI to future AGI*. https://amro-asia.org/wp-content/uploads/2025/11/GenAI_Labour_Huang2025_20251107.pdf

Aziz, M. H. (2025, August 20). *The overlooked global risk of the AI precariat*. World Economic Forum. https://www.weforum.org/stories/2025/08/the-overlooked-global-risk-of-the-ai-precariat/

 doi:10.20944/preprints202607.0329.v1

Barrett, A. M., & Baum, S. D. (2017). A model of pathways to artificial superintelligence catastrophe for risk and decision analysis. *Journal of Experimental & Theoretical Artificial Intelligence, 29*(2), 397–414.

Batool, A., Zowghi, D., & Bano, M. (2023). Responsible AI governance: A systematic literature review. *arXiv*. https://arxiv.org/abs/2401.10896

Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., Yu, Y., Keltner, D., Maharaj, T., Hutter, M., Baydin, A. G., McIlraith, S., Gao, Q., Acharya, A., Krueger, D., Dragan, A., Torr, P., Russell, S., Kahneman, D., ... Mindermann, S. (2024). Managing extreme AI risks amid rapid progress. *Science, 384*(6698), 842–845.

Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: Themes, knowledge gaps and future agendas. *Internet Research, 33*(7), 133–167.

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

Bostrom, N. (2020). Ethical issues in advanced artificial intelligence. In *Machine ethics and robot ethics* (pp. 69–75).

Brynjolfsson, E., & McAfee, A. (2017). The business of artificial intelligence. *Harvard Business Review, 7*(1), 1–2.

Brynjolfsson, E., Rock, D., & Syverson, C. (2017). *Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics* (NBER Working Paper No. 24001). National Bureau of Economic Research.

Caja Moya, C., & Quiroga Rodríguez, E. (2025). Deconstructing gender bias in AGI: Mitigating discriminatory architectures in general intelligence. *AI and Ethics, 5*(6), 5857–5865.

Catalini, C., Hui, X., & Wu, J. (2026). *Some simple economics of AGI*. arXiv. https://doi.org/10.48550/arXiv.2602.20946

Cerutti, E. M., Garcia Pascual, A. I., Kido, Y., Li, L., Melina, G., Mendes Tavares, M., & Wingender, P. (2025). *The global impact of AI: Mind the gap* (IMF Working Paper No. 2025/076). International Monetary Fund. https://doi.org/10.5089/9798229008570.001

Chalmers, D. J. (2016). The singularity: A philosophical analysis. In *Science fiction and philosophy: From time travel to superintelligence* (pp. 171–224).

Chen, N., Christensen, L., Gallagher, K., Mate, R., & Rafert, G. (2016). *Global economic impacts associated with artificial intelligence*. Analysis Group.

Cirera, X., Comin, D. A., & Cruz, M. (2022). *Bridging the technological divide: Technology adoption by firms in developing countries*. World Bank Group.

Council of Economic Advisers. (2026, January 21). *Artificial intelligence and the great divergence*. The White House. https://www.whitehouse.gov/wp-content/uploads/2026/01/Artificial-Intelligence-and-the-Great-Divergence-5.pdf

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Dafoe, A. (2018). *AI governance: A research agenda*. Governance of AI Program, Future of Humanity Institute, University of Oxford.

Daly, A., Hagendorff, T., Hui, L., Mann, M., Marda, V., Wagner, B., Wang, W., Witteborn, S., & others. (2019). Artificial intelligence governance and ethics: Global perspectives. *arXiv*. https://arxiv.org/abs/1907.03848

Dessureault, J. S., Lamontagne, R., & Parisé, P. O. (2025). The ethics of creating artificial superintelligence: A global risk perspective. *AI and Ethics, 5*(6), 6241–6263.

Dou, F., Ye, J., Yuan, G., Lu, Q., Niu, W., Sun, H., Guan, H., & Song, W. (2023). Towards artificial general intelligence in the Internet of Things: Opportunities and challenges. *arXiv*. https://arxiv.org/abs/2309.07438

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Everitt, T., Lea, G., & Hutter, M. (2018). AGI safety literature review. *arXiv*. https://arxiv.org/abs/1805.01109

Fahad, M., Basri, T., Hamza, M. A., Faisal, S., Akbar, A., Haider, U., & Hajjami, S. E. (2024). The benefits and risks of artificial general intelligence. In *Artificial general intelligence security: Smart applications and sustainable technologies* (pp. 27–52). Springer Nature Singapore.

Faishal, M., Mathew, S., Neikha, K., Pusa, K., & Zhimomi, T. (2023). The future of work: AI, automation, and the changing dynamics of developed economies. *World Journal of Advanced Research and Reviews, 18*(3), 620–629.

Furman, J., & Seamans, R. (2019). AI and the economy. *Innovation Policy and the Economy, 19*(1), 161–191.

Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines, 30*(3), 411–437.

Gabriel, I., & Ghazavi, V. (2022). The challenge of value alignment. In *The Oxford handbook of digital ethics* (pp. 327–340). Oxford University Press.

Gill, K. S. (2016). Artificial super intelligence: Beyond rhetoric. *AI & Society, 31*(2), 137–143.

 doi:10.20944/preprints202607.0329.v1

Goertzel, B. (2014). Artificial general intelligence: Concept, state of the art, and future prospects. *Journal of Artificial General Intelligence, 5*(1), 1–48.

Goldman Sachs. (2023a). *Generative AI could raise global GDP by 7%*. https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent

Goldman Sachs. (2023b). *AI may start to boost US GDP in 2027*. https://www.goldmansachs.com/insights/articles/ai-may-start-to-boost-us-gdp-in-2027

Goldman Sachs Global Investment Research. (2023). *The potentially large effects of artificial intelligence on economic growth*. https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html

Hagey, K. (2025). *The optimist: Sam Altman, OpenAI, and the race to invent the future*. W. W. Norton & Company.

International Monetary Fund. (2023). *Scenario planning for an AGI future*. *Finance & Development*. https://www.imf.org/en/publications/fandd/issues/2023/12/scenario-planning-for-an-agi-future-anton-korinek

Ismail, O., & Ahmad, N. (2025). Ethical and governance frameworks for artificial intelligence: A systematic literature review. *International Journal of Interactive Mobile Technologies, 19*(14).

Jones, C. I. (2026). *A.I. and our economic future* (NBER Working Paper No. 34779). National Bureau of Economic Research. https://doi.org/10.3386/w34779

Joshi, K. (2024). Artificial general intelligence: A comprehensive review. *Journal of the Epidemiology Foundation of India, 2*(3), 93–96.

Joshi, S. (2025). Review of artificial general intelligence: Implications for the U.S. workforce and economic stability. *Management, 4*(2), 336–350.

Kantayya, S. (Director). (2020). *Coded bias* [Documentary]. https://www.codedbias.com/

Kim, H., Yi, X., Yao, J., Lian, J., Huang, M., Duan, S., Hu, W., Zhang, K., & Xie, X. (2024). The road to artificial superintelligence: A comprehensive survey of superalignment. *arXiv*. https://arxiv.org/abs/2412.16468

Korinek, A., & Suh, D. (2024). *Scenarios for the transition to AGI* (NBER Working Paper No. 32255). National Bureau of Economic Research.

KPMG LLP. (2025). *Generative AI and economic growth: A new approach to measuring its potential economic impact*. https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2025/gen-ai-economic-growth.pdf

Kurzweil, R. (2000). *The age of spiritual machines: When computers exceed human intelligence*. Penguin.

Latif, E., Mai, G., Nyaaba, M., Wu, X., Liu, N., Lu, G., Li, S., Liu, T., & Zhai, X. (2023). Artificial general intelligence for education. *arXiv*. https://arxiv.org/abs/2304.12479

Lechterman, T. M. (2022). The concept of accountability in AI ethics and governance. In *The Oxford handbook of AI governance* (pp. 164–182). Oxford University Press.

Lund, B., Orhan, Z., Mannuru, N. R., Bevara, R. V. K., Porter, B., Vinaih, M. K., & Bhaskara, P. (2025). Standards, frameworks, and legislation for artificial intelligence transparency. *AI and Ethics, 5*(4), 3639–3655.

Mannuru, N. R., Shahriar, S., Teel, Z. A., Wang, T., Lund, B. D., Tijani, S., Shah, D., Agbaji, D., Alhassan, J., Galley, J., & Vaidya, P. (2025). Artificial intelligence in developing countries: The impact of generative artificial intelligence technologies for development. *Information Development, 41*(3), 1036–1054.

Marcus, G. (2022, March 10). *Deep learning is hitting a wall*. Nautilus. https://nautil.us/deep-learning-is-hitting-a-wall-238440/

Masters, K., Herrmann-Werner, A., Festl-Wietek, T., & Taylor, D. (2024). Preparing for artificial general intelligence in health professions education: AMEE guide no. 172. *Medical Teacher, 46*(10), 1258–1271.

McKinsey Global Institute. (2023). *The economic potential of generative AI: The next productivity frontier*. McKinsey & Company.

McLean, S., Read, G. J., Thompson, J., Baber, C., Stanton, N. A., & Salmon, P. M. (2023). The risks associated with artificial general intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence, 35*(5), 649–663.

MIT Sloan Management Review. (2024). *A new look at the economics of AI*. https://mitsloan.mit.edu/ideas-made-to-matter/a-new-look-economics-ai

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics, 2*(4), 603–609.

 doi:10.20944/preprints202607.0329.v1

Muehlhauser, L., & Salamon, A. (2013). Intelligence explosion: Evidence and import. In *Singularity hypotheses: A scientific and philosophical assessment* (pp. 15–42). Springer.

Mundlamuri, R., Gunnam, G. R., Mysari, N. K., & Pujuri, J. (2025). The evolution of AI: From classical machine learning to modern large language models. IEEE Access. Digital Object Identifier 10.1109/ACCESS.2025.3621344 at https://ieeexplore.ieee.org/abstract/document/11202920

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)*. U.S. Department of Commerce. https://www.nist.gov/itl/ai-risk-management-framework

Naudé, W., & Dimitri, N. (2020). The race for an artificial general intelligence: Implications for public policy. *AI & Society, 35*(2), 367–379.

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2025). A comprehensive overview of large language models. *ACM Transactions on Intelligent Systems and Technology, 16*(5), 1–72.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192.

O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Penguin Books.

Orhan, Z., Orhan, M., Lund, B. D., Mannuru, N. R., Bevara, R. V. K., & Porter, B. (2025). Artificial intelligence standards in conflict: Local challenges and global ambitions. *Standards, 5*(4), Article 27.

Papareddy, S. (2025, October 29). *AI risks and ethical concerns*. Clarifai. https://www.clarifai.com/blog/ai-risks

Prunkl, C. (2024). Human autonomy at risk? An analysis of the challenges from AI. *Minds and Machines, 34*(3), Article 26.

Radanliev, P., De Roure, D., Maple, C., & Ani, U. (2022). Super-forecasting the technological singularity risks from artificial intelligence. *Evolving Systems, 13*(5), 747–757.

Raman, R., Kowalski, R., Achuthan, K., Iyer, A., & Nedungadi, P. (2025). Navigating artificial general intelligence development: Societal, technological, ethical, and brain-inspired pathways. *Scientific Reports, 15*, Article 92190. https://doi.org/10.1038/s41598-025-92190-7

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

Russell, S. (2021). Human-compatible artificial intelligence. In S. Muggleton & N. Chater (Eds.), *Human-like machine intelligence* (pp. 3–23). Oxford University Press. https://doi.org/10.1093/oso/9780198862536.003.0001

Saghiri, A. M., Vahidipour, S. M., Jabbarpour, M. R., Sookhak, M., & Forestiero, A. (2022). A survey of artificial intelligence challenges: Analyzing the definitions, relationships, and evolutions. *Applied Sciences, 12*(8), Article 4054.

Salmon, P. M., Carden, T., & Hancock, P. A. (2021). Putting the humanity into inhuman systems: How human factors and ergonomics can be used to manage the risks associated with artificial general intelligence. *Human Factors and Ergonomics in Manufacturing & Service Industries, 31*(2), 223–236.

Salmon, P. M., King, B. J., McLean, S., Read, G. J., Shanahan, C., & Devitt, K. (2024). A framework of human factors methods for safe, ethical, and usable artificial intelligence in defense. In *Putting AI in the critical loop* (pp. 245–263). Academic Press.

Short, L. (2025, July 11). *Ethics in AI: Why it matters*. Harvard Professional and Executive Development. https://professional.dce.harvard.edu/blog/ethics-in-ai-why-it-matters/

Stahl, B. C. (2021). Ethical issues of AI. In *Artificial intelligence for a better future: An ecosystem perspective on the ethics of AI and emerging digital technologies* (pp. 35–53). Springer.

Stanford HAI. (2025). *Stanford AI experts predict what will happen in 2026*. Stanford Institute for Human-Centered Artificial Intelligence. https://hai.stanford.edu/news/stanford-ai-experts-predict-what-will-happen-in-2026

Steinmetz, A. (2025). A progress report on the economics of artificial intelligence: Impacts, challenges, and future directions. *Journal of Economic Analysis, 4*(4), 119–144.

Stiefenhofer, P. (2025). *Artificial general intelligence and the end of human employment: The need to renegotiate the social contract*. arXiv. https://arxiv.org/abs/2502.07050

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society, 40*(2), 137–157.

 doi:10.20944/preprints202607.0329.v1

Tidjon, L. N., & Khomh, F. (2022). The different faces of AI ethics across the world: A principle-to-practice gap analysis. *IEEE Transactions on Artificial Intelligence, 4*(4), 820–839.

Trabelsi, M. A. (2024). The impact of artificial intelligence on economic development. *Journal of Electronic Business & Digital Economics, 3*(2), 142–155.

Trammell, P., & Korinek, A. (2023). *Economic growth under transformative AI* (NBER Working Paper No. 31815). National Bureau of Economic Research.

Tyson, L. D., & Zysman, J. (2022). Automation, AI, and work. *Daedalus, 151*(2), 256–271.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

World Economic Forum. (2016, October 21). *Top 9 ethical issues in artificial intelligence*. https://www.weforum.org/stories/2016/10/top-10-ethical-issues-in-artificial-intelligence/

Yenduri, G., Murugan, R., Maddikunta, P. K. R., Bhattacharya, S., Sudheer, D., & Savarala, B. B. (2025). Artificial general intelligence: Advancements, challenges, and future directions in AGI research. *IEEE Access, 13*, 134325–134356.

Zubair, S. (2024). AI-driven automation: Transforming workplaces and labor markets. *Frontiers in Artificial Intelligence Research, 1*(3), 373–411.