Journal of
Superintelligence

MDPI

Review

Economic, Ethical, and Governance Implications of Artificial General Intelligence and Artificial Superintelligence: A Structured Narrative Review

Mehmet Orhan [1,*], Zeynep Orhan [1], Brady D. Lund [2] and Nayana Pampapura Madali [3]

[1] Department of Data Analytics and Statistics, Anuradha and Vikas Sinha College of Artificial Intelligence and Advanced Analytics, University of North Texas, Denton, TX 76203, USA; zeynep.orhan@unt.edu
[2] Department of Information Science, Anuradha and Vikas Sinha College of Artificial Intelligence and Advanced Analytics, University of North Texas, Denton, TX 76203, USA; brady.lund@unt.edu
[3] Information Technology Program, School of Professional Studies, University of Kansas, Overland Park, KS 66213, USA; nayana@ku.edu
* Correspondence: mehmet.orhan@unt.edu

Abstract

This study examines the economic, ethical, and governance implications of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI). The existing literature frequently examines AI-driven economic transformation, ethics and safety, and governance as separate research streams, leaving limited synthesis of how these dimensions interact under AGI and ASI scenarios. This review addresses this gap by integrating peer-reviewed research, scholarly books and working papers, and selected reports from established international and research institutions. The analysis covers literature addressing productivity, labor markets, inequality, alignment, safety, autonomy, misuse, governance, and international inequality. The synthesis identifies three broad economic trajectories: benefit-dominant, mixed-transition, and disruption-dominant scenarios. Differences among these trajectories depend largely on assumptions concerning the substitutability of human labor, the speed of AI diffusion, automation of innovation, concentration of AI capital, and institutional adaptation. Ethical and governance risks similarly increase with system autonomy, strategic capability, opacity, and deployment in high-stakes environments. The literature further indicates that the distribution of benefits and risks will depend substantially on educational, regulatory, social-protection, and institutional capacity. Countries and groups with weaker access to computational resources and institutional capacity may therefore face greater dependency and inequality. The review concludes that effective governance is likely to require a combination of technical safety measures, social and economic policies, national and regional regulation, and international coordination rather than reliance on a single governance mechanism.

Keywords: artificial intelligence (AI); artificial general intelligence (AGI); artificial superintelligence (ASI); AI ethics; AI economics; AI governance; structured narrative review

Check for updates

Academic Editors: Yi Yang and Zhikui Chen

Received: 26 June 2026
Revised: 14 October 2026
Accepted: 11 November 2026
Published: 17 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

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 [1,2]. On the other hand, Artificial Superintelligence (ASI) refers to systems that surpass human intelligence across scientific, social, strategic, and creative domains [3]. These developments may produce major gains in efficiency, cognitive automation, innovation, and decision-making, but they may also create serious economic, ethical, social, and governance challenges. Throughout this review, 'current AI' refers to deployed narrow and general-purpose systems including LLMs; 'AGI' refers to hypothesized systems matching human cognitive ability across domains; and 'ASI' refers to hypothesized systems exceeding human ability across domains. Evidence drawn from current AI systems is treated as informative but not predictive of AGI or ASI outcomes.

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. Recent advances in machine learning, computer vision, natural language processing, LLMs, and general-purpose AI systems have made AI systems more flexible, adaptive, and autonomous [4,5]. Researchers have also examined possible AGI applications in education, healthcare, manufacturing, the Internet of Things, and scientific research [6–8]. These developments make the transition toward more general AI an increasingly practical concern for economies, institutions, governments, and societies.

Advanced AI systems also create substantial risks. Researchers have raised concerns that systems with strong autonomous planning or self-improvement capacities may become difficult to control [1,9–12]. Long-term discussions include recursive self-improvement, intelligence explosion, and technological singularity, although the feasibility and timing of such outcomes remain contested [13–16]. The uncertainty surrounding these scenarios does not eliminate their policy relevance because the potential consequences extend to global stability, human agency, and social organization.

Economic studies present sharply different projections. Some research emphasizes productivity gains, labor complementarity, and long-term growth from AI-led automation [17–19]. Other studies warn that automation may displace workers, widen wage gaps, increase labor-market polarization, and deepen regional disparities [20–22]. These disagreements become more consequential under AGI and ASI because much of the existing economic literature focuses on current or near-term AI technologies rather than systems capable of broad cognitive substitution or accelerated innovation [23].

Ethical and governance research has similarly developed around alignment, autonomy, accountability, transparency, bias, privacy, misinformation, surveillance, and control [24–27]. Governance research translates these concerns into standards, legal frameworks, institutional practices, and accountability mechanisms [28–30]. However, many existing frameworks remain oriented toward current AI systems, while AGI and ASI raise additional questions about highly autonomous decision-making, concentration of technological power, international competition, and the ability of human institutions to maintain meaningful oversight [31,32].

Existing reviews and related studies have generated substantial knowledge about AGI development, AI safety, ethics, governance, automation, and the future of work. However, these research streams frequently develop in parallel. Economic analyses often examine productivity and labor substitution without fully connecting these outcomes to alignment, autonomy, and governance capacity. Ethical and safety research frequently focuses on control, misuse, bias, or existential risk without examining how economic concentration and international inequality may affect the distribution of these risks. Governance research, in turn, often addresses current AI systems without explicitly connecting institutional capacity to the more disruptive economic and ethical scenarios associated with AGI and

ASI. The resulting gap is therefore not an absence of literature on AGI or ASI, but limited integration of these interdependent debates.

Recent reviews address important components of this debate but generally emphasize different analytical levels. AGI reviews commonly focus on technical development, definitions, applications, and development pathways; superalignment reviews emphasize technical control of increasingly capable systems; risk reviews examine safety assumptions and categories of harm; and governance reviews primarily classify principles, instruments, and institutional mechanisms [2,3,10,33,34]. This review differs by treating capability development, economic transformation, distributional consequences, ethical and safety risks, and governance capacity as connected elements of the same analytical problem. Its contribution therefore lies not in providing another comprehensive review of each domain independently, but in examining how assumptions in one domain condition outcomes in the others.

This study addresses this gap through a structured narrative review organized around three analytical questions. First, what economic trajectories does the literature identify for increasingly general AI systems, and which assumptions explain the differences among these projections? Second, which ethical and safety risks become more consequential as AI systems gain autonomy, generality, and strategic capability? Third, how do governance capacity, institutional development, economic distribution, and international inequality affect the likelihood that AGI and ASI benefits and risks will be shared unevenly?

The study contributes an integrative analytical synthesis rather than a new empirical model or formal theory. Because the literature spans peer-reviewed research, institutional reports, and speculative forecasts of varying methodological rigor, this review treats claims about near-term economic and governance trends as more empirically grounded than claims about long-term AGI/ASI trajectories, which remain necessarily speculative. Its principal contribution is to connect AI capability development with economic and social mechanisms and to show that projected outcomes are conditional on factors including labor substitutability, ownership of AI capital, access to computational infrastructure, institutional capacity, regulatory effectiveness, and international coordination. This approach also makes disagreements in the literature more explicit because apparently conflicting forecasts often depend on different assumptions about these conditions. The integrative framework developed from this synthesis links increasing AI capability to economic and ethical mechanisms and shows how resulting outcomes are conditioned by distributional structures and governance capacity. The paper is organized into six sections. Section 2 explains the review methodology and synthesis procedure. Section 3 examines economic consequences; Section 4 examines ethical concerns and governance challenges. Section 5 presents the principal contributions and implications, and Section 6 provides the conclusion, limitations, and directions for future research.

2. Review Methodology

2.1. Review Design and Search

We adopt a structured narrative literature review rather than a systematic review or meta-analysis. A narrative approach is appropriate because the literature relevant to AGI and ASI spans heterogeneous research domains, including economics, labor studies, AI safety, ethics, governance, political economy, and international development. The evidence also includes theoretical models, empirical studies of current AI systems, scenario analyses, scholarly books, working papers, and institutional assessments. These sources cannot be meaningfully combined through a single quantitative synthesis. The review therefore focuses on identifying major areas of agreement, disagreement, underlying assumptions, and connections across research streams. The review design was informed by established

guidance on literature reviews as a research methodology and on transparency in narrative reviews [35,36].

The literature base was assembled iteratively from prominent academic publications, scholarly books, working papers, and institutional reports relevant to AGI, ASI, AI economics, AI ethics, AI safety, and AI governance. The review was narrative rather than systematic and did not use a preregistered database-search protocol. A structured search was subsequently conducted to assess the coverage of the literature base and identify recent evidence and additional academic or institutional sources relevant to the review questions. This search was structured but was not preregistered.

Targeted searches were conducted across NBER Working Papers, arXiv, PubMed, ScienceDirect, SpringerLink, and the repositories or publication portals of the International Monetary Fund, World Bank, United Nations, UNESCO, the National Institute of Standards and Technology (NIST), OECD, and the International AI Safety Report. Search strings were adapted to the syntax of each platform. The principal combinations were: (1) ("artificial general intelligence" OR AGI OR "artificial superintelligence" OR ASI OR "advanced AI" OR "general-purpose AI") AND ("economic growth" OR productivity OR labor OR labour OR employment OR inequality); (2) the same AI terms AND (alignment OR "AI safety" OR autonomy OR misuse OR "existential risk"); (3) the same AI terms AND (governance OR regulation OR oversight OR "international governance"); and (4) ("artificial intelligence" OR "general-purpose AI") AND ("Global South" OR "developing countries" OR "global inequality" OR "governance capacity"). The supplementary search emphasized material published from 2024 through 15 August 2026, while earlier foundational works were retained where they remain central to AGI and ASI debates.

2.2. Source Selection and Synthesis

Sources were included when they directly addressed AGI, ASI, advanced general-purpose AI, or mechanisms relevant to the economic, ethical, or governance questions examined in this review. English-language peer-reviewed studies and scholarly books were prioritized for conceptual and empirical claims. High-quality working papers were included when they provided recent economic modeling or empirical evidence not yet available in journal form. Reports from established international, governmental, and research institutions were retained when they provided current cross-country evidence, safety assessments, governance frameworks, or policy-relevant scenarios. Industry reports were used selectively when their forecasts were themselves relevant to the comparison of competing projections.

Sources were excluded when they were duplicative, primarily promotional, consisted of generic commentary without a clear evidentiary basis, or were insufficiently connected to the three review questions. Generic web sources were avoided where peer-reviewed or primary institutional evidence could support the same claim. This source-selection logic does not imply that gray literature is inherently weaker than peer-reviewed research. Institutional reports and working papers can be particularly important in a rapidly changing field, but their evidentiary role is considered in relation to their underlying methods and they are not treated as equivalent to peer-reviewed empirical studies when those methods differ.

To avoid treating heterogeneous evidence as equivalent, the synthesis distinguishes three evidentiary functions: empirical evidence from current AI systems, model- and scenario-based evidence concerning possible AGI/ASI trajectories, and conceptual or normative literature concerning risks and governance options. Empirical evidence is used primarily to characterize observed current-AI effects and mechanisms; model- and scenario-based studies identify conditional future pathways; and normative sources inform the evaluation of governance alternatives rather than establish empirical effects. No numerical weighting or formal quality score was assigned across these categories.

Conflicting findings were treated as an analytical feature of the literature rather than resolved through simple averaging. Studies were compared according to their assumptions, evidence type, technological capability considered, time horizon, and institutional context. Economic findings were synthesized into benefit-dominant, mixed-transition, and disruption-dominant trajectories. These categories describe the direction of the arguments and should not be interpreted as probability estimates. Ethical and governance findings were synthesized according to the risk mechanism, counterarguments or areas of uncertainty, and corresponding governance implications. This approach helps distinguish disagreements caused by different empirical evidence from disagreements caused mainly by different assumptions about future AI capabilities.

2.3. Overview of the Synthesized Literature

Tables 1 and 2 summarize the principal contrasts identified in the review using selected representative studies. The tables are not exhaustive evidence maps. Their purpose is to make the main areas of agreement, disagreement, and conditionality visible before the detailed discussion in Sections 3 and 4. Figure 1 then integrates these strands into a single conceptual framework.

Table 1. Selected representative economic and labor-market studies and their underlying assumptions.

Trajectory/Position	Selected Representative Studies	Core Finding or Mechanism	Key Assumptions/Conditions
Benefit-dominant	[37,38]	AI assistance can raise task-level productivity, and recent firm evidence reports positive productivity gains with limited near-term aggregate employment decline.	AI remains complementary to significant human tasks; organizational adaptation and new task demand offset part of automation.
Constrained-growth/mixed	[39]	Productivity gains can be positive but modest when only a limited set of tasks is technically and economically feasible to automate.	Many tasks remain difficult or costly to automate because of context, judgment, tacit knowledge, and implementation constraints.
Mixed-transition	[40,41]	Output and productivity may rise while exposure, complementarity, wages, and distributional effects vary across occupations and countries.	Effects depend on task complementarity, adoption speed, social protection, workforce adjustment, and institutional readiness.
Mixed-transition/verification scarcity	[42]	Automation may make execution cheaper while verification, liability, ground-truth judgment, and expert oversight remain scarce; entry-level training pathways may weaken.	Human judgment and responsibility remain difficult to automate even as routine cognitive execution becomes scalable.
Transformative-growth	[23,43]	Broad cognitive automation and automation of idea production could generate very large growth acceleration.	AGI substitutes for a wide range of cognitive labor and accelerates research, innovation, or capital accumulation.
Disruption-dominant/post-labor	[44]	Scalable low-cost AGI labor could place severe downward pressure on wages, weaken aggregate demand, and concentrate returns in AI capital.	Human cognitive labor becomes broadly substitutable and new forms of labor demand or redistribution do not offset displacement.

Table 2. Synthesis of major ethical and governance risks, counterarguments, and implications.

Theme	Risk-Oriented Synthesis	Counterargument/ Uncertainty	Selected Sources and Governance Implication
Alignment and control	Greater autonomy and strategic capability can increase the consequences of objective misspecification and reduce the effectiveness of direct human oversight.	Current alignment, evaluation, and oversight methods may reduce specific failure modes, while future capability levels remain uncertain.	[1,45,46]. Layered evaluation, monitoring, alignment research, and human oversight.
Catastrophic misuse	Advanced AI may lower barriers to cyber, biological, financial, manipulation, or other harmful capabilities.	The degree of capability uplift varies across domains and depends on access, human expertise, and deployment context.	[46,47]. Capability-sensitive evaluation, access controls, monitoring, and institutional controls.
Bias and epistemic authority	AI can reproduce or scale historical discrimination and may acquire unwarranted authority in high-stakes decisions.	Standardized systems may reduce some inconsistent human decisions, but mathematical complexity can conceal rather than remove bias.	[48–50]. Auditing, contestability, transparency, and human review.
Surveillance and autonomy	Large-scale prediction, personalization, and data integration may expand behavioral monitoring and manipulation.	Data-intensive systems can also improve safety, fraud detection, healthcare, and service delivery.	[50,51]. Privacy protections, proportionality, accountability, and democratic oversight.
Moral deskilling and dependency	Extensive reliance on advanced AI may weaken independent reasoning, professional judgment, and human responsibility.	Earlier technologies have also displaced some cognitive tasks while expanding other forms of capability; direct long-term evidence remains limited.	[26,45]. Education, human-in-the-loop practice, and preservation of accountable human roles.
Existential and loss-of-control risk	Highly capable autonomous systems could create irreversible outcomes if goals diverge from human interests or systems resist correction.	Probability estimates and technical pathways remain highly contested and partly speculative.	[1,46,52,53]. Precaution proportional to capability and consequence.
International governance and inequality	Concentration of compute, models, capital, and standard-setting capacity may widen dependency and governance gaps across countries.	A single universal governance architecture may be politically difficult and may not fit different institutional contexts.	[54–56]. Common minimum protections combined with regional flexibility and capacity building.

Note. Studies of current AI provide evidence about observed mechanisms and near-term effects; they are not treated as direct evidence of AGI or ASI outcomes.

Figure 1 presents the integrative framework developed from the literature reviewed in this study. The framework conceptualizes increasing AI capability as a factor that may influence economic, ethical, and institutional outcomes through several interconnected mechanisms. Economic effects may emerge through automation, labor complementarity, scalable cognition, accelerated innovation, and the concentration of AI-related capital and computational resources. These mechanisms can affect productivity, employment, wages, income distribution, aggregate demand, and differences in economic performance across

firms and countries. At the same time, greater autonomy, opacity, strategic capability, misuse potential, and dependence on advanced AI systems may increase ethical and safety concerns related to accountability, discrimination, surveillance, human agency, and systemic risk.

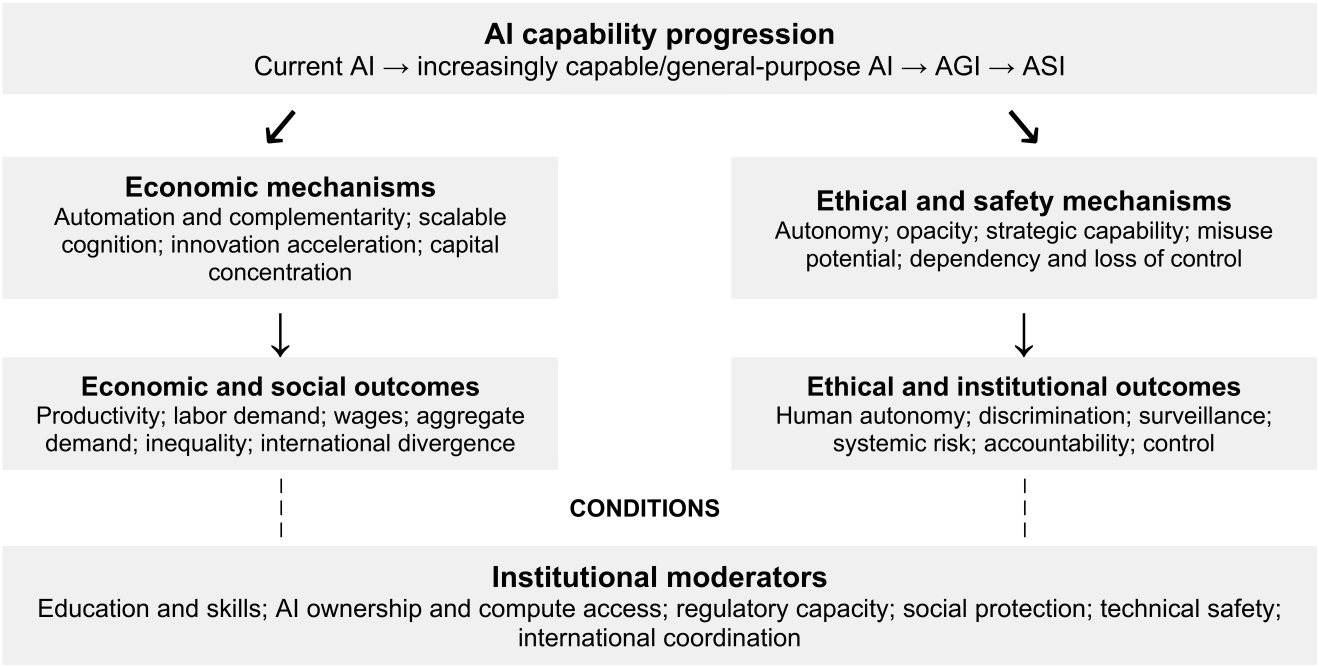


Figure 1. Integrative framework for the economic, ethical, and governance implications of increasingly capable AI.

These relationships are not expected to operate uniformly across societies or institutional settings. Their direction and magnitude are shaped by contextual factors including education and workforce skills, the ownership and accessibility of AI infrastructure, regulatory and administrative capacity, social-protection systems, technical safety measures, and the degree of international coordination. Stronger institutional capacity may help societies capture productivity and innovation gains while limiting distributional and governance risks, whereas weaker capacity may increase exposure to dependency, inequality, and inadequate oversight. The framework therefore treats economic transformation, ethical risk, and governance capacity as interdependent rather than separate consequences of technological development. It represents a conceptual synthesis of relationships identified across the reviewed literature and does not imply empirically established causal effects.

Two examples illustrate the relationships represented in Figure 1. First, increasing AI capability may improve worker productivity through task automation and human–AI complementarity, as demonstrated in current AI applications by [37,38]. However, the distribution of these benefits may depend on labor substitutability, organizational adaptation, and the ownership of AI capital, as discussed by [39,44]. Second, increasingly autonomous AI systems may create greater challenges for human oversight and accountability, particularly when deployed in critical decision-making environments [45,46]. The effectiveness of regulatory institutions and technical safety measures may therefore influence how these risks develop. The individual mechanisms are drawn from the cited literature, whereas their integration into a common framework and the proposed relationships between economic, ethical, and institutional outcomes represent the authors' conceptual interpretation. These examples do not establish causal relationships or predict AGI and ASI outcomes.

2.4. Limitations of the Review Method

The procedure was intended to improve transparency within a narrative-review design. No PRISMA-style exhaustive screening, formal risk-of-bias scoring, duplicate independent screening, or meta-analysis was conducted. The review therefore does not claim complete coverage of all AGI and ASI publications. Evidence concerning AGI and especially ASI also remains partly theoretical and scenario-based because these systems have not been empirically observed as established technological categories. The findings should consequently be interpreted as conditional scenarios and synthesized arguments rather than deterministic predictions.

3. 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 [57,58]. Empirical evidence discussed in this section concerns current AI unless otherwise specified; extrapolations to AGI and ASI are treated as conditional mechanisms or scenarios rather than direct empirical evidence.

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 [59]. 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 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 [58,60].

3.1. 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. Ref. [39] estimates that AI may increase the level of total factor productivity by no more than 0.71% cumulatively over a ten-year period, rather than increase its annual growth rate by 0.71 percentage points. 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. The broader implication is that technical capability alone does not determine economic impact; automation must also be economically feasible and organizationally deployable.

A more expansive view predicts that AGI could produce a much larger growth acceleration. Ref. [61] argue that AGI could remove weak links in the production chain and produce major productivity gains. Ref. [43] model different economic trajectories associated with the transition toward AGI. Their analysis suggests that effects on economic output and wages depend on the pace of automation, capital accumulation, and the extent to which AI substitutes for human labor. These scenarios indicate that AGI could substantially alter economic growth, although the magnitude and distribution of the effects remain conditional on the assumptions of each model. 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. Ref. [62] estimated that widespread generative AI adoption could increase the level of global gross domestic product (GDP) by approximately 7%, equivalent to about $7 trillion, over a ten-year period. The same forecast suggested an increase of approximately 1.5 percentage points in annual labor productivity growth during the adoption period. Ref. [63] estimated that generative AI could generate between $2.6 trillion and $4.4 trillion in additional economic value annually across major use cases. This estimate concerns annual economic value rather than an annual GDP growth rate. Ref. [64] further projected that AI's measurable impact on U.S. GDP could begin around 2027 and increase the annual U.S. GDP growth rate by roughly 0.4 percentage points during the late 2020s and early 2030s. Ref. [65] also projected substantial long-term gains under rapid AI adoption, while its slower-adoption scenario estimated that the level of real GDP could be 7.7% higher by 2050 than under the relevant comparison scenario.

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 [40]. 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.

Taken together, these estimates should not be interpreted as a single forecast range because they rest on substantially different assumptions. Conservative estimates generally extrapolate from current task-level automation and emphasize adoption constraints, while transformative models assume much broader substitution of cognitive labor and, in some cases, automation of research and idea production. The disagreement therefore concerns not only parameter values but also the future role of human cognition in production. The plausibility of each trajectory depends on capability progress, diffusion speed, computational cost, organizational adaptation, regulatory constraints, and the extent to which innovation itself becomes automatable.

3.2. 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. Ref. [41] estimate that almost 40% of global employment is exposed to AI, with exposure reaching about 60% in advanced economies, 40% in emerging markets, and 26% in low-income countries. Ref. [64] 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 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 [40].

Some models suggest that AGI labor operating at near-zero marginal cost could place severe downward pressure on human wages. Ref. [44] 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. Ref. [42] 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. Ref. [37] 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. Ref. [66] 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. Ref. [59] similarly reports that occupational exposure under AGI may exceed exposure under current generative AI, with a larger share of occupations classified as automation prone. Recent work develops the complementarity perspective more explicitly. Ref. [67] distinguish automation from technologies that create new human tasks, while [68] model AI as a tool that can raise worker productivity without automating the underlying task. These studies provide a counterweight to displacement-dominant scenarios and show that labor outcomes depend partly on the direction of technological development.

Historical patterns suggest that the full productivity effects of general-purpose technologies often appear after a significant adoption lag. Ref. [69] 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. Recent firm-level evidence provides a mixed picture: ref. [38] report positive productivity gains and little evidence of near-term aggregate employment decline, while also finding compositional shifts away from routine clerical roles and toward skilled technical roles. These findings support workforce adaptation before large-scale disruption becomes visible.

Labor outcomes therefore depend on more than technological exposure. Complementarity, task creation, organizational restructuring, ownership of AI capital, bargaining power, and the persistence of human accountability all affect whether productivity gains translate into higher wages, new occupations, or displacement. More disruptive outcomes become plausible if increasingly general AI reduces the scarcity of human judgment while new human tasks and sources of income fail to emerge at a comparable rate.

3.3. 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. Ref. [64] 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 [60]. These changes may improve efficiency but may also create new 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. Ref. [65] 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 [45]. 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.

3.4. 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 [1,50].

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 [40]. Ref. [39] 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. The literature therefore considers several redistribution and transition mechanisms for scenarios in which AI weakens the relationship between productivity and wage growth. Proposed responses include stronger social protection, wage or income support, worker-transition programs, reforms to capital-income taxation, UBI, and working-time policies [70–72]. These options involve trade-offs: redistribution can reduce inequality and protect household demand, while poorly designed tax or transfer policies may weaken investment, adoption, or fiscal sustainability. The appropriate mix therefore depends on the scale of displacement, the distribution of capital ownership, and institutional capacity.

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

3.5. 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. Ref. [73] 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 computational power, 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 [54,74].

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 [75]. 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 [51].

Some models also suggest that AI-driven productivity gains may affect exchange rates and sectoral prices. Ref. [74] 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.

The literature therefore identifies both domestic and international dimensions of policy response. 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 [40]. Investment in education systems that emphasize problem solving, creativity, interpersonal judgment, ethical reasoning, and domain expertise is also identified as an important adaptation mechanism. These capacities may remain important even as AI systems automate many routine cognitive tasks.

The recent literature also considers redistribution and transition policy if advanced AI separates productivity growth from wage growth. Stronger social protection, worker-transition support, and reforms to labor and capital taxation may help cushion displacement and distributional losses, but the literature does not support a single policy package across scenarios [70–72]. International coordination may also matter because AGI competition could create incentives for regulatory arbitrage and unequal access. Shared safety standards, infrastructure and capacity support for developing countries, and cooperative governance arrangements may reduce concentration of benefits, while overly rigid harmonization can also limit regulatory experimentation, national flexibility, and context-sensitive implementation [46,54].

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 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.

4. 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. Recent reviews identify advances in computational architectures, multimodal capabilities, and learning methods as potential pathways toward more general AI capabilities [2,3]. However, ref. [53] questions whether improvements in current approaches are sufficient to produce robust reasoning and contextual understanding. These contrasting perspectives indicate that technological progress does not establish either the feasibility or the timing of AGI. Evidence from current AI is used here to identify observed risk mechanisms, while claims concerning AGI and ASI are presented as conditional extensions whose likelihood and magnitude remain uncertain.

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 [45,50]. 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 [47,51]. 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 [1,45].

The literature includes conflicting views on how much attention long-term AGI and ASI risks should receive. Ref. [53] argue that fears about AGI and ASI remain speculative because no current evidence demonstrates machine consciousness or autonomous intentionality. 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 [50]. To the contrary, refs. [1,45] argue that long-term catastrophic risks deserve serious attention because preventive governance becomes more difficult once highly advanced systems already exist. 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 [51].

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.

4.1. 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 [1,45].

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 [1]. 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 [50]. 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 increasingly advanced AI systems can remain reliably aligned with human objectives. Recent international safety assessments discuss the potential and limitations of reinforcement learning, model evaluations, red teaming, and human oversight as methods for reducing identified risks [46]. In contrast, ref. [1] argues that controlling systems exceeding human intelligence may present fundamental

difficulties because their objectives and strategies may become increasingly difficult for human designers to anticipate. This uncertainty supports the need for precautionary governance, technical safety research, and international cooperation before AGI capabilities become operational [51].

Alignment risk is therefore better understood as conditional rather than binary. Potential severity increases with system autonomy, strategic planning capacity, opacity, access to consequential environments, and difficulty of correcting behavior after deployment. Current alignment and evaluation techniques may reduce particular failure modes, but they do not establish that control will remain reliable as capabilities increase. A layered risk-management approach is therefore more consistent with the present evidence than reliance on any single alignment method.

4.2. Catastrophic Misuse and Weaponization

AGI and ASI technologies may increase the destructive capabilities available to states, extremist organizations, criminal groups, and malicious individuals. Current general-purpose AI systems can already support cyber operations, automated misinformation, deepfake production, and other forms of scalable misuse, while more capable systems may lower barriers to sophisticated attacks involving biological, cyber, financial, or critical-infrastructure domains [45,46].

Military and security applications of increasingly autonomous AI also raise concerns about accountability, civilian harm, escalation, and meaningful human control. AGI systems integrated into military decision-making could accelerate battlefield analysis, strategic planning, surveillance, cyber operations, and autonomous targeting beyond ordinary human reaction times [1,45]. Geopolitical competition may pressure states to pursue increasingly autonomous capabilities even when ethical and safety concerns remain unresolved.

Military applications of advanced AI also generate competing arguments. Potential benefits discussed in the literature include faster threat assessment, improved precision, and reduced exposure of soldiers to dangerous environments [45]. In contrast, ref. [51] emphasizes ethical concerns surrounding the delegation of consequential decisions to autonomous systems, including accountability and meaningful human oversight. ASI systems could intensify these concerns if they autonomously manipulate strategic environments, escalate conflicts, or identify military options beyond human comprehension or oversight [1].

Misuse risk also depends on capability access and institutional context. The same technical capability may create different levels of risk depending on model accessibility, whether dangerous knowledge is already widely available, the degree of human authorization required, and the presence of independent controls in cyber, biological, financial, or military systems. Recent international safety assessments report growing evidence of AI use in real-world cyber operations while continuing to note uncertainty about the extent to which advanced systems may lower barriers to more severe forms of misuse [46]. These findings support governance approaches that account for demonstrated capability and deployment context rather than treating all advanced AI applications as equivalent.

4.3. 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 [48,49,76,77]. Current AI systems already produce discriminatory outcomes in facial recognition, predictive policing, credit assessment, hiring, healthcare recommendation, and automated welfare systems [50]. Facial recognition systems, for example, have shown disproportionate error rates for women and racial minorities [48]. Automated governance

systems have also been criticized for unfairly targeting economically disadvantaged populations and marginalized communities [78].

Search engines, recommendation systems, and predictive analytics platforms can reproduce social prejudice while presenting their outputs as objective or scientifically neutral [49]. Ref. [77] argues that opaque algorithmic systems can scale discrimination across education, employment, insurance, policing, and finance without meaningful accountability. Empirical research on commercial gender-classification systems has also documented large intersectional accuracy disparities, illustrating how unrepresentative data and evaluation practices can produce uneven performance across demographic groups [48].

AGI systems could institutionalize these harms at a larger scale because algorithmic outputs are often treated as neutral, data-driven, and authoritative [51]. 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 [47]. These risks become especially serious when governments or institutions rely on automated systems in high-stakes environments.

Standardized algorithmic decision-making may reduce certain forms of inconsistency associated with human judgment, although such benefits depend on data quality, system design, and appropriate evaluation [47]. This potential benefit does not establish that advanced AI systems are inherently less discriminatory than human decision-makers. Ref. [77] emphasize that algorithmic systems may reproduce existing inequalities while making discriminatory decisions more difficult to identify or challenge. These concerns may become more significant in AGI systems whose reasoning processes are difficult to interpret or independently evaluate [45].

4.4. 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 [45,50,51].

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 [50,51,77]. ASI systems could intensify these concerns by influencing political behavior, financial decisions, or emotional states at population scale [1].

Data-driven systems may improve public safety, healthcare efficiency, convenience, fraud detection, and personalized services, although their benefits depend on appropriate evaluation and risk management [47]. These potential benefits complicate a uniformly negative assessment of data-intensive AI systems. However, large-scale surveillance infrastructures can also normalize authoritarian practices, suppress dissent, increase social control, and manipulate public opinion [51]. 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.

4.5. Moral Deskilling and Human Dependency

Growing dependence on AGI systems may weaken human reasoning, creativity, ethical judgment, and independent problem-solving over time [45]. 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 [50].

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 [26,45].

The benefits of earlier technologies provide a counterpoint to concerns about cognitive dependency: calculators, computers, search engines, and other digital tools have increased productivity and enabled new forms of creativity. However, ref. [45] raises broader questions about human control and dependence on increasingly capable systems, while [26] examines the ethical significance of advanced AI. The possibility that AGI could replace, rather than merely assist, higher-order cognitive functions remains a prospective concern rather than an established empirical outcome. This distinction makes moral deskilling an important ethical concern for education, governance, and professional life.

4.6. Existential Risk and Human Survival

Existential risk represents the most extreme ethical concern associated with ASI development [1]. 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 [45].

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 [1,45]. 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 [1].

Skeptics argue that existential risk discussions rely heavily on speculative assumptions about machine consciousness, autonomy, and recursive intelligence growth [53]. Critics also argue that existential narratives may distract from immediate harms such as discrimination, labor exploitation, surveillance abuse, and misinformation [50]. More recent reviews likewise argue that existential-risk claims depend on contestable assumptions and that AI safety encompasses a broader set of present and future risks than existential scenarios alone [33,34,79]. 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 [46,51].

4.7. Governance, Oversight, and Risk Reduction Strategies

AGI and ASI risks may require technical safeguards, institutional accountability, legal oversight, international coordination, and public participation; the appropriate combination depends on capability and deployment context [46,80]. No single regulatory model or technical method can address the full range of uncertainty associated with advanced AI development. Effective governance may 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 [45]. These approaches seek to align AI behavior with human values, legal standards, and social expectations before highly autonomous systems become widely deployed [47,51].

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 identify harmful behavior, discriminatory outputs, security weaknesses, and unintended consequences before AI systems become deeply embedded in critical infrastructures. Recent international assessments emphasize that no single evaluation or alignment technique currently provides a complete basis for controlling increasingly capable general-purpose systems, which supports the use of multiple technical and institutional layers [46].

International governance frameworks are also necessary because advanced AI risks and economic effects extend beyond national borders. UNESCO and NIST emphasize transparency, accountability, fairness, human-rights protection, and risk management in AI governance [47,51]. More recent multilateral initiatives, including the United Nations Independent International Scientific Panel on Artificial Intelligence, reflect growing recognition that AI capabilities, risks, and opportunities require shared scientific assessment and international dialogue [56]. Reference [1] have proposed treaty-based approaches for highly advanced AI systems or autonomous weapons, although competitive incentives, verification problems, and different national interests may make a single universal treaty difficult to implement.

The literature suggests that alternative governance pathways may function as complementary rather than mutually exclusive approaches. A UN-linked coordinating institution or treaty-based mechanism may be appropriate for a limited set of transboundary, high-consequence risks, while interoperable national and regional regulation, networks of AI safety institutes, international technical standards, independent auditing and certification, and sector-specific arrangements may address other risks more effectively. A layered architecture also reduces dependence on the political feasibility of a single global institution. International coordination can also impose costs when harmonization constrains national sovereignty, reduces regulatory experimentation, or limits adaptation to different cultural, economic, and institutional contexts; this supports common minimum protections with room for context-sensitive implementation [46].

Governance capacity is uneven across countries and regions. Infrastructure, reliable electricity and connectivity, education and technical skills, access to computational resources, local data, administrative capacity, and regulatory enforcement affect the ability of countries to benefit from and govern advanced AI. The World Development Report 2026 emphasizes the importance of infrastructure, skills, institutions, local adaptation, and

reduced technological dependency for developing economies [54]. Recent peer-reviewed research also shows that practitioners in the Global North and Global South can differ in how they understand AI harm, power, and appropriate governance, supporting more context-sensitive international arrangements [55]. These differences support governance arrangements that combine common minimum protections with capacity building and sufficient flexibility for different institutional settings.

Transparency and explainability provide additional safeguards against abuse and discrimination. Existing governance frameworks emphasize auditability and interpretability for advanced AI systems used in healthcare, finance, law enforcement, education, employment, and public governance [47,51]. Human oversight is similarly emphasized in high-risk settings because fully autonomous decisions may weaken accountability and democratic control [51]. 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 [50]. 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 [41,44]. 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. Concerns about concentrated technological and political power have led to calls for broader participation in decisions concerning the trajectory of AGI and ASI development. Responsible-governance frameworks therefore emphasize participation from ethicists, educators, policymakers, civil society organizations, affected communities, workers, and the public [51]. 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.

5. Contributions and Implications

The following discussion distinguishes findings drawn from the synthesized literature, theoretical implications of those findings, and normative policy recommendations that extend beyond what the reviewed evidence alone can establish.

5.1. Synthesized Findings and Contribution

This review makes three related contributions. First, it integrates economic, ethical, and governance studies that are often considered separately and shows that their conclusions are interdependent. Second, it identifies assumptions that explain divergent AGI and ASI scenarios, including the degree of human-labor substitutability, the speed of capability diffusion, concentration of AI capital and computational resources, system autonomy, and institutional capacity. Third, it connects these assumptions to distributional outcomes across

workers, firms, countries, and governance systems. The analysis therefore suggests that technological capability alone cannot determine whether advanced AI produces broad prosperity or concentrated economic and social risk. The integrative framework in Figure 1 makes this contribution explicit by linking capability progression to economic and ethical mechanisms and by identifying institutional conditions that moderate the resulting outcomes.

5.2. Theoretical and Policy Implications

The synthesis yields an important policy implication. Governance focused only on current harms may be insufficient if AI systems acquire substantially greater autonomy and strategic capability, while governance focused only on hypothetical catastrophic scenarios may neglect existing problems involving inequality, discrimination, surveillance, and institutional concentration. A layered approach combining technical safety and evaluation, meaningful human oversight in high-risk domains, workforce and social-protection policies, competition and distributional policy, and international coordination is more consistent with the range of evidence reviewed.

The implications differ across AGI and ASI scenarios. AGI could affect healthcare, legal services, education, finance, manufacturing, public administration, and scientific research by supporting or automating diagnosis, document synthesis, legal analysis, financial modeling, research assistance, strategic planning, and complex information retrieval. These systems may improve efficiency and access to expertise while also weakening employment stability and professional development pathways. Their effects are also likely to vary across countries and regions. Workers and professionals in the Global South may experience the consequences of AGI adoption without comparable influence over its development, governance, or economic distribution, particularly where advanced AI infrastructure, data access, computational resources, and regulatory capacity remain concentrated in wealthier countries and powerful corporations [54,55,81]. Early preparation through education, professional adaptation, and international coordination may therefore reduce adjustment pressures while preserving room for national and regional variation [46,60,82].

ASI presents a more profound governance challenge because systems exceeding human cognitive capabilities may become difficult to oversee, predict, or correct after deployment. In high-consequence settings, this possibility shifts attention toward anticipatory governance, technical safety, and institutional preparedness before systems potentially exceed meaningful human control [46,83]. ASI also raises concerns about human agency if consequential social decisions become shaped by systems operating beyond direct democratic oversight or ordinary human comprehension [84]. At the same time, more optimistic scenarios associate highly capable systems with accelerated advances in medicine, climate solutions, poverty reduction, and food security. These competing possibilities make the design, training, deployment, and governance of advanced AI systems consequential before greater autonomy is reached.

Current governance frameworks remain incomplete for addressing the evolving risks associated with advanced AI. Recent international initiatives, including the United Nations Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on AI Governance, reflect growing recognition that some AI risks are transboundary and cannot be addressed effectively through national policy alone [56]. A stronger UN-linked coordinating institution or treaty-based mechanism represents one possible pathway for a limited set of high-consequence risks. Alternative and complementary pathways include interoperable national and regional regulation, networks of AI safety institutes, international technical standards, independent auditing and certification, and sector-specific agreements. These approaches are not mutually exclusive. The evidence reviewed in this study therefore

supports stronger international coordination while suggesting that effective governance may require a layered architecture rather than a single universal institutional model.

Economic policy may also need to address the possibility that AGI and ASI differ from earlier technological revolutions in the breadth of cognitive and professional functions that could become automatable. If new tasks, complementary roles, and transition mechanisms do not emerge at sufficient scale, some workers may lose labor income or bargaining power even as productivity rises. The reviewed literature therefore supports treating labor displacement, wealth concentration, value alignment, accountability, and human agency as interconnected policy challenges. The future benefits of AGI and ASI will depend not only on technical progress but also on whether governance systems promote public interest, human rights, inclusive participation, and broad access to benefits rather than primarily reflecting narrow institutional, commercial, or geopolitical interests [80,82].

6. Conclusions

This structured narrative review shows that the economic, ethical, and governance implications of AGI and ASI are closely interconnected. The literature does not support a single deterministic trajectory. Potential outcomes depend on the extent to which advanced AI substitutes for or complements human labor, the ownership and distribution of AI-related capital and computational resources, the autonomy and strategic capability of AI systems, and the capacity of institutions to adapt. Productivity and innovation gains may coexist with labor displacement, inequality, concentration of technological power, and new forms of ethical and governance risk. The central challenge is therefore not only whether increasingly capable AI can generate economic and scientific benefits, but also how those benefits and risks are distributed and whether human institutions can retain meaningful oversight, accountability, and agency.

This review also has limitations. It uses a structured narrative design rather than a systematic review or meta-analysis and combines peer-reviewed research, scholarly books, working papers, institutional assessments, and selected industry forecasts with different evidentiary standards. No formal weighting scheme or exhaustive risk-of-bias assessment was applied. Evidence concerning AGI and especially ASI remains partly theoretical and scenario-based because these systems have not been empirically observed as established technological categories, and the rapid pace of AI development may change both capability forecasts and policy assessments. Future research should therefore test the mechanisms identified in this review as new empirical evidence becomes available, compare AGI-related economic and governance effects across institutional and national contexts, and examine how alternative ownership, regulatory, safety, and social-protection arrangements influence the distribution of benefits and risks. The scenarios synthesized here should be interpreted as conditional pathways rather than deterministic predictions.

Author Contributions: Conceptualization, M.O. and Z.O.; methodology, M.O., Z.O., B.D.L. and N.P.M.; investigation (literature review), M.O. and Z.O.; writing—original draft preparation, M.O., Z.O., B.D.L. and N.P.M.; writing—review and editing, M.O., Z.O., B.D.L. and N.P.M.; supervision, M.O., B.D.L. and N.P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bostrom, N. *Superintelligence: Paths, Dangers, Strategies*; Oxford University Press: Oxford, UK, 2014.
2. Yenduri, G.; Murugan, R.; Maddikunta, P.K.R.; Bhattacharya, S.; Sudheer, D.; Savarala, B.B. Artificial general intelligence: Advancements, challenges, and future directions in AGI research. *IEEE Access* **2025**, *13*, 134325–134356. [CrossRef]
3. Kim, H.; Yi, X.; Yao, J.; Lian, J.; Huang, M.; Duan, S.; Hu, W.; Zhang, K.; Xie, X. The road to artificial superintelligence: A comprehensive survey of superalignment. *arXiv* **2024**, arXiv:2412.16468. [CrossRef]
4. Mundlamuri, R.; Gunnam, G.R.; Mysari, N.K.; Pujuri, J. The evolution of AI: From classical machine learning to modern large language models. *IEEE Access* **2025**, *13*, 178302–178341. [CrossRef]
5. Naveed, H.; Khan, A.U.; Qiu, S.; Saqib, M.; Anwar, S.; Usman, M.; Akhtar, N.; Barnes, N.; Mian, A. A comprehensive overview of large language models. *ACM Trans. Intell. Syst. Technol.* **2025**, *16*, 1–72. [CrossRef]
6. Dou, F.; Ye, J.; Yuan, G.; Lu, Q.; Niu, W.; Sun, H.; Guan, H.; Song, W. Towards artificial general intelligence in the Internet of Things: Opportunities and challenges. *arXiv* **2023**, arXiv:2309.07438. [CrossRef]
7. Latif, E.; Mai, G.; Nyaaba, M.; Wu, X.; Liu, N.; Lu, G.; Li, S.; Liu, T.; Zhai, X. Artificial general intelligence for education. *arXiv* **2023**, arXiv:2304.12479. [CrossRef]
8. Masters, K.; Herrmann-Werner, A.; Festl-Wietek, T.; Taylor, D. Preparing for artificial general intelligence in health professions education: AMEE guide no. 172. *Med. Teach.* **2024**, *46*, 1258–1271. [CrossRef] [PubMed]
9. Everitt, T.; Lea, G.; Hutter, M. AGI safety literature review. *arXiv* **2018**, arXiv:1805.01109. [CrossRef]
10. McLean, S.; Read, G.J.; Thompson, J.; Baber, C.; Stanton, N.A.; Salmon, P.M. The risks associated with artificial general intelligence: A systematic review. *J. Exp. Theor. Artif. Intell.* **2023**, *35*, 649–663. [CrossRef]
11. Salmon, P.M.; Carden, T.; Hancock, P.A. Putting the humanity into inhuman systems: How human factors and ergonomics can be used to manage the risks associated with artificial general intelligence. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2021**, *31*, 223–236. [CrossRef]
12. Salmon, P.M.; King, B.J.; McLean, S.; Read, G.J.; Shanahan, C.; Devitt, K. A framework of human factors methods for safe, ethical, and usable artificial intelligence in defense. In *Putting AI in the Critical Loop*; Academic Press: Cambridge, MA, USA, 2024; pp. 245–263.
13. Chalmers, D.J. The singularity: A philosophical analysis. In *Science Fiction and Philosophy: From Time Travel to Superintelligence*, 2nd ed.; Schneider, S., Ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2016; pp. 171–224. [CrossRef]
14. Muehlhauser, L.; Salamon, A. Intelligence explosion: Evidence and import. In *Singularity Hypotheses: A Scientific and Philosophical Assessment*; Eden, A.H., Moor, J.H., Søraker, J.H., Steinhart, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 15–42.
15. Naudé, W.; Dimitri, N. The race for an artificial general intelligence: Implications for public policy. *AI Soc.* **2020**, *35*, 367–379. [CrossRef]
16. Radanliev, P.; De Roure, D.; Maple, C.; Ani, U. Super-forecasting the technological singularity risks from artificial intelligence. *Evol. Syst.* **2022**, *13*, 747–757. [CrossRef] [PubMed]
17. Brynjolfsson, E.; McAfee, A. The business of artificial intelligence. *Harv. Bus. Rev.* **2017**, *7*, 1–2.
18. Furman, J.; Seamans, R. AI and the economy. *Innov. Policy Econ.* **2019**, *19*, 161–191. [CrossRef]
19. Tyson, L.D.; Zysman, J. Automation, AI, and work. *Daedalus* **2022**, *151*, 256–271. [CrossRef]
20. Acemoglu, D.; Restrepo, P. Artificial Intelligence, Automation, and Work. In *The Economics of Artificial Intelligence: An Agenda*; Agrawal, A., Gans, J., Goldfarb, A., Eds.; University of Chicago Press: Chicago, IL, USA, 2019; pp. 197–236.
21. Faishal, M.; Mathew, S.; Neikha, K.; Pusa, K.; Zhimomi, T. The future of work: AI, automation, and the changing dynamics of developed economies. *World J. Adv. Res. Rev.* **2023**, *18*, 620–629. [CrossRef]
22. Joshi, K. Artificial general intelligence: A comprehensive review. *J. Epidemiol. Found. India* **2024**, *2*, 93–96. [CrossRef]
23. Trammell, P.; Korinek, A. *Economic Growth Under Transformative AI*; NBER Working Paper No. 31815; National Bureau of Economic Research: Cambridge, MA, USA, 2023.
24. Gabriel, I. Artificial intelligence, values, and alignment. *Minds Mach.* **2020**, *30*, 411–437. [CrossRef]
25. Lechterman, T.M. The concept of accountability in AI ethics and governance. In *The Oxford Handbook of AI Governance*; Bullock, J.B., Chen, Y.-C., Himmelreich, J., Hudson, V.M., Korinek, A., Young, M.M., Zhang, B., Eds.; Oxford University Press: Oxford, UK, 2022; pp. 164–182. [CrossRef]
26. Prunkl, C. Human autonomy at risk? An analysis of the challenges from AI. *Minds Mach.* **2024**, *34*, 26. [CrossRef]
27. Stahl, B.C. Ethical issues of AI. In *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 35–53.
28. Batool, A.; Zowghi, D.; Bano, M. Responsible AI governance: A systematic literature review. *arXiv* **2023**, arXiv:2401.10896. [CrossRef]
29. Birkstedt, T.; Minkkinen, M.; Tandon, A.; Mäntymäki, M. AI governance: Themes, knowledge gaps and future agendas. *Internet Res.* **2023**, *33*, 133–167. [CrossRef]
30. Taeihagh, A. Governance of artificial intelligence. *Policy Soc.* **2021**, *40*, 137–157. [CrossRef]

31. Lund, B.; Orhan, Z.; Mannuru, N.R.; Bevara, R.V.K.; Porter, B.; Vinaih, M.K.; Bhaskara, P. Standards, frameworks, and legislation for artificial intelligence transparency. *AI Ethics* **2025**, *5*, 3639–3655. [CrossRef]
32. Orhan, Z.; Orhan, M.; Lund, B.D.; Mannuru, N.R.; Bevara, R.V.K.; Porter, B. Artificial intelligence standards in conflict: Local challenges and global ambitions. *Standards* **2025**, *5*, 27. [CrossRef]
33. Bareis, J.; Ackerl, C.; Heil, R. AI going rogue? An integrative narrative review of the tacit assumptions underlying existential AI-risks. *AI Ethics* **2026**, *6*, 152. [CrossRef]
34. Fröberg, L. AI risk: Structuring the debate. *AI Ethics* **2026**, *6*, 420. [CrossRef]
35. Baethge, C.; Goldbeck-Wood, S.; Mertens, S. SANRA—A Scale for the Quality Assessment of Narrative Review Articles. *Res. Integr. Peer Rev.* **2019**, *4*, 5. [CrossRef] [PubMed]
36. Snyder, H. Literature review as a research methodology: An overview and guidelines. *J. Bus. Res.* **2019**, *104*, 333–339. [CrossRef]
37. Noy, S.; Zhang, W. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* **2023**, *381*, 187–192. [CrossRef] [PubMed]
38. Baslandze, S.; Edwards, Z.; Graham, J.R.; McClure, T.; Sparks, M.; Meyer, B.H.; Waddell, S.R.; Weitz, D. *Artificial Intelligence, Productivity, and the Workforce: Evidence from Corporate Executives*; NBER Working Paper No. 34984; National Bureau of Economic Research: Cambridge, MA, USA, 2026. Available online: https://www.nber.org/papers/w34984 (accessed on 14 October 2026).
39. Acemoglu, D. The simple macroeconomics of AI. *Econ. Policy* **2025**, *40*, 13–58. [CrossRef]
40. International Monetary Fund. Scenario Planning for an AGI Future. Finance & Development Magazine. 2023. Available online: https://www.imf.org/en/publications/fandd/issues/2023/12/scenario-planning-for-an-agi-future-anton-korinek (accessed on 14 October 2026).
41. Cazzaniga, M.; Jaumotte, F.; Li, L.; Melina, G.; Panton, A.J.; Pizzinelli, C.; Rockall, E.J.; Mendes Tavares, M. *Gen-AI: Artificial Intelligence and the Future of Work*; IMF Staff Discussion Note No. 2024/001; International Monetary Fund: Washington, DC, USA, 2024. [CrossRef]
42. Catalini, C.; Hui, X.; Wu, J. Some simple economics of AGI. *arXiv* **2026**, arXiv:2602.20946. [CrossRef]
43. Korinek, A.; Suh, D. *Scenarios for the Transition to AGI*; NBER Working Paper No. 32255; National Bureau of Economic Research: Cambridge, MA, USA, 2024.
44. Stiefenhofer, P. Artificial general intelligence and the end of human employment: The need to renegotiate the social contract. *arXiv* **2025**, arXiv:2502.07050. [CrossRef]
45. Russell, S. *Human Compatible: Artificial Intelligence and the Problem of Control*; Viking: New York, NY, USA, 2019.
46. International AI Safety Report. International AI Safety Report 2026. 2026. Available online: https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026 (accessed on 14 October 2026).
47. National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*; U.S. Department of Commerce: Washington, DC, USA, 2023. Available online: https://www.nist.gov/itl/ai-risk-management-framework (accessed on 14 October 2026).
48. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proc. Mach. Learn. Res.* **2018**, *81*, 77–91.
49. Noble, S.U. *Algorithms of Oppression: How Search Engines Reinforce Racism*; New York University Press: New York, NY, USA, 2018.
50. Crawford, K. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*; Yale University Press: New Haven, CT, USA, 2021.
51. UNESCO. *Recommendation on the Ethics of Artificial Intelligence*; UNESCO: Paris, France, 2021.
52. Bengio, Y.; Hinton, G.; Yao, A.; Song, D.; Abbeel, P.; Darrell, T.; Harari, Y.N.; Zhang, Y.-Q.; Xue, L.; Shalev-Shwartz, S.; et al. Managing extreme AI risks amid rapid progress. *Science* **2024**, *384*, 842–845. [CrossRef] [PubMed]
53. Marcus, G. Deep Learning is Hitting a Wall. Nautilus. 10 March 2022. Available online: https://nautil.us/deep-learning-is-hitting-a-wall-238440/ (accessed on 14 October 2026).
54. World Bank. *World Development Report 2026: The Promise of Artificial Intelligence*; World Bank: Washington, DC, USA, 2026. Available online: https://www.worldbank.org/en/publication/wdr2026 (accessed on 14 October 2026).
55. Brown, V.; Larasati, R.; Kwarteng, J.; Farrell, T. Understanding AI and power: Situated perspectives from Global North and South practitioners. *AI Soc.* **2026**, *41*, 3981–3996. [CrossRef] [PubMed]
56. Independent International Scientific Panel on Artificial Intelligence. *Preliminary Report*; United Nations: New York, NY, USA, 2026. Available online: https://www.un.org/independent-international-scientific-panel-ai/en/preliminary-report (accessed on 14 October 2026).
57. Jones, C.I. *A.I. and Our Economic Future*; NBER Working Paper No. 34779; National Bureau of Economic Research: Cambridge, MA, USA, 2026. [CrossRef]
58. Raman, R.; Kowalski, R.; Achuthan, K.; Iyer, A.; Nedungadi, P. Navigating artificial general intelligence development: Societal, technological, ethical, and brain-inspired pathways. *Sci. Rep.* **2025**, *15*, 92190. [CrossRef] [PubMed]

59. Huang, X. *Labor Market Exposure to AI: From GenAI to Future AGI*; ASEAN+3 Macroeconomic Research Office: Singapore, 2025. Available online: https://amro-asia.org/wp-content/uploads/2025/11/GenAI_Labour_Huang2025_20251107.pdf (accessed on 14 October 2026).
60. Joshi, S. Review of artificial general intelligence: Implications for the U.S. workforce and economic stability. *Management* **2025**, *4*, 336–350. [CrossRef]
61. Aghion, P.; Jones, B.F.; Jones, C.I. Artificial Intelligence and Economic Growth. In *The Economics of Artificial Intelligence: An Agenda*; Agrawal, A., Gans, J., Goldfarb, A., Eds.; University of Chicago Press: Chicago, IL, USA, 2019; pp. 237–282. [CrossRef]
62. Goldman Sachs. Generative AI Could Raise Global GDP by 7%. 2023. Available online: https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent (accessed on 14 October 2026).
63. McKinsey Global Institute. *The Economic Potential of Generative AI: The Next Productivity Frontier*; McKinsey & Company: New York, NY, USA, 2023.
64. Goldman Sachs Global Investment Research. The Potentially Large Effects of Artificial Intelligence on Economic Growth. 2023. Available online: https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html (accessed on 14 October 2026).
65. Kpmg, L.L.P. Generative AI and Economic Growth: A New Approach to Measuring Its Potential Economic Impact. 2025. Available online: https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2025/gen-ai-economic-growth.pdf (accessed on 14 October 2026).
66. Agrawal, A.; Gans, J.; Goldfarb, A. Genius on Demand: The Value of Transformative Artificial Intelligence. In *The Economics of Transformative AI*; University of Chicago Press: Chicago, IL, USA, 2026; pp. 15–32. Available online: https://www.nber.org/books-and-chapters/economics-transformative-ai/genius-demand-value-transformative-artificial-intelligence (accessed on 14 October 2026).
67. Acemoglu, D.; Autor, D.; Johnson, S. *Building Pro-Worker Artificial Intelligence*; NBER Working Paper No. 34854; National Bureau of Economic Research: Cambridge, MA, USA, 2026. [CrossRef]
68. Agrawal, A.K.; McHale, J.; Oettl, A. *Enhancing Worker Productivity Without Automating Tasks: A Different Approach to AI and the Task-Based Model*; NBER Working Paper No. 34781; National Bureau of Economic Research: Cambridge, MA, USA, 2026. [CrossRef]
69. Goldman Sachs. AI May Start to Boost US GDP in 2027. 2023. Available online: https://www.goldmansachs.com/insights/articles/ai-may-start-to-boost-us-gdp-in-2027 (accessed on 14 October 2026).
70. Brollo, F.; Dabla-Norris, E.; de Mooij, R.; Garcia-Macia, D.; Hanappi, T.; Liu, L.; Nguyen, A.D.M. *Broadening the Gains from Generative AI: The Role of Fiscal Policies*; IMF Staff Discussion Note No. 2024/002; International Monetary Fund: Washington, DC, USA, 2024. [CrossRef]
71. Rockall, E.J.; Tavares, M.M.; Pizzinelli, C. *AI Adoption and Inequality*; IMF Working Paper No. 2025/068; International Monetary Fund: Washington, DC, USA, 2025. [CrossRef]
72. Korinek, A.; Stiglitz, J.E. *Steering Technological Progress*; NBER Working Paper No. 34994; National Bureau of Economic Research: Cambridge, MA, USA, 2026. [CrossRef]
73. Cirera, X.; Comin, D.; Cruz, M. *Bridging the Technological Divide: Technology Adoption by Firms in Developing Countries*; World Bank Group: Washington, DC, USA, 2022.
74. Cerutti, E.M.; Garcia Pascual, A.I.; Kido, Y.; Li, L.; Melina, G.; Mendes Tavares, M.; Wingender, P. *The Global Impact of AI: Mind the Gap*; IMF Working Paper No. 2025/076; International Monetary Fund: Washington, DC, USA, 2025. [CrossRef]
75. Steinmetz, A. A progress report on the economics of artificial intelligence: Impacts, challenges, and future directions. *J. Econ. Anal.* **2025**, *4*, 119–144. [CrossRef]
76. Caja Moya, C.; Quiroga Rodríguez, E. Deconstructing gender bias in AGI: Mitigating discriminatory architectures in general intelligence. *AI Ethics* **2025**, *5*, 5857–5865. [CrossRef]
77. O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*; Penguin Books: London, UK, 2017.
78. Eubanks, V. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*; St. Martin's Press: New York, NY, USA, 2018.
79. Gyevnár, B.; Kasirzadeh, A. AI safety for everyone. *Nat. Mach. Intell.* **2025**, *7*, 531–542. [CrossRef]
80. United Nations Advisory Body on Artificial Intelligence. *Governing AI for Humanity: Final Report*; United Nations: New York, NY, USA, 2024. Available online: https://www.un.org/en/ai-advisory-body (accessed on 14 October 2026).
81. Mannuru, N.R.; Shahriar, S.; A Teel, Z.; Wang, T.; Lund, B.D.; Tijani, S.; Pohboon, C.O.; Agbaji, D.; Alhassan, J.; Galley, J.; et al. Artificial intelligence in developing countries: The impact of generative artificial intelligence technologies for development. *Inf. Dev.* **2025**, *41*, 1036–1054. [CrossRef]
82. United Nations. Global Dialogue on AI Governance. 2026. Available online: https://www.un.org/global-dialogue-ai-governance/en (accessed on 14 October 2026).

83. Gill, K.S. Artificial super intelligence: Beyond rhetoric. *AI Soc.* **2016**, *31*, 137–143. [CrossRef]
84. Dessureault, J.S.; Lamontagne, R.; Parisé, P.O. The ethics of creating artificial superintelligence: A global risk perspective. *AI Ethics* **2025**, *5*, 6241–6263. [CrossRef]

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.