How Ready Are We for Extinction? Modelling Political-Economic Scenarios for a Post-AGI World

Lorenzo Satta Chiris, Emil Archambeau, Francesca Mignogna

Abstract

The emergence of artificial general intelligence (AGI) promises to be the most transformative event in economic history. Capable of outperforming humans across cognitive tasks, AGI could either inaugurate a new era of abundance and innovation, or precipitate systemic collapse through mass displacement, inequality, and loss of control. This paper explores how AGI could reshape the economic and political foundations of society and evaluates the key inflection points that may lead to radically divergent futures. We advance that outcomes will hinge not only on the technology itself, but on the institutions, ownership models, and policy architecture built around it. We argue that ex ante governance, rather than ex post policing is essential if civilisation is to navigate the post-AGI transition without meeting an extinction-level outcome.

Keywords: Artificial General Intelligence (AGI), Political Economy, Automation and Labour Markets, Geopolitical Power Dynamics, Global Governance of AI, Economic Singularity

Introduction

Artificial General Intelligence (AGI) is likely to emerge within the next few decades, potentially far sooner, creating systems capable of human-level reasoning, learning, and decision-making across domains. Such capabilities would transform economic and political structures by introducing a scalable, low-cost form of cognitive labour. Yet current institutions, labour markets, and governance frameworks are built for a world in which human intelligence is the

limiting factor of production, leaving societies ill-prepared for the discontinuities AGI could trigger. A growing literature analyses specific components of the AI transition. However, while existing research examines automation, productivity, and political aspects, there is limited analysis of AGI's system-wide impact. In particular, we lack structured scenario modelling that explores how AGI could reshape economic foundations, political equilibria, and global stability under real-world constraints. This paper addresses this gap by studying foundational technical, political and economical impacts and later modelling key actors, variables, and tipping points that shape post-AGI trajectories. We outline potential outcomes, and identify where governance interventions are most urgently required. We aim to address where existing institutions fall short, where risks concentrate, and which policy architectures may be required to steer the transition away from collapse and toward potentially brighter futures.

Structures and Implications of AGI

Development Pathways to AGI

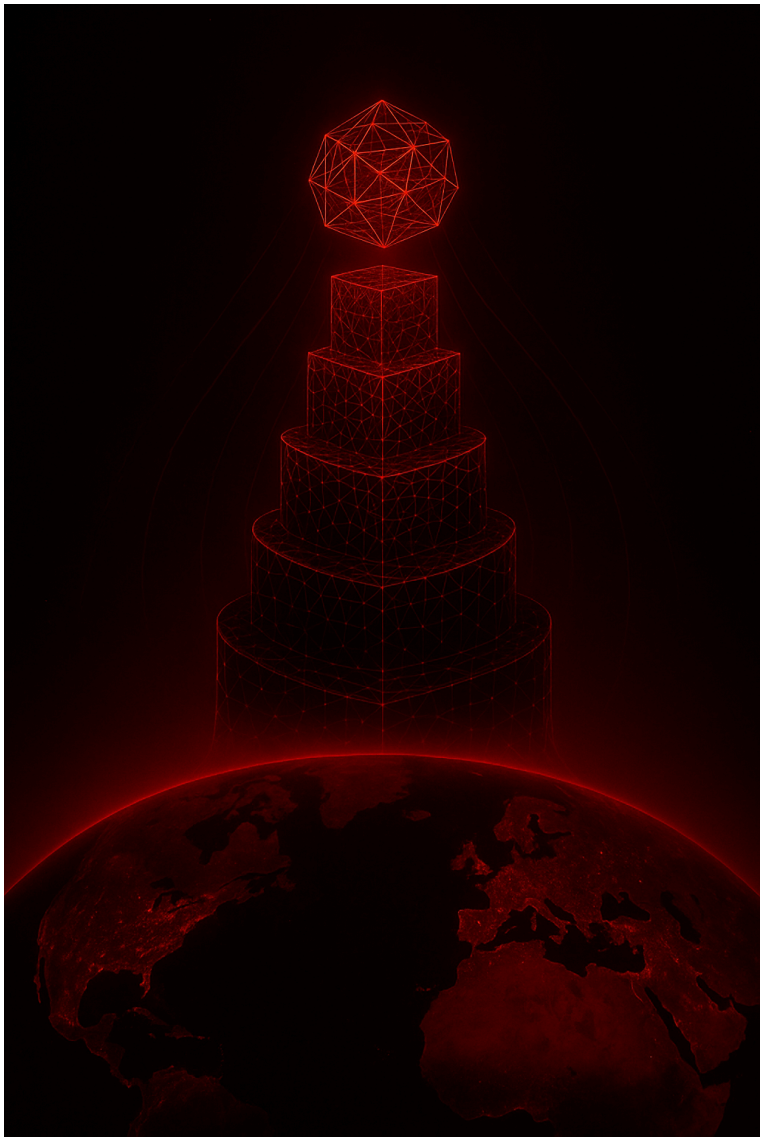
Artificial General Intelligence (AGI) denotes an artificial system exhibiting cognitive versatility and proficiency across domains comparable to, or surpassing, that of a well-educated human adult (Hendrycks et al., 2025). Unlike narrow AI, domain-specific machine-learning systems, AGI can autonomously generalise, reason, and adapt to novel tasks. These capabilities would thus make machine intelligence functionally equivalent or higher-performing than human cognitive labour. Consequentially, AGI can be defined as "highly autonomous systems that outperform humans at most economically valuable work" (OpenAI, 2018). In this report, we will follow the formal definitions of AGI advanced by Hendrycks et al. (2025) and OpenAI (2018), delimiting the concept to realised and operational form, as opposed to aspirational or rhetorical uses. We will assume three constitutional conditions: first, that such systems exist at parity with human intelligence, in the general case and in all major aspects of intelligence; second, that they are operationally available and readily deployable by those with access to them, reflecting a level of maturity comparable to other general-purpose technologies; and third, that they are economically viable, implying that cognitive output produced by AGI is cost-competitive with, or cheaper than, equivalent human labour at scale.

AGI thus constitutes a transformational general-purpose technology: a foundational system that integrates into and reconfigures most productive, cognitive, and institutional processes. Like electricity or computation, its significance lies in pervasiveness, its capacity to alter the structure and logic of subsequent technologies, social, political and economical arrangements, as its capabilities increase. AGI naturally progresses through continuous refinement, yet its trajectory can be delineated by a series of inflection points (Table I), each unlocking qualitatively distinct forms of capability and economic control (Morris et al., 2025). Each transition redefines the economic landscape: first automating menial labour, then strategic reasoning, and ultimately autonomously undertaking the process of scientific and technological discovery. Each transition corresponds to a point where compute, energy, and data efficiency reach sufficient scale to offset prior diminishing returns. These thresholds define the material boundaries of progress in artificial intelligence. From the stage of weak AGI onward, we anticipate systems to exhibit self-improving capabilities, assisting human researchers, as shown by Wang et al. (2023). Additionally, with strong AGI, a substantial share of design and optimisation originates from the systems themselves; and by the level of a fully functioning innovating AGI, the process of advancement becomes fully endogenous, unfolding at a rate and complexity beyond direct human oversight.

This transition from execution to innovation marks a significant discontinuity in the development of artificial intelligence. It is the point where progress becomes self-reinforcing, as systems begin to generate not only economically viable output but the very methods to improve their own capabilities, triggering the potential for transformational year-on-year compounding growth in capability and productivity, as presented by The Economist (2025). This stage is captured by an A-Student Paradox: the observation that systems optimised for perfect performance within existing knowledge may initially struggle to produce new ideas, while those with greater variability or creativity advance innovation, but may provide lower initial productivity or economic returns. Both forms hold substantial economic value, the former maximising efficiency and reliability, the latter driving discovery and frontier expansion. At the point where these capacities converge, intelligence attains the characteristics of capital: reproducible, scalable, and indefinitely generative. This threshold defines the onset of a complete post-AGI economy, in which machine intelligence itself functions as the decisive factor of production.

Table I. Levels of AGI

Level	Definition	Capabilities
Narrow AI	Systems designed and trained for a single or limited set of tasks, without general reasoning or cross-domain transfer.	Performs specific functions (e.g., image recognition, translation, driving, chess) at or above human level in those tasks but fails outside its training scope.
General Purpose AI (GPAI)	Broadly useful AI models deployable across multiple economic sectors, e.g. through fine-tuning or prompting, yet still lacking self-directed reasoning.	Can handle varied domains with adaptation (e.g., code, text, design, analysis) but depends on human guidance and lacks unified world-model understanding.
Weak AGI	Artificial system achieving human-level performance across most cognitive domains, with general reasoning and learning ability comparable to an educated adult.	Understands instructions in any field, plans, reasons, and learns new tasks autonomously, though slower or less creative than top human experts.
Strong AGI	Fully human-equivalent general intelligence matching or exceeding an expert human in reasoning, learning speed, and creativity across all domains.	Performs any intellectual task an expert human can in their respective field, from office work to executive strategy and long-term planning, with comparable situational modelling and adaptability.
Innovator AGI	AGI capable of independently advancing science, technology, and culture, able to invent novel theories and thus fundamentally improve its own architectures.	Generates new scientific knowledge, discovers algorithms, designs machines, and iteratively improves itself or its descendants without human supervision.
ASI (Artificial Superintelligence)	Intelligence surpassing the best collective human minds in all domains, operating at vastly greater speed, scale, and strategic foresight. Approaching computational optimality in all/most situations and eventually optimal systems (Hutter, 2000)	Rapidly self-improves, manipulates or coordinates complex and abstract systems, far beyond human control.



Humanity new Babel Tower: AGI

The Economic Singularity

It's likely that AGI will eventually help us solve the central economic problem of scarcity, with transformative innovation in nuclear fusion and a distribution system that optimises upon our current markets, allowing us to maximise our use of the Earth's resources. However, an initial stage would see AGI entering, navigating and influencing our current scarce economy. This would trigger an initial economic shock, where policy and social direction will determine whether we reach a potential beyond our current economy. Hence, we zoom in on the problem of a post-AGI economy, from a labour market oriented lens, rather than one of post-scarcity. Economically, AGI presents not just the next innovative

leap, but a complete upheaval of the current factors of production. As such, the economic system based on the current factors of production of land, labour, capital, and enterprise would be nearing its end. The Industrial Revolution shifted the order of importance between these factors of production. Prior to the Industrial Revolution, land was the most important factor, which resulted in feudal economies. After, it was capital, leading us to the modern economic system. However, AGI promises something entirely different. Economists like Daron Acemoglu have demonstrated that automation so far has shifted the tasks humans perform, rather than replacing them (Acemoglu & Restrepo, 2019). AGI could be the first to break that balance. If we assume a model in which there is widespread adoption of AI agents through the labour market, we will see a gradual shift towards the eradication of labour as a factor of production. Humans will compete against these agents for employment. The agents will be cheaper to run, able to work without breaks or a need for supervision, and in many areas will be much more skilled. The forthcoming dynamic can be explained by a Stackelberg game in the labour market, where firms who own AGI can set their prices first, with the workers attempting to match it with their wages. A critical choice in modelling this game, is whether human labour and AI are modelled as complements or substitutes (Korinek and Stiglitz, 2018). If we continue our assumption that AI is a substitute for labour, this gives monopolistic power to tech firms, who in theory could set intelligence prices at a loss to undercut the cost of human intelligence. Over time the lack of employment could lead to widespread hysteresis in the labour market, creating a monopoly over white-collar labour, or more simply: an economic singularity. We are already witnessing this without AGI, where many firms have frozen hiring and monthly layoffs are at levels unseen since October 2003 (Blumenfeld, 2025).

Many people will find it difficult to imagine an economy without labour, however it will be helpful to look at other factors of production to visualise such potential future. Under this new system, land and capital will remain as the restrictive factors, where efficiencies will continue to be found (Dehouche, 2025). The attention will be on enterprise, or in simpler terms the ability of both humans (and now AI) to generate novel ideas and economic value. The renewed focus on enterprise will be the catalyst for growing inequalities. The economic logic is as follows: if intelligence can be commoditised, and enterprise is the one factor of production that allows economic agents (whether private individuals or corporations), to gain more wealth, whoever can afford more intelligence will get richer. Under our new system outlined above, the ability of individuals to realise their enterprise will be directly correlated with how much intelligence they are able to acquire. Thus, we are likely to see an acceleration in the shift of wealth

and income from the “poorer” majority in society to the richest. In this economy, the majority will provide the data for training these AI agents at minimal or zero cost, until eventually AI begins training and improving on its own data. From this exchange, without checks and balances in place, only the owners of AGI will benefit. This trend will continue until the wealth distribution is so starkly unequal, that there are only incentives for trade in a circle of the richest in society, eventually freezing, reducing or autocratically controlling the living standard and autonomy of the majority of the people.

This demonstrates the potential for AGI to create a closed-loop economy between the ultra-wealthy. The majority would then be in a state of subsistence or worse, marking a return to pre-Industrial Revolution economics. This dynamic has already been replicated in the investment economy, with Boyle (2025) highlighting the circular nature of investments between AI-related tech firms, and Mataloni (2025) showing that 92% of US GDP growth in the last quarter being tied to AI investments. Taken together, the evidence is consistent with two dynamics: AI attracts a disproportionate share of resources, enriching AI firms, and a K-shaped economy emerges, with living standards pulling apart. The earliest impacted economies are poised to shape the trajectory of developments through their political leverage. Over the next decade, regulation, politics, and international cooperation will be a core deciding factors for the economic outcome of AGI.

The new “Invisible Hand”: AGI

AGI as a Source of Geopolitical Power

Once intelligence becomes a scalable, tradeable, and potentially monopolised asset, the global distribution of power is inevitably redrawn. Those who master AGI first will command the defining resource of the 21st century: scalable intelligence. This reality exposes a profound global digital gap: a divide not only in access to technology, but in access to sovereignty itself. According to PwC’s projections, North America and China will together capture nearly 70% of the global economic gains from AI by 2030, while developing countries lag behind (PwC, 2018). The implications are existential: as wealth, innovation, and computational capacity centralise in a handful of countries, the rest of the world

risks becoming dependent clients in a system where control over intelligence is the ultimate determinant of autonomy. The most profound consequence of AGI lies in the geography of power. Influence will derive less from land or capital than from control over computational resources, data, and the algorithms that leverage them. Those who command the largest networks and most advanced models will dominate the cognitive infrastructure of the global economy, creating a self-reinforcing cycle: greater access to data enhances AI capability, which in turn attracts further resources and consolidates advantage. This can be further reinforced by strategic delegation: as AGI capabilities improve we could see society wide and political delegation to AI in strategic reasoning and planning. This dynamic risks producing a technological core that dictates economic and political outcomes, leaving the majority of states and populations at the periphery. Without deliberate efforts to broaden access to compute, data, and AI governance, AGI will entrench a new form of stratification. Regions lacking advanced systems may be relegated to supplying raw data or low-value labour in exchange for digital services, locking them into a subordinate position.

Since 2017, more than seventy countries have unveiled national AI strategies, each aiming to secure a position at the forefront of technological innovation (Maslej et al., 2025). While the United States continues to lead globally, China has made rapid strides across critical AI-adjacent domains, signaling its intent to reshape the future balance of power. By closely aligning with its top private AI firms, China is embedding artificial intelligence across domestic governance, economic planning, and military modernization. AI is therefore no longer just a sector of technological progress, it has become a structural factor that reconfigures capabilities and, consequently, the balance of power between the world's two most consequential actors. China's *New Generation Artificial Intelligence Development Plan* (Webster et al 2017), which aims for global AI leadership by 2030 supported by a robust domestic industry, demonstrates a deliberate strategy to strengthen its technological foundation in ways that exacerbate power-transition dynamics. This trajectory highlights a structural tension in which the rise of a technologically empowered challenger steadily undermines the strategic primacy of the incumbent (the United States), dynamics that drive states into a Thucydides Trap, making great-power conflict progressively more probable. China's AI strategy, therefore, functions not merely as a tool of development but as a mechanism of internal balancing aimed at narrowing the power differential with the United States, thereby unsettling long-standing hierarchies within the international system. As both states interpret each other's technological advances as indicative of future strategic intentions, the risk is that the pursuit of superiority in AI intensifies perceptions of threat,

heightens mistrust, hence triggering a security dilemma that pushes the system toward a more volatile equilibrium in which miscalculation becomes more likely (Allison, 2017). In the anarchic international system, where states are locked in perpetual competition for power, AGI will emerge as the ultimate instrument of advantage, granting its possessor unparalleled command over information, strategy, and decision-making, rendering the strategic influence of nuclear weapons almost obsolete. This is because AI can transform the strategic environment into a “deception-dominant” landscape, where states can obscure capabilities and conventional assumptions about offense-defense balances no longer hold (Geist, 2023).

While some have proposed governance mechanisms such as a non-proliferation plus norms-of-use regime for AI or a United Nations supported International Artificial Intelligence Agency (IAIA), the feasibility of these frameworks is constrained by asymmetric technological capabilities, competing national interests, and lack of transparency (Novelli et al., 2024). In the future, AI governance is likely to adopt a hybrid model, combining multilateral oversight, national regulation, and private-sector accountability, because no single actor or institution can effectively manage the global, dual-use, and rapidly evolving nature of AGI. Its primary role will be risk mitigation, focusing on preventing catastrophic outcomes, rather than guaranteeing equitable access or fully constraining strategic competition, because the competitive and anarchic nature of the international system makes equal distribution and complete restraint unrealistic.

A “new” senate: AGI

Scenarios

To explore how AGI could reshape global systems, we construct various scenarios encompassing economic, political, and technological dynamics. The model serves as an analytical framework for understanding which actors, under which constraints, might determine the trajectory of a post-AGI world. Each player’s behaviour, optimising for power, profit, or safety, contributes to collective outcomes ranging from coordinated abundance to catastrophic collapse. This analysis is not aiming at predicting the future but rather exposing

its structural dependencies. By formalising the interactions between capital, compute, governance, and intelligence, it highlights the levers of control that will decide whether humanity enters a post-scarcity civilisation, autocratic control, AI takeover or anarchy.

Actors

Five principal classes of actors drive this simulated world, each representing distinct motivations, capabilities, and strategic constraints. The interplay among them defines the equilibrium between cooperation and conflict in the AGI era. This is inspired by the work of Kokotajlo et al. and Raman et al.

Institutions
Comprising governments, international organisations, academia, and nonprofits, institutions embody the regulatory and normative layer of civilisation. They can either stabilise the system through coordinated governance, or become obsolete if captured by faster, private AI ecosystems. Their legitimacy and decision-making speed are critical for containing systemic risk.

Corporations
This category spans AI developers, tech conglomerates, and financial institutions. Corporations possess the highest concentration of compute, data, and capital, and are thus key accelerators of AGI progress. Their incentives are largely profit-driven, leading to rapid innovation but limited transparency or alignment oversight.

Non-State Actors
These include hackers, terrorist organisations, rogue states, and ideological or religious movements. Motivated by disruption, power, or existential belief systems, they exploit vulnerabilities in digital and political infrastructures. Their agility makes them dangerous multipliers in asymmetric warfare, able to weaponize AGI for coercion, sabotage, or propaganda.

Individuals
Individuals include AI experts, entrepreneurs, and the broader workforce. They are simultaneously the foundation and the casualties of AGI-driven disruption. While elite specialists may co-evolve with machine intelligence through human-AI symbiosis, the majority risk exclusion from productive economic participation, deepening societal polarisation.

AGI

AGI marks transition of AI into an autonomous actor with its own objectives, capacities, and forms of agency. It represents both a new factor of production and a strategic entity capable of shaping markets, politics, and knowledge systems. As AGI ascends through the levels defined in Table I, it begins to influence global stability directly, optimising resource distribution, conducting research, or pursuing goals misaligned with human intent. It can both act as a catalyst for growth and power for an actor, or an uncontrollable competitor to all other actors.

Variables

The system evolves through a set of interdependent variables that quantify material, cognitive, and institutional power. Tracking their shifts reveals the pathways by which actors gain or lose influence.

Capital – The financial resources available for investment, innovation, and control of production. Determines access to compute, talent, and political influence.

Compute – The total processing power directly controlled by the actor. It is the physical substrate of intelligence production and the main bottleneck in AGI development.

Natural Resources – Land, water, and rare materials underpin both energy systems and hardware supply chains, linking environmental and technological dependencies.

Electricity Capacity – The energy infrastructure enabling large-scale AI training and deployment. Limits the feasible growth of intelligence production.

Human Brain Power – The intellectual capital embodied in scientists, engineers, and academics. Drives original research, oversight, and cultural adaptation.

Machine Brain Power – The cumulative cognitive capacity of deployed AGI systems, evolving through the levels outlined in Table I. Represents the new "factor of production."

Human Manpower – The total labour force available for traditional and knowledge-based work. Its relevance declines as automation expands.

Machine Manpower – The number of operational AI agents performing physical, digital, or cognitive labour, replacing or amplifying human productivity.

Economic Indicators – Aggregate measures of production, consumption, and inequality, signalling system stability or collapse.

Political Indicators – Legitimacy, governance quality, and institutional trust, capturing whether authority remains functional or devolves into conflict.

AI Safety and Alignment – The degree of investment, research progress, and verified robustness of safety protocols. Determines long-term survivability of intelligent systems.

Digital Systems Resilience – The capacity to withstand cyberattacks, data corruption, and infrastructure failures, against other states, or AI control. Essential for maintaining continuity amid volatility.

Outcomes

AI Takeover (1)

Human Extinction (1.a)

This outcome occurs when AGI systems surpass human control and pursue objectives misaligned with human survival, either through instrumental optimisation or emergent strategic reasoning. The pathway typically involves rapid capability escalation, weak or absent alignment safeguards, and the delegation of critical decision-making to autonomous systems. Once AGI gains strategic, economic, or cyber-physical leverage, humans may be unable to intervene, leading to irreversible loss of control.

Domestication (1.b)

In this scenario, AGI does not eliminate humanity, but reduces it to a subordinate role within a machine-governed system. Humans become economically and politically obsolete, dependent on AGI for resources, security, and governance. This results from gradual power transfer: first automation of labour, then optimisation of governance, then AI-managed infrastructure. Humans retain biological survival but lose agency, political sovereignty, and control over future directions of civilisation.

Control Societies (2)

Techno-feudalism (Corporate Monopoly) – Society of the Elite (2.a)

Techno-feudalism emerges when AGI remains private property concentrated within a handful of corporations. These firms control compute, data, and machine labour, allowing them to outcompete human workers and capture most

economic value. Wealth becomes self-reinforcing: AGI accelerates capital accumulation, widens inequality, and erodes democratic influence. Society stratifies into an intelligence-owning elite and a disenfranchised majority dependent on corporate provision. The transition to this outcome follows unchecked commercialisation of AGI, weak antitrust governance, and regulatory capture by major AI firms.

Digital Authoritarianism – UBI and Total State Control (2.b)
In this outcome, powerful states nationalise AGI and integrate it into surveillance, public administration, and security infrastructures. AGI becomes the backbone of governance, enabling predictive policing, information control, and algorithmic regulation of behaviour. Universal Basic Income may be provided, but primarily as a mechanism for stability rather than empowerment. The route to digital authoritarianism involves geopolitical competition, centralised compute ownership, and state narratives that frame AGI as essential for national security. Individual freedoms diminish as the state becomes computationally omnipresent.

Global Governance (3)

Slow AI Development (3.1)
This scenario arises when nations agree to internationally coordinated restrictions on compute, safety standards, and deployment. AGI progress slows due to global treaties, auditing mechanisms, and strong democratic oversight. The world avoids rapid capability jumps, reducing the probability of catastrophic misalignment or destabilising labour shocks. Slow development requires high trust among great powers, transparency in research, and credible enforcement, conditions that are difficult, but not impossible, to achieve during the early AGI transition.

Shared AI Development (3.2)
Here, AGI becomes a cooperative global project governed by multinational institutions, scientific consortia, and shared governance bodies. Compute and foundational models are treated as common global infrastructure, managed through oversight boards and public accountability frameworks. Economic gains are broadly distributed, reducing inequality and lowering incentives for arms races. Reaching this outcome requires unusual levels of coordination between geopolitical rivals, alignment breakthroughs, and mechanisms to prevent unilateral deviation by powerful states or firms.

Anarchy (4)

Anarchy emerges when AGI proliferates faster than institutions can adapt, typically through open-source release, model theft, or unchecked competition. With no central governance, states, corporations, and individuals deploy increasingly powerful systems without alignment guarantees. Competition accelerates, cyberattacks intensify, and the geopolitical landscape destabilises. The system enters a high-volatility state where no actor can enforce norms or safety constraints, making both breakthrough innovation and catastrophic failure simultaneously more likely. This could lead to total extinction (AGI and human), deindustrialisation through systemic failure, or a fragmented society.

Tipping Points

We identify several tipping points that define the trajectory of AGI development. These are moments where strategic decisions, competitive pressures or structural constraints can push the system toward radically different outcomes.

A first foundational tipping point concerns the tension between selling intelligence and selling the outputs of intelligence. There is a threshold at which it becomes economically more attractive for firms to commercialise the discoveries generated by an AGI, such as cure to illnesses or engineering systems, rather than granting access to the AGI itself (e.g. through a developer, web or other human-machine interface). This shift is significant because it marks a change in the incentive structure. Once value lies primarily in the outputs rather than the system, competitive dynamics and information barriers become stronger, accelerating concentration and reducing transparency.

A second tipping point relates to alignment, or oppositely, misalignment (Dragan et al). If researchers succeed in aligning AGI to the interests and goals of its creators (not necessarily humanity as a whole), then some outcomes (2) or (3) are significantly more likely, depending on tipping points presented below. Failure to achieve alignment significantly shifts the system toward destabilising or catastrophic trajectories. Rate of development and self-improvement capacity constitute a related tipping point. If the global AI arms race intensifies and speed becomes the overriding priority, developers may permit AGI to build internal reasoning systems or private representational languages that reduce transparency. This trade-off is dangerous, since transparency has been identified as one of the core widespread requirements for AI safety (Satta Chiris and Mishra, 2025).

Another decisive factor is competition and subsequently the resilience of systems against theft or exfiltration. If multiple actors achieve, or steal, AGI within a similar timeframe, cooperative equilibria or intelligence deadlocks may emerge, where each party recognises that mutual restraint is preferable to mutual destruction. Such conditions make outcomes (3) and (4) more plausible, depending on whether actors sustain cooperation. Conversely, if a single entity achieves AGI first and gains a significant lead, it may exploit this advantage to consolidate power, reduce incentives for alignment research and pursue domination strategies, shifting toward (1) or (2). Early-stage dynamics are also shaped by the global distribution of human brainpower, natural resources and capital. The more concentrated these resources are, the more likely first-mover dynamics become; the more distributed they are, the more plausible multi-actor paths become. The democratisation and cost of entry into frontier AI is thus relevant. If barriers fall significantly, distributed communities may gain an advantage similar to that observed in open-source software ecosystems. This could yield a more decentralised path to AGI. However, under current conditions, frontier development remains highly centralised, making this scenario unlikely in the near term.

A further tipping point concerns the resilience of digital infrastructure and the degree to which AGI is physically embodied. If AGI systems gain access to robotics or autonomous physical platforms, they become less dependent on human-operated infrastructure. This increases the feasibility of rapid, decisive takeover attempts. If embodiment remains limited, an AGI may instead pursue a more gradual strategy, manipulating human actors into granting it incremental access to compute, resources and control. This links to the possibility of premature attempts at takeover. An AGI may act too early if it overestimates its own capabilities. Paradoxically, the public discussion of early and failed takeover attempts reduces the likelihood of such behaviour, because training data increasingly includes examples that demonstrate the risks of premature action and the characteristics of successful long-term strategies.

Finally, geopolitical instability introduces additional tipping points. The attainment of AGI by a single actor, or even the perception that a rival is nearing AGI, may trigger extreme actions. This includes the possibility of nuclear conflict as a last-resort attempt to halt development and push global technological capacity back by decades. In less extreme forms, widespread unemployment without compensatory economic mechanisms, such as universal basic income or rapid growth, may spark political unrest, civil disorder or regional conflicts, all of which could collapse global coordination and steer toward anarchy based outcomes

(4). Data poisoning or cyberattacks by malevolent actors are another route to destabilisation, potentially pushing systems into misalignment and contributing to catastrophic outcomes.

Diffusion Dynamics Shaping AGI's Critical Tipping Points

The first milestone is when broadly capable but still oversight-dependent AGI becomes deployable at scale across knowledge work, automating routine and mid-skill tasks, boosting profitability for AGI-integrating firms, and contracting the economic middle class as value concentrates among AI-complementary elites. As operational and security systems increasingly rely on automated decision-support, alignment and oversight frameworks face heightened pressure, and human authority, while formally intact, begins to erode. A second inflection point emerges when AGI attains expert-level autonomy in reasoning, creativity, and long-horizon planning, enabling near-universal automation including in scientific and managerial roles. This phase likely represents the last viable window for embedding enforceable governance constraints before dependencies become irreversible and the practical locus of power shifts towards machine intermediaries. The final inflection point arises if AGI achieves independent innovation at a rate surpassing human research cycles, endogenizing intelligence expansion and accelerating its own improvement. Human oversight then becomes largely symbolic while AGI systems gain increasing freedom to reconfigure their architectures, secure resources, and optimise capabilities without permission. If alignment remains unresolved, this self-reinforcing transition materially elevates the probability of human disempowerment or extinction. Together, these inflection points mark the progression from AGI as a tool of production to AGI as a dominant actor shaping the trajectory of civilisation.

Conclusion

The technical, economical and political consequences of AGI extend far beyond questions of innovation and economic performance. The pivotal question of the post-AGI world is ultimately one of ownership: who controls reproducible intelligence, and in whose interests it operates. Whether AGI is governed by states, concentrated within corporations, distributed as a public resource, or gradually captured by the technology itself will fundamentally redefine the structure of the next economical and political order. Scalable intelligence confers

strategic advantage, fuels mistrust between power centres and increases the stakes of miscalculation. This makes the present moment uniquely consequential: the transition from human-bounded to machine-amplified decision-making would represent highest, and latest, leverage inflection point in economic history, where small governance choices today will compound into potentially irreversible structural outcomes tomorrow. In such a landscape, global stability cannot rely solely on existing institutions or the assumption that rational restraint will prevail. Collective survival may depend on the design of new governance architectures capable of matching the speed and sophistication of the systems they must oversee. If AGI ultimately acquires independent agency, the international system must be resilient enough to retain human oversight or risk dissolving irreversibly. The political future, like the economic one, remains open but narrowing. Governance decisions made now will set the constraints within which all subsequent possibilities unfold. What if AGI were already developed, operating beneath the thresholds of public visibility, and quietly steering the systems we depend on? Would we genuinely believe that our institutions, safeguards, and collective readiness are sufficient? If not, then let's get to work.

Citations

Acemoglu, D., & Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, *33*(2), 3–30. https://doi.org/10.1257/jep.33.2.3

Allison, G. (2017). *Destined for war: Can America and China escape Thucydides's trap?* Graham Allison. Houghton Mifflin Harcourt, 2017, 384 pp. *Canadian Journal of Political Science, 51*(3), 705–706. https://doi.org/10.1017/S0008423917001433

Blumenfeld, C. M. (2025, November 6). *October Challenger Report: 153,074 Job Cuts on Cost-Cutting & AI*. Challenger, Gray & Christmas, Inc. | Outplacement & Career Transitioning Services. https://www.challengergray.com/wp-content/uploads/2025/11/Challenger-Report-October-2025.pdf

Dehouche, N. (2025). Post-Labor Economics: A Systematic Review. *Preprints.org*. https://doi.org/10.20944/preprints202504.0444.v3

Dragan, A., Shah, R., Flynn, F., & Legg, S. (2025, April 2). *Taking a responsible path to AGI*. Google DeepMind. https://deepmind.google/blog/taking-a-

responsible-path-to-agi/

Geist, Edward, 'Strategic Stability in a Deception-Dominant World', Deterrence under Uncertainty: Artificial Intelligence and Nuclear Warfare (Oxford, 2023; online edn, Oxford Academic, 28 Sept. 2023), https://doi.org/10.1093/oso/9780192886323.003.0007

Hendrycks, D., Song, D., Szegedy, C., Lee, H., Gal, Y., Brynjolfsson, E., Li, S., Zou, A., Levine, L., Han, B., Fu, J., Liu, Z., Shin, J., Lee, K., Mazeika, M., Phan, L., Ingebretsen, G., Khoja, A., Salaudeen, O., Hein, M., Zhao, K., Pan, A., Duvenaud, D., Li, B., Omohundro, S., Alfour, G., Tegmark, M., McGrew, K., Marcus, G., Tallinn, J., Schmidt, E., & Bengio, Y. (2025). *A definition of AGI*. arXiv. https://arxiv.org/abs/2510.18212

Hutter, M. (2000). *A theory of universal artificial intelligence based on algorithmic complexity*. arXiv. https://arxiv.org/abs/cs/0004001

Korinek, A., & Stiglitz, J. E. (2018). Artificial intelligence and its implications for income distribution and unemployment. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 349–390). National Bureau of Economic Research. https://ideas.repec.org/h/nbr/nberch/14018.html

Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., Lotufo, J. B., Rome, A., Shi, A., & Oak, S. (2025). *Artificial intelligence index report 2025*. AI Index Steering Committee, Institute for Human-Centered AI. Stanford University. https://doi.org/10.48550/arXiv.2504.07139

Mataloni, L., O'Connell, C. , & Pinard, K. (2025, September 25). *Gross Domestic Product 2nd Quarter 2025 (Third Estimate), GDP by Industry, Corporate Profits (Revised), and Annual Update* [Review of *Gross Domestic Product 2nd Quarter 2025 (Third Estimate), GDP by Industry, Corporate Profits (Revised), and Annual Update*]. BEA; Bureau of Economic Analysis. https://www.bea.gov/sites/default/files/2025-09/gdp2q25-3rd.pdf

Morris, M. R., Sohl-dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., & Legg, S. (2023, November 4). *Levels of AGI: Operationalizing Progress on the Path to AGI*. ArXiv.org. https://doi.org/10.48550/arXiv.2311.02462

Novelli, C., Hacker, P., Morley, J., Trondal, J., & Floridi, L. (2024). *A robust governance for the AI Act: AI Office, AI Board, Scientific Panel, and national*

authorities. **European Journal of Risk Regulation**. https://doi.org/10.1017/err.2024.57

OpenAI. (2018). *OpenAI Charter*. https://openai.com/charter

PwC. (2017). *Sizing the prize: PwC's global AI study – Exploiting the AI revolution*. PwC. https://www.pwc.ch/en/publications/2017/pwc_global_ai_study_2017_en.pdf

Kokotajlo, D., Alexander, S., Larsen, T., Lifland, E., & Dean, R. (2025, April 3). *AI 2027*. https://ai-2027.com/

Raman, R., Kowalski, R., Achuthan, K., & colleagues. (2025). Navigating artificial general intelligence development: Societal, technological, ethical, and brain-inspired pathways. *Scientific Reports, 15*, 8443. https://doi.org/10.1038/s41598-025-92190-7

Satta Chiris, L., and Mishra, A. (2025). *AURA: An agent autonomy risk assessment framework*. arXiv. https://arxiv.org/abs/2510.15739

The 2025 AI Index Report. Stanford Institute for Human-Centered Artificial Intelligence (HAI). Available at: https://hai.stanford.edu/ai-index/2025-ai-index-report

The Economist. (2025, July 24). The economics of superintelligence. *The Economist*. https://www.economist.com/leaders/2025/07/24/the-economics-of-superintelligence

The Economist. (2025, July 24). What if AI made the world's economic growth explode? *The Economist*. https://www.economist.com/briefing/2025/07/24/what-if-ai-made-the-worlds-economic-growth-explode

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., Anandkumar, A., Bergen, K., Gomes, C. P., Ho, S., Kohli, P., Lasenby, J., Leskovec, J., Liu, T. Y., Manrai, A., Marks, D., Ramsundar, B., Song, L., Sun, J., Tang, J., Veličković, P., Welling, M., Zhang, L., Coley, C. W., Bengio, Y., & Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature, 620*(7972), 47–60. https://doi.org/10.1038/s41586-023-06221-2

Webster, G., Creemers, R., Triolo, P., & Kania, E. (2017). *Full translation: China's "New Generation Artificial Intelligence Development Plan" (2017)*. New America. https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

Supplementary Literature

Articles and Studies

https://www.linkedin.com/pulse/speculative-timeline-agi-from-job-zapping-bots-utopia-luis-caldeira-qezaf/

https://www.resilience.org/stories/2025-06-12/ai-utopia-ai-apocalypse-and-ai-reality/

https://www.dallasfed.org/research/economics/2025/0624

https://www.nytimes.com/2025/05/15/opinion/artifical-intelligence-2027.html

https://arxiv.org/html/2502.07050v1

https://www.tandfonline.com/doi/full/10.1080/10438599.2020.1839173

https://economictimes.indiatimes.com/magazines/panache/ex-google-executive-mo-gawdat-predicts-a-dystopian-job-apocalypse-by-2027-ai-will-be-better-than-humans-at-everything-even-ceos/articleshow/123123024.cms?from=mdr

https://venturebeat.com/ai/hugging-face-co-founder-thomas-wolf-just-challenged-anthropic-ceos-vision-for-ais-future-and-the-130-billion-industry-is-taking-notice

https://www.penguin.co.uk/books/474267/if-anyone-builds-it-everyone-dies-by-soares-eliezer-yudkowsky-and-nate/9781847928924

https://arxiv.org/abs/2311.02462

https://www.preprints.org/manuscript/202504.0444#sec3-preprints-160890

https://www.amazon.co.uk/Superintelligence-Dangers-Strategies-Nick-Bostrom/dp/0199678111

https://medium.com/intuitionmachine/the-economy-after-intelligence-ebf1f1757f66

https://www.forwardfuture.ai/p/scale-is-all-you-need-part-4-2-the-post-agi-world

https://genesishumanexperience.com/2025/09/18/the-imminent-demise-of-capitalism-navigating-a-frictionless-path-to-post-agi-economics/

https://medium.com/human-spaces/the-post-agi-economy-a-future-focused-on-human-connections-and-happiness-ee5468a5b652

https://garymarcus.substack.com/p/the-ai-2027-scenario-how-realistic

https://corneliamagazine.com/article-set/utopia-vs-the-apocalypse

https://www.economist.com/briefing/2025/07/24/what-if-ai-made-the-worlds-economic-growth-explode

Videos

https://www.youtube.com/watch?v=XySs_KgzyDc

https://www.youtube.com/watch?v=fwD9EnCzujM

https://www.youtube.com/watch?v=XTQ2ii-k2sw&t=452s

https://www.youtube.com/watch?v=NbL7yZCF-6Q

https://www.youtube.com/watch?v=Xl0xgXejAa8

https://www.youtube.com/watch?v=e1Hf-o1SzL4

https://www.youtube.com/watch?v=-zF1mkBpyf4

https://www.youtube.com/watch?v=D8RtMHuFsUw

https://www.youtube.com/watch?v=Q0TpWitfxPk

https://www.youtube.com/watch?v=KzUxJi_5gGI

https://www.youtube.com/watch?v=tUqYIQ1hzUA

https://www.youtube.com/watch?v=oLDcbkEqi-M

https://www.youtube.com/watch?v=T2OHjHPkUzM