FROM AGI TO ASI: MAPPING THE SOCIETAL AND SYSTEMIC IMPACT OF SUPERINTELLIGENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

We examine the implications of Artificial General Intelligence (AGI) and its potential transition to Artificial Superintelligence (ASI) across economics, law, health, education, and belief systems. Treating advanced AI as a systemic force, we synthesize domain-specific disruptions and emphasize cross-domain feedback loops that can amplify both benefits and risk. We discuss impacts for developing nations and present compact dynamical-systems sketches illustrating cascades and recovery in fragile socio-technical infrastructures. We conclude with governance priorities for navigating rapid capability gains and dependency risks.

1 INTRODUCTION

Artificial General Intelligence (AGI) denotes machine capability to perform any intellectual task a human can do (Goertzel, 2007). If such systems progress to Artificial Superintelligence (ASI) (Bostrom, 2014), institutional disruptions may be rapid and tightly coupled across sectors. We provide a brief synthesis of plausible impacts and highlight systemic risk mechanisms, focusing on feedback, dependency, and concentration of power.

2 DOMAIN IMPACTS (BRIEF SYNTHESIS)

Economics. Automation of cognitive labor may reconfigure labor markets and bargaining power (Acemoglu & Restrepo, 2018). If control of frontier systems concentrates, societies may face extreme inequality and techno-feudal dynamics (Zuboff, 2019). Redistribution, ownership, and safety regulation become central.

Law and governance. AGI strains legal notions of responsibility, intent, and liability (Pagallo, 2013). Governments may adopt autonomous decision support, raising transparency and legitimacy concerns; the same tools can accelerate malicious use and strategic instability (Brundage et al., 2018). International governance remains difficult under competitive pressures.

Health and wellbeing. Advanced AI can improve diagnostics and treatment (Topol, 2019), but may also intensify surveillance, bias, and dependency. In mental health, systems that optimize engagement or compliance can undermine autonomy if incentives are misaligned.

Education. Personalized tutoring and assessment could scale high-quality instruction (Luckin et al., 2016). However, heavy reliance on AI may deskill educators, erode epistemic agency, and introduce ideological or commercial bias, particularly where curricula are outsourced.

Consciousness and moral status. If machine consciousness is possible (Chalmers, 1996), moral status and welfare risks (including simulated suffering) become relevant (Bostrom & Yudkowsky, 2011). Even without consciousness, persuasive systems may change social norms and reshape moral reasoning.

Religion and philosophy. Superhuman systems can challenge belief structures grounded in human exceptionalism Holm & Banerjee (2024). New technotheological movements may arise; institutional supports may be needed to avoid crises of meaning and conflict. Some perspectives argue for drawing on existing traditions to articulate norms for coexistence with superintelligence Banerjee (2025).

3 Systemic risk and developing-nation exposure

Risks are amplified by *interdependence*: failures propagate when health, infrastructure, governance, and markets share AI services, data pipelines, and optimization objectives. Developing nations may experience higher vulnerability due to thinner institutional buffers, infrastructure fragility, and limited bargaining power in global AI supply chains. At the same time, leapfrogging benefits are plausible if access is equitable (e.g., low-cost tutoring, diagnostics, and agricultural planning), suggesting that governance should prioritize resilience, redundancy, and local capacity.

3.1 Failure modes and levers

Dependency traps. When critical services (payments, logistics, triage, benefits delivery) are optimized around a single AI layer, outages and model regressions can create sudden, correlated disruptions. Organizations may lose the ability to operate manually, making recovery slower than the initial failure.

Power concentration. Frontier compute, data advantages, and deployment channels can produce winner-take-most dynamics. This can shift governance from public institutions to private infrastructures, complicating accountability and due process.

Information integrity. As synthetic media and automated persuasion scale, societies may face persistent uncertainty about authenticity, which can undermine elections, public health messaging, and conflict de-escalation.

Policy levers. Practical mitigations include stress-testing and red-teaming of AI-enabled institutions, enforced fallback procedures, procurement rules that require interoperability, and investments in local technical capacity. For developing nations, shared public infrastructure (e.g., open evaluation suites, auditing support, and regional incident reporting) can reduce dependence on external vendors.

3.2 Benefit pathways

AGI also offers plausible upside if governed well: rapid translation and tutoring can broaden educational access; diagnostic support can extend scarce clinical expertise; and planning tools can improve agricultural yields and disaster response. The central question is not whether capability grows, but whether institutions can *absorb* it without locking in brittle dependencies or inequitable control.

3.3 Failure modes and levers

Dependency traps. When critical services (payments, logistics, triage, benefits delivery) are optimized around a single AI layer, outages and model regressions can create sudden, correlated disruptions. Organizations may lose the ability to operate manually, making recovery slower than the initial failure.

Power concentration. Frontier compute, data advantages, and deployment channels can produce winner-take-most dynamics. This can shift governance from public institutions to private infrastructures, complicating accountability and due process.

Information integrity. As synthetic media and automated persuasion scale, societies may face persistent uncertainty about authenticity, which can undermine elections, public health messaging, and conflict de-escalation.

Policy levers. Practical mitigations include stress-testing and red-teaming of AI-enabled institutions, enforced fallback procedures, procurement rules that require interoperability, and investments in local technical capacity. For developing nations, shared public infrastructure (e.g., open evaluation suites, auditing support, and regional incident reporting) can reduce dependence on external vendors.

3.4 BENEFIT PATHWAYS

AGI also offers plausible upside if governed well: rapid translation and tutoring can broaden educational access; diagnostic support can extend scarce clinical expertise; and planning tools can improve agricultural yields and disaster response. The central question is not whether capability grows, but whether institutions can *absorb* it without locking in brittle dependencies or inequitable control.

4 COMPACT DYNAMICAL SKETCH (CASCADE & RECOVERY)

To illustrate cascades, consider a minimal coupled model with (i) AI reliability $A(t)$, (ii) infrastructure robustness $I(t)$, and (iii) policy capacity $P(t)$:

$$\dot{A} = r_A A(1 - A) - \lambda_A (1 - I)A - s(t), \tag{1}$$

$$\dot{I} = r_I (1 - I) - \lambda_I (1 - A)I + \eta_P P, \tag{2}$$

$$\dot{P} = r_P (1 - P) - \lambda_P (1 - I)P, \tag{3}$$

where $s(t)$ is an exogenous shock (e.g., sudden degradation or misuse). The key qualitative point is the *positive feedback*: reduced A degrades I, which further reduces A via dependency; stronger policy capacity P can dampen cascades by sustaining I.

4.1 INTERPRETATION AND GOVERNANCE MAPPING

The model is intentionally abstract, but it suggests three governance-relevant design principles. First, *increase separability*: avoid single points of AI failure by diversifying suppliers, models, and operating procedures so that A does not fully determine I. Second, *make recovery faster than propagation*: invest in incident response, monitoring, and rollback so that shocks $s(t)$ are detected early and the effective coupling λ_I, λ_A is reduced during crises. Third, *treat policy capacity as infrastructure*: the term $\eta_P P$ captures that competent oversight (audits, procurement, standards, and enforcement) can directly raise robustness even when AI reliability fluctuates.

In practice, these translate to (i) mandatory fallback pathways for critical services, (ii) regular stress tests and public reporting for AI-dependent systems, and (iii) international coordination on evaluation and misuse prevention to reduce arms-race incentives.

4.2 INTERPRETATION AND GOVERNANCE MAPPING

The model is intentionally abstract, but it suggests three governance-relevant design principles. First, *increase separability*: avoid single points of AI failure by diversifying suppliers, models, and operating procedures so that A does not fully determine I. Second, *make recovery faster than propagation*: invest in incident response, monitoring, and rollback so that shocks $s(t)$ are detected early and the effective coupling λ_I, λ_A is reduced during crises. Third, *treat policy capacity as infrastructure*: the term $\eta_P P$ captures that competent oversight (audits, procurement, standards, and enforcement) can directly raise robustness even when AI reliability fluctuates.

In practice, these translate to (i) mandatory fallback pathways for critical services, (ii) regular stress tests and public reporting for AI-dependent systems, and (iii) international coordination on evaluation and misuse prevention to reduce arms-race incentives.

Figure 1 shows an illustrative stress scenario (shock at t=10) where infrastructure partially recovers while the AI subsystem degrades, consistent with asymmetric recovery dynamics in coupled systems.

5 CONCLUSION

AGI/ASI should be treated as a systemic intervention: it can increase capability across domains while introducing correlated failures, power concentration, and dependency. Priority actions include (i) robust evaluation and monitoring, (ii) governance mechanisms with international coordination,

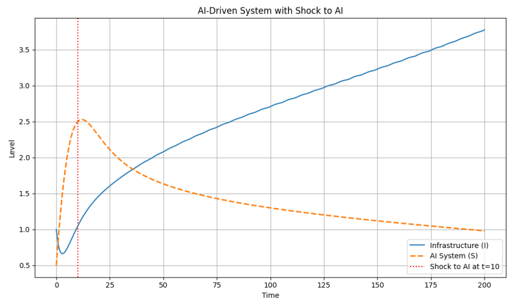

Figure 1: Stress scenario simulation of AI-driven systemic risk. A shock to AI reliability at $t = 10$ reduces AI contributions and affects infrastructure (partial recovery), while the AI subsystem continues to degrade.

(iii) resilience investments (redundancy, fallback procedures, and human-in-the-loop safeguards), and (iv) equitable access and local capacity-building to reduce developing-nation fragility.

References

Daron Acemoglu and Pascual Restrepo. Artificial intelligence, automation, and work. *NBER Working Paper No. 24196*, 2018.

Soumya Banerjee. Ai and the future of the technosphere: A path towards co-existence with superintelligence. In *AAAI 1st Workshop on Post-Singularity Symbiosis*, 2025. URL `https://osf.io/8gauc`.

Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.

Nick Bostrom and Eliezer Yudkowsky. The ethics of artificial intelligence. 2011.

Miles Brundage et al. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228*, 2018.

David J. Chalmers. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, 1996.

Ben Goertzel. Artificial general intelligence: Concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, 2007.

Halfdan Holm and Soumya Banerjee. Intelligence in animals, humans and machines: a heliocentric view of intelligence? *AI and Society*, pp. 1–3, 4 2024. ISSN 14355655. doi: 10.1007/S00146-024-01931-1/METRICS. URL `https://link.springer.com/article/10.1007/s00146-024-01931-1`.

Rose Luckin, Wayne Holmes, Mark Griffiths, and Laurie Forcier. *Intelligence Unleashed: An Argument for AI in Education*. Pearson, 2016.

Ugo Pagallo. The laws of robots: Crimes, contracts, and torts. *Springer*, 2013.

Eric Topol. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books, 2019.

Shoshana Zuboff. *The Age of Surveillance Capitalism*. PublicAffairs, 2019.