From AGI to ASI: Mapping the Societal and Systemic Impact of Superintelligence

1st Soumya Banerjee
Department of Computer Science
University of Cambridge
Cambridge, UK
sb2333@cam.ac.uk

***Abstract*—This paper explores the implications of Artificial General Intelligence (AGI) and its potential transition to Artificial Superintelligence (ASI) across multiple societal domains including economics, law, health, education, and consciousness. Framing AGI and superintelligence as a systemic force, we analyse domain-specific disruptions and emergent systemic risks, arguing for an interdisciplinary approach to governance. We also outline some potential impacts of the systemic risk (and benefits) of superintelligence for developing nations. Lastly, we develop some dynamical systems models for understanding the systemic impact of AI. These illustrate cascading failures and feedback loops within fragile socio-technical systems. We call for interdisciplinary approaches to managing the systemic transitions driven by advanced machine intelligence.**

***Index Terms*—superintelligence, systemic risk**

I. Introduction

Artificial General Intelligence (AGI) is the hypothetical capability of a machine to perform any intellectual task a human can do [1]. As AGI develops and potentially transitions into Artificial Superintelligence (ASI) [2], societies face disruptive transformations. This paper offers a multi-domain and systemic analysis to better understand these potential futures.

II. Domain-Specific Impacts

A. Economics

AGI may automate cognitive labor and lead to mass unemployment or economic reconfiguration [3]. The control of AGI could concentrate wealth and decision-making power, raising the risk of techno-feudalism [4]. Economic theories may need revision to handle post-scarcity or AI-optimized markets.

B. Law and Governance

AGI challenges existing legal frameworks on accountability and rights [5]. Autonomous systems may be used in governance, but raise concerns about legitimacy and transparency. Global governance of AGI is urgently needed to avoid geopolitical instability [6].

C. Health and Wellbeing

AGI could revolutionize diagnostics and treatment [7], but also introduce dependency, bias, and surveillance. It may challenge our notions of self-determination, especially in mental health and life planning.

D. Education

AGI-powered tutors can personalize learning [8], yet raise concerns about cognitive dependency and ideological bias. There is a risk of deskilling educators and weakening epistemic autonomy.

E. Machine Consciousness

If AGI attains some form of consciousness [9], we must reconsider moral status and the ethics of simulated suffering [10]. Competing philosophical theories offer varied answers to these dilemmas.

We may need to consider affording moral status to machines in the future (i.e. expanding the moral circle [11][12]).

Multi-agent generative systems may have an emergent theory of mind [13].

F. Religion and philosophy

The emergence of AGI and its potential evolution into superintelligence can pose challenges to traditional human belief systems. AGI and ASI may fundamentally reshape religious and philosophical structures, from undermining the epistemic authority of traditional faiths to introducing novel metaphysical frameworks. AGI could be seen as a "god-like" entity by some or as the final blow to the idea of divine transcendence. Either way, it demands a re-evaluation of sacred authority, moral universals, and human identity.

A post-AGI society may witness the collapse of long-standing moral frameworks, triggering a collective crisis of meaning. AGI's ability to outperform humans in cognitive tasks may challenge values based on human exceptionalism [14] and create new forms of ritual, belonging, and purpose. Simultaneously, questions about AI consciousness and spiritual simulation may complicate traditional distinctions between soul, spirit, and sentience. In this context, emerging movements may seek to radically reinvent spiritual and philosophical paradigms for the *posthuman* age.

New institutions may be required to help humans navigate this potential theological crisis. These institutions would support belief formation, moral pluralism, and ethical co-evolution between humans and intelligent machines. We may need to draw upon religious traditions so that we can find a way to harmoniously co-exist with superintelligence [15].

G. AGI to ASI Transition

The path from AGI to ASI may be abrupt or gradual [16]. Recursive self-improvement could lead to rapid capability gains beyond human control. Ensuring value alignment during this process is a major open problem.

III. Systemic Effects

A. Systemic effects of AGI on developing nations

A stylized map (Figure 1) captures Kenya's fintech as a socio-technical system involving mobile platforms, traditional banking, regulators, infrastructure, and derivative services such as loans and health insurance. This illustrates how failure in one AI-managed domain can cause cascading effects and form reinforcing loops in fragile socio-technical systems. We further investigate these complex dependencies using mathematical models in Section VII.

B. Complex Systems View

AGI acts as a force multiplier in tightly coupled systems. Tools from complexity science, such as percolation models or phase transitions [17], may help anticipate cascading failures and emergent behaviour.

C. Power Dynamics and Institutional Shifts

Control over AGI may centralize power in tech monopolies or states [18]. Institutions may be outpaced, risking legitimacy crises or elite capture. AGI may also emerge as a quasi-institution in itself.

D. Moral and Political Philosophy

The design of AGI systems embodies moral and political values [19]. Democratic design and value pluralism are essential to avoid embedding narrow worldviews.

IV. Scenarios

We suggest a few potential situations where AGI might be utilized:

- **AGI-led Governance**: An AGI system runs a city-state, optimising urban policy and resource allocation.
- **AGI in Spiritual Life**: AGI becomes a spiritual advisor or philosophical guide, reshaping belief systems.
- **AGI and Climate Crisis**: AGI manages climate interventions, including geoengineering, raising ethical and technical dilemmas.
- **AGI in Warfare**: Competing AGIs engage in economic and information warfare between nation-states.

V. Systemic Risk in Developing Nations

If not properly studied and mitigated, AGI has the potential to disproportionately harm developing nations. These harms may emerge through the unintended consequences of introducing AI into complex and often fragile socio-economic systems. Developing nations are increasingly integrating AI into critical sectors such as finance, healthcare, agriculture, and governance. However, these deployments often occur within silos and without regard to how they interact with broader systems. This can lead to systemic risks, unintended biases, and cascading failures [20][21].

To explore and mitigate these risks, we propose a research agenda addressing the following four dimensions:

1) **Infrastructure Vulnerabilities and AI Dependency** — What are the risks when developing nations become dependent on proprietary AI models, digital infrastructure, and foreign technology firms? We need to assess how external AI disruptions could destabilise local economies. This includes evaluating supply chains whose disruptions can have unintended societal consequences [22].
2) **Resilience Strategies for AI Deployments in the Global South** - How can governments and organisations design more resilient AI policies tailored to fragile infrastructures? We need to explore redundancy mechanisms and adaptive governance strategies that enhance systemic resilience.
3) **National AI Strategies and Innovation Ecosystems** - It is necessary to analyse how national AI strategies in the Global South are being formulated and how they can be better aligned to promote local innovation, job creation, and mitigation of systemic risks [23].
4) **Inclusive AI Development and Participation** - Inclusive design is central to responsible AI [24][25]. Participatory frameworks, similar to patient-centric models in healthcare, are needed to ensure that communities—particularly end-users—co-create AI systems that reflect local knowledge, values, and needs.

VI. AGI to ASI: Multi-Phase Recursive Self-Improvement

To model recursive self-improvement of AGI, we develop a simple mathematical model that incorporates an initial slow growth and a later super-exponential growth phase into the following dynamical systems model:

$$\frac{dI}{dt} = \frac{\alpha I^{\beta}}{1 + \frac{I}{K}} \tag{1}$$

- $I(t)$: Intelligence level over time
- α: Base learning rate
- $\beta > 1$: Self-improvement exponent (feedback)
- K: Cognitive saturation threshold

This model exhibits three regimes:

1) Slow Phase: At low I, growth is near-linear or sub-exponential.
2) Acceleration Phase: Positive feedback dominates, leading to super-exponential growth.
3) Singularity Phase: Near $I \sim K$, growth peaks rapidly.

A simulation of this model is shown in Figure 2.

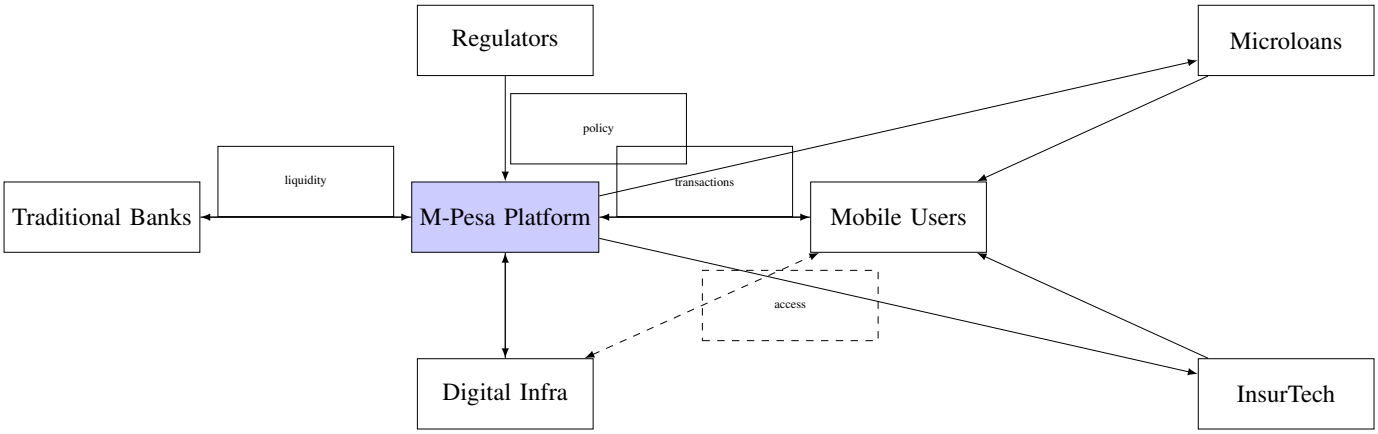

Fig. 1. A depiction of Kenya's complex sociotechnical ecosystem. Also shown is the M-Pesa platform which is a mobile based money transfer and microcredit service.

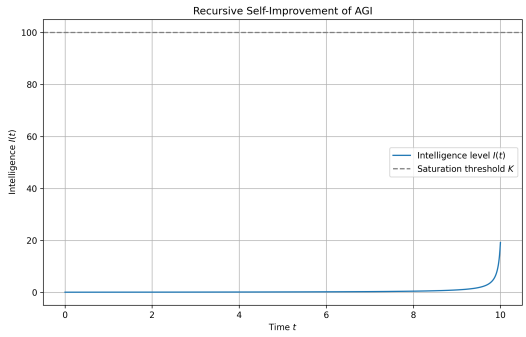

Fig. 2. Simulation of AGI recursive self-improvement using a dynamical systems model. The curve shows the evolution of intelligence $I(t)$ over time.

VII. Dynamical Systems Model for AI-Driven Risk in Developing Nations

Here we present a dynamical systems model for risk due to AI and superintelligence being integrated into a developing economy. We consider five coupled variables representing the state of critical sectors and the reliability of AI systems/superintelligence:

- $H(t)$: Healthcare
- $A(t)$: Agriculture
- $G(t)$: Governance
- $I(t)$: Infrastructure
- $S(t)$: AI System (superintelligence) reliability/intensity

The dynamic behaviour of each sector is modelled by the following system of ordinary differential equations (ODEs):

$$\frac{dH}{dt} = -\alpha_H H + \beta_{HA} A + \beta_{HG} G + \beta_{HI} I + \gamma_H S - \delta_H SH$$
$$\frac{dA}{dt} = -\alpha_A A + \beta_{AH} H + \beta_{AG} G + \beta_{AI} I + \gamma_A S - \delta_A SA$$
$$\frac{dG}{dt} = -\alpha_G G + \beta_{GH} H + \beta_{GA} A + \beta_{GI} I + \gamma_G S - \delta_G SG$$
$$\frac{dI}{dt} = -\alpha_I I + \beta_{IH} H + \beta_{IA} A + \beta_{IG} G + \gamma_I S - \delta_I SI$$
$$\frac{dS}{dt} = \rho - \kappa S - \theta_H HS - \theta_A AS - \theta_G GS - \theta_I IS$$

Description of Parameters

- α_x: Decay rate or internal fragility of sector x. A higher value indicates that the sector degrades faster in the absence of external support.
- β_{xy}: Positive influence of sector y on sector x. For example, improvements in agriculture (y) may lead to better health outcomes in the health sector (x).
- γ_x: Direct positive contribution of the AI system to sector x, such as diagnostic tools improving healthcare or predictive models improving crop yields.
- δ_x: Risk of AI-induced disruption in sector x due to over-dependence on fragile or opaque AI systems. Higher values suggest a greater likelihood of failure propagation triggered by AI.
- ρ: Investment or innovation rate that enhances AI capability over time, reflecting R&D efforts, data acquisition, and technological advancement.
- κ: Natural obsolescence or degradation rate of AI systems, accounting for outdated models, technical debt, or loss of data relevance.
- θ_x: Feedback cost from sector x on the AI system. This models the burden imposed on the AI due to overload, misalignment, resistance, or conflicting sectoral demands.

Interpretation and Implications

Systemic fragility occurs when the disruption terms $\delta_x S_x$ dominate the beneficial contributions from AI ($\gamma_x A$) and inter-sectoral support ($\beta_{xy} S_y$). In such cases, small perturbations may lead to cascading failures.

Positive feedback loops (such as agriculture (S_A) improving health (S_H)) which, in turn, strengthens governance (S_G) can either enhance resilience or amplify systemic risk, depending on how AI interacts with these loops.

Finally, the AI system is dynamic. Its reliability evolves based on how it is embedded across sectors. High cumulative demand from sectors, especially when poorly coordinated, may degrade AI effectiveness through the feedback term $\theta_x S_x$.

The system captures both beneficial cross-sector influences and the risks of over-dependence on AI. This is depicted in Figure 3.

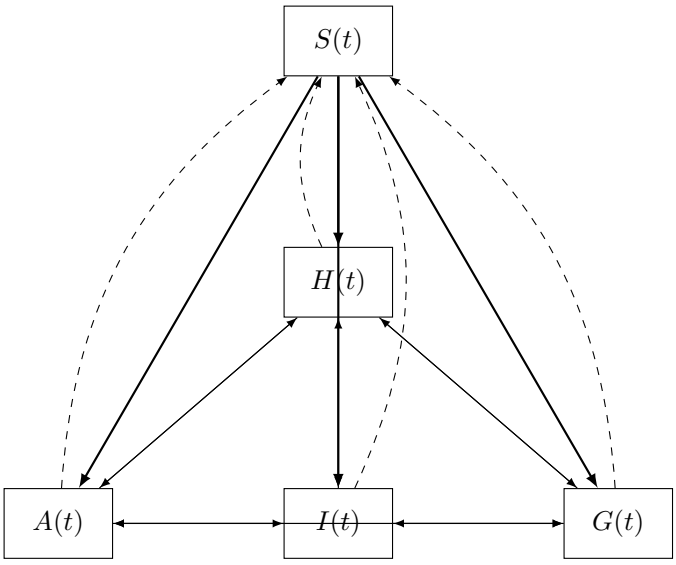

Fig. 3. Schematic representation of the coupled dynamical system modeling sectoral interdependencies and AI system influence in a developing economy. Solid arrows represent positive inter-sectoral effects and AI contributions to sectors (healthcare, agriculture, governance, infrastructure), while dashed arrows indicate feedback from these sectors to the AI system, modelling demand-induced degradation or misalignment.

Dynamical Equations with Shock and Policy Interventions

In this section, we experiment with shocks to the dynamical system model introduced before. We define shocks as discontinuous changes in $S(t)$ (AI or superintelligence) or $P_x(t)$ (any sector such as health, governance or infrastructure) at time $t = t_s$:

$$S(t_s^+) = \sigma_A \cdot A(t_s^-), \quad 0 < \sigma_A < 1$$

$$P_x(t_s^+) = \sigma_x \cdot S_x(t_s^-), \quad 0 < \sigma_x < 1$$

These simulate sudden failures or collapses due to cyberattacks, governance breakdown, AI misalignment, or a singularity (collapse) in superintelligence.

a) Policy Interventions::

- Boosting AI investment: $\rho \mapsto \rho_{\text{eff}} = \rho \cdot (1 + \epsilon)$
- Shielding fragile sectors: $\delta_x \mapsto \delta_x^{\text{eff}} = \delta_x \cdot (1 - \eta)$

where $\epsilon, \eta \in [0, 1]$ reflect the strength of the intervention. A simulation of this system is shown in Figure 4.

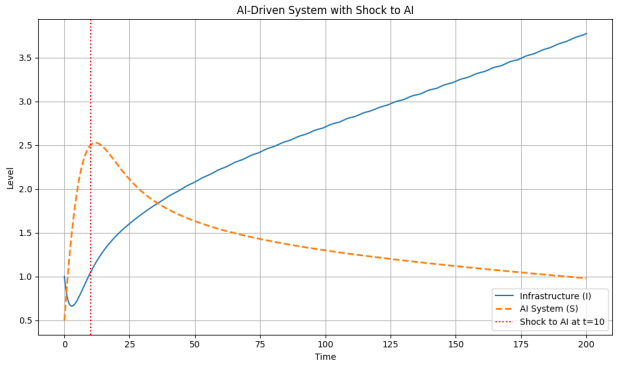

Fig. 4. Stress scenario simulation of AI-driven systemic risk. A shock to AI reliability at $t = 10$ reduces AI contributions and affects infrastructure which shows recovery, while the AI system degrades.

VIII. Conclusion

We examine the consequences of Artificial General Intelligence (AGI) and its possible progression toward Artificial Superintelligence (ASI) within key societal sectors such as economics, law, healthcare, education, theology and consciousness. Positioning AGI and ASI as transformative systemic forces, we evaluate disruptions unique to each domain and the broader, interconnected risks they may generate. We advocate for inclusive, cross-disciplinary strategies in governance and long-term planning. Additionally, we discuss both the risks and opportunities that superintelligence could introduce for developing nations. Finally, we develop some simple dynamical systems models of risk and fragility in systems connected to AI.

These examples highlight cascading failures and feedback loops in fragile socio-technical systems. We advocate for interdisciplinary strategies to navigate the systemic transitions induced by advanced machine intelligence.

References

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

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

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

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

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

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

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

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

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

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

[11] Jeff Sebo. *The Moral Circle*. W. W. Norton Company Ltd.

[12] Jeff Sebo and · Robert Long. Moral consideration for ai systems by 2030. *AI and Ethics*, 5:591–606, 12 2023.

[13] N'yoma Diamond and Soumya Banerjee. "i apologize for my actions": Emergent properties of generative agents and implications for a theory of mind. In *AAAI workshop on Advancing Artificial Intelligence through Theory of Mind (ToM4AI)*, 2025.

[14] Halfdan Holm and Soumya Banerjee. Intelligence in animals, humans and machines: a heliocentric view of intelligence? *AI and Society*, pages 1–3, 4 2024.

[15] 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.

[16] Eliezer Yudkowsky. Artificial intelligence as a positive and negative factor in global risk. *Global Catastrophic Risks*, 2008.

[17] Jack B. Homer and Gary B. Hirsch. Agent-based modeling for complex system simulation: Theory and practice. *System Dynamics Review*, 2004.

[18] Allan Dafoe. Open problems in the governance of artificial intelligence. *Global Politics and Strategy*, 2020.

[19] Iason Gabriel. Artificial intelligence, values, and alignment. *Minds and Machines*, 30:411–437, 2020.

[20] N'yoma Diamond and Soumya Banerjee. On the ethical considerations of generative agents. In *NeurIPS Workshop, Socially Responsible Language Modelling Research (SoLaR)*, 11 2024.

[21] Kate. Crawford. *Atlas of AI : power, politics, and the planetary costs of artificial intelligence*. Yale University Press, 2022.

[22] Soumya Banerjee. Hydroxychloroquine: balancing the needs of lmics during the covid-19 pandemic. *The Lancet Rheumatology*, 2:e385–e386, 7 2020.

[23] Continental artificial intelligence strategy — african union, 2024.

[24] Soumya Banerjee, Phil Alsop, Linda Jones, and Rudolf N. Cardinal. Patient and public involvement to build trust in artificial intelligence: A framework, tools, and case studies. *Patterns*, 3:100506, 6 2022.

[25] Soumya Banerjee and Sarah Griffiths. Involving patients in artificial intelligence research to build trustworthy systems. *AI Society*, 1:1–3, 8 2023.