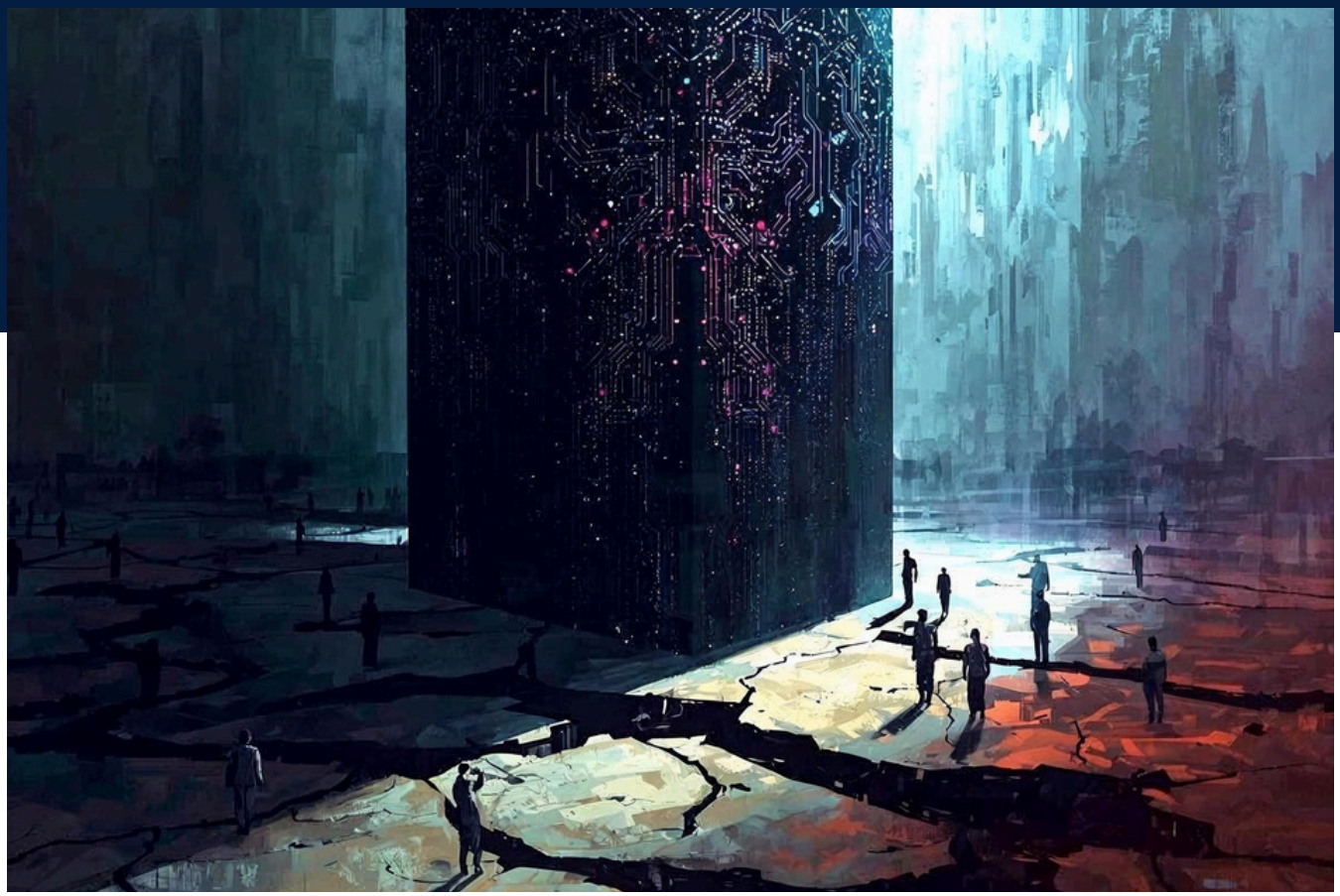
RESEARCH REPORT
2025 OCTOBER

AI GOVERNANCE INITIATIVE

OXFORD MARTIN SCHOOL

UNIVERSITY OF OXFORD

Agentic Inequality



Authors: Matthew Sharp, Omer Bilgin, Iason Gabriel, Lewis Hammond

Agentic Inequality

Matthew Sharp[1,2] Omer Bilgin[3] Iason Gabriel[4] Lewis Hammond[3,5]

21 October 2025

[1]Access Partnership [2]Oxford Martin AI Governance Initiative [3]University of Oxford
[4]School of Advanced Study [5]Cooperative AI Foundation

Abstract

The development of autonomous AI agents, capable of complex planning and action, represent a significant technological evolution beyond current generative tools. As these systems become integrated into political and economic life, their distribution and capabilities will be highly consequential. This paper introduces and explores "agentic inequality" – the potential disparities in power, opportunity, and outcomes stemming from differential access to, and capabilities of, AI agents. We analyse the dual potential of this technology, exploring how agents could both exacerbate existing divides and, under the right conditions, serve as a powerful equalising force. To this end, the paper makes three primary contributions. First, it establishes an analytical framework by delineating the three core dimensions through which this inequality can manifest: disparities in the *availability*, *quality*, and *quantity* of agents. Second, it argues that agentic inequality is distinct from prior technological divides. Unlike tools that primarily augment human abilities, agents act as autonomous delegates, creating novel power asymmetries through scalable goal delegation and direct agent-to-agent competition that are poised to reshape outcomes across economic and socio-political spheres. Finally, it provides a systematic analysis of the technical and socioeconomic drivers – from model release strategies to market incentives – that are likely to shape the distribution of agentic power, concluding with a research agenda for navigating the complex governance challenges ahead.

1. Introduction

Generative AI models – typified by systems that convert prompts into prose, code or images – are now serving as the foundation for AI agents: systems that can pursue complex goals with significant autonomy by perceiving their environment, planning, and executing multi-step tasks, often using software tools across a range of digital environments (Kasirzadeh and Gabriel, 2025; Wang et al., 2024). While still in their infancy, AI agents are beginning to execute economically valuable workflows, and ongoing research aims to significantly extend their capabilities. This shift from AI as a tool, which executes commands, to AI as an autonomous actor, which pursues goals independently, is significant because it directly impacts human agency – the capacity of individuals to pursue goals that they value (Sen, 1999; Prunkl, 2024). As a powerful resource and means of pursuing goals, AI agents can amplify this capacity. Furthermore, the societal allocation of AI agents could reshape the distribution of human agency, posing new challenges for distributive justice that demand novel frameworks for analysis and governance.[1]

Correspondence:
matthew.sharp@accesspartnership.com;
omer.bilgin@reuben.ox.ac.uk; iason.gabriel@sas.ac.uk;
lewis.hammond@cooperativeai.org.

[1] In the context of justice, the philosopher John Rawls famously argued for a focus on the distribution of "social primary goods". These goods encompass things that serve as "all-purpose means" i.e. goods and resources

This paper introduces and explores "agentic inequality" – the potential disparities in power, opportunity, and outcomes stemming from differential access to, and capabilities of, AI agents. We argue this raises questions that reach beyond those encountered by prior technological divides. Its unique character arises from two novel mechanisms: the creation of new power asymmetries through the scalable delegation of complex tasks, and new competitive dynamics forged through direct agent-to-agent interactions. While this analysis relates to broader concerns about the political economy of AI, our focus is narrowly on the specific disparities that emerge from the unique characteristics of AI agents. As these systems are not yet widely deployed, a window of opportunity exists to proactively shape their development and ensure their benefits are distributed equitably.

This paper builds on a long tradition of scholarship examining how technology and society coevolve: creating, reinforcing, or disrupting social hierarchies (Mumford, 1970; Jasanoff, 2004). It extends research on the socioeconomic impacts of large language models (LLMs), (Acemoglu and Restrepo, 2020; Brynjolfsson, Li and Raymond, 2025; Noy and Zhang, 2023) using this work as a foundation to analyse the newer, less-adopted technology of AI agents. The paper also extends research on the "digital divide", which has focused on disparities in technology access (Warschauer, 2003).

In this paper, we develop an analytical framework for understanding this emerging challenge, exploring both the risks of new forms of inequality and the potential for AI agents to mitigate existing divides. Indeed, the very sophistication that allows agents to be a powerful resource for some could also allow them to be powerful equalisers for others. For example, universally accessible agents could enhance baseline access to information, make complex processes more accessible, or automate routine tasks for all, thereby mitigating certain pre-existing disadvantages. A critical step involves shifting our evaluative lens to view AI agents not just as individual products, but as a form of "infrastructure" that is likely to mediate access to essential goods, services, and opportunities (Plantin et al., 2018; Rothschild et al., 2025; Bergman et al., 2024; Chan et al., 2025; Lazar, 2025). Viewing agents through this infrastructural lens reveals how their design and distribution can systematically channel benefits and create barriers.

To analyse this emerging challenge, we proceed as follows.[2] In Section 2, we delineate the key dimensions through which agentic inequality may manifest. In Section 3, we subsequently examine its potential impacts across crucial societal domains. In Section 4, we explore the technical and socio-political drivers that could either propel or reduce this inequality. Finally, in Section 5, we analyse the complexities of governing agentic inequality and propose a research agenda exploring potential interventions, such as public investment in agent infrastructure and the development of common standards for interoperability.

2. Dimensions of Agentic Inequality

To systematically analyse agentic inequality, we first establish a framework that sets out its constitutive dimensions. More precisely, this section distinguishes between three core dimensions: (1) *availability* of an agent, (2) the *qualities* of agents, and (3) the *quantity* of agents. Understanding these dimensions is helpful to clarify the differential extents to which individuals and organisations can derive benefits from agentic AI.

that people can use to pursue their conception of the good life irrespective of its content (Rawls, 2001). Income and wealth are important because they can be used by people to achieve a higher level of goal fulfillment, across a wide spectrum of worldviews and beliefs. Moving forward, AI agents may fulfil a similar role - allowing people to get more of what they value across the board. On the Rawlsian view, the distribution of AI agents would then become a matter of distributive justice.

[2] While we focus our analysis primarily on individuals and organisations/firms, we acknowledge that the implications of agentic inequality extend to all aspects of society.

2.1. Availability

Availability is the most foundational dimension, signifying a binary divide between those who can utilise even a single AI agent and those who cannot. This "access gap" threatens to amplify existing digital divides, where individuals or organisations are precluded from the opportunities agentic AI affords due to technological, geographic, or socioeconomic barriers. A crucial distinction must be made between access to an underlying foundation model and access to a functional agent. While foundation models are increasingly accessible, deploying a capable agent is comparatively difficult. The necessary agentic infrastructure – such as high-quality scaffolding – is still new, complex, and often restricted to a handful of providers, requiring specialised expertise to deploy effectively (Chan et al., 2025; Liu et al., 2025). This creates significant awareness and knowledge gaps that constitute a primary barrier to entry. It is also worth noting that inequalities in availability can be a matter of choice, if individuals or entities intentionally refuse to employ agents for ethical or practical reasons. In either case, some groups in society could be positioned to leverage agents for productivity and efficiency gains, while others will not.

2.2. Quality

Beyond mere availability, the *quality* of an agent constitutes a second critical dimension. This refers to what an individual agent can do and the operational characteristics that determine its behaviour. Agent *quality* can manifest along several axes:

- **Core intelligence:** This comprises an agent's world knowledge, its multimodal proficiency with language and forms of data, and its raw capacity for complex reasoning, planning, and problem-solving. It is largely determined by the sophistication of the underlying foundation model and its training data (Chen, 2025; Zhu et al., 2025).
- **Operational speed and throughput:** An agent's efficiency is determined by its latency and data processing capabilities, which are a function of the computational resources (e.g. hardware accelerators and cloud infrastructure) underpinning its operations (Sastry et al., 2024).
- **Reliability:** An agent's robustness and failure rate on set tasks is another key quality, dictating its consistency and trustworthiness.
- **Tool use:** This refers to an agent's ability to access and meaningfully utilise external affordances, such as APIs, proprietary databases, real-time data feeds, or even physical actuators (Chan et al., 2025).
- **Disposition:** An agent has tendencies or propensities to behave in certain ways which influence its performance in different environments. Whether a certain disposition is advantageous is highly context-sensitive. For example, an aggressive disposition may be advantageous for a deal-negotiating agent but detrimental in a customer service role where a polite and helpful disposition is more valuable.

2.3. Quantity

The final dimension is the *quantity* of agents an individual or organisation can deploy. This captures the power derived from scale, which is distinct from the qualities of any single agent. The ability to deploy and coordinate large "swarms" of agents can enable users to tackle problems of greater size and complexity through parallelised task execution (Jimenez-Romero et al., 2025; Mamie and Rao, 2025). For example, a team of agents could run millions of parallel simulations for drug discovery (Song et al., 2025), a task that is fundamentally infeasible in both scale and speed for a single agent. Such "quantity gaps" could therefore allow some actors to derive far greater benefits from agentic technology.

2.4. Compounding Effects and User-Dependent Value

Crucially, the impact of these dimensions is not independent; rather, they may compound. Ac-

cess to a large *quantity* of high-*quality* agents provides a profound advantage over an individual with access to only a single, less capable agent, creating synergistic effects that create a wider gap between the two parties. Furthermore, the ultimate utility derived from agents is not fixed but is also contingent on the user. Even with equal access to identical agents, significant performance gaps can emerge based on a user's ability to operate them effectively, which can be constrained by hardware, digital literacy, and other resource barriers (Weidinger et al., 2021; Bommasani et al., 2022). As a result, some users may elicit greater performance from an agent than others who have access to the very same one, meaning societal inequalities may still emerge even if perfect agentic equality were secured.

3. Implications of Agentic Inequality

The dimensions of agentic inequality are not abstract; they have profound societal consequences. Moreover, outcomes in different domains are likely to be shaped by a fundamental tension: whether agents primarily serve to concentrate power in the hands of a few well-resourced actors or to empower individuals widely. How this tension resolves will depend, in turn, on how disparities in agent *availability*, *quality*, and *quantity* manifest across the economic and socio-political landscape – and on the policies put in place to manage these effects.

3.1. Economic Impacts

3.1.1 Labour Market Effects

The impact of agents on labour is deeply contested, posing a tension between capital-labour substitution and intra-firm levelling (Brynjolfsson, Li and Raymond, 2025). A primary economic risk is that "agentic capital" – AI systems capable of autonomously managing complex business processes – will accelerate the shift of national income from labour to capital. This dynamic is characteristic of transformative technologies, which historically have tended to be skill-biased and capital-augmenting (Brynjolfsson and McAfee, 2014). Yet, the existing evidence on the labour market impact of generative AI is complex. While broad, economy-wide analyses show relative stability (Gimbel et al., 2025), disaggregated studies of exposed industries have found a decline in hiring for junior roles, alongside stable or growing employment for more senior workers (Brynjolfsson, Chandar and Chen, 2025; Lichtinger and Hosseini Maasoum, 2025). Separately, some firm-level studies have documented a "levelling-up" effect, where the technology boosts the productivity of less-experienced staff (Peng et al., 2023; Noy and Zhang, 2023; Cui et al., 2024; Brynjolfsson, Li and Raymond, 2025). The central uncertainty is whether these patterns will hold as the technology shifts from assistive tools to autonomous agents. For example, the observed levelling-up is a feature of human-tool augmentation; it is unclear if this positive effect can survive a transition to agentic capital designed to execute and manage entire workflows independently.

3.1.2 Industrial Organisation and Market Structures

In industry, the core tension is between market concentration and more open competition. Disparities in agent deployment risk accelerating the rise of "superstar firms" through two mechanisms (Agrawal et al., 2022; Kapoor et al., 2025). First, dominant firms can leverage proprietary data to create more capable agents, which improve services, attract more users, and generate yet more data – a powerful feedback loop that smaller rivals may find difficult to replicate. Second, large firms may be able to deploy a greater *quantity* of agents to automate internal processes and accelerate innovation at a scale that is hard to attain for smaller competitors. However, the declining cost and increasing availability of powerful agentic platforms could also act as an equalising force by lowering barriers to entry, enabling startups to orchestrate sophisticated operations that previously required significant human capital.

3.1.3 Consumer Welfare and Negotiations

The widespread adoption of AI agents is set to reshape consumer markets, introducing a new

and complex dynamic between individual empowerment and the power of corporate firms. A primary risk is that consumers will be outmatched by sophisticated corporate agents designed to maximise profit through manipulative "dark patterns" (Kolt, 2025; Zuboff, 2019). This power imbalance is particularly stark in negotiations, as studies show that a sophisticated corporate agent can consistently exploit a less capable consumer agent to secure better deals (Chen, 2025; Zhu et al., 2025). On the other hand, the universal availability of powerful consumer-side agents could act as a significant check on the power of market actors by automating comparison shopping, negotiating prices, and identifying deceptive practices on a massive scale (Van Loo, 2019). This points to a deeper transformation from human psychology to high-speed, direct agent-to-agent competition, where outcomes could be determined by relative strategic capabilities.

3.2. Social and Political Impacts

3.2.1 Access to Essential Services

Agents could remake the citizen-state relationship, creating the possibility of a stratified system or alternatively wider and better quality access. A significant risk is that affluent individuals could use premium agents capable of autonomously navigating complex bureaucratic workflows and optimising applications to secure better outcomes in healthcare or legal systems (Bovens and Zouridis, 2002). This could be exacerbated if public services become optimised for such agentic interaction, marginalising citizens without access to these capabilities. Conversely, "public-good" agents could democratise access by automating form-filling and proactively pursuing entitlements on behalf of citizens, reducing the administrative hurdles that disproportionately harm lower income groups.

3.2.2 Political Discourse and Participation

In the political sphere, the tension is between amplifying elite influence and empowering grassroots movements. Disparities in agent *quantity* and *quality* could allow well-resourced actors to deploy sophisticated agent "swarms" to execute coordinated, multi-step influence campaigns, such as overwhelming public consultations or generating personalised propaganda at scale (Gabriel et al., 2024; Persily and Tucker, 2020). However, widespread access to agents that can autonomously draft policy proposals, manage activist campaigns, and engage with political processes could also empower citizens. This shift from human persuasion to machine-mediated influence challenges democratic legitimacy itself, as the "public will" could become increasingly shaped by the scalable delegation of political action.

3.2.3 Social Stratification

Agents may introduce market-like competition into social spheres, posing a tension between amplifying the "Matthew Effect" – where initial advantages accumulate over time (Merton, 1968; DiPrete and Eirich, 2006) – and enhancing social mobility. Unlike tools that merely augment human effort, agents could enable the scalable delegation of complex social strategies – from autonomously managing a professional's networking to optimising educational opportunities for one's children. This could create a foundational divide between those able to leverage such autonomous delegates and those who cannot, with the continuous accrual of small advantages systematically widening gaps in social capital. It is also plausible, however, that agents could enhance social mobility by providing access to automated coaching and opportunity-sourcing that was previously unaffordable, transforming the basis of social stratification from one based on inherited, tacit knowledge to one based on explicit, agentic capability.

4. Forces Shaping Agentic Inequality

The impacts of AI agents are not inevitable outcomes of technology alone; they are likely to be shaped by a confluence of powerful underlying forces and the decisions we make. The critical question is whether these forces will steer

development in a direction that deepens existing divides, or whether they can be guided to ensure AI agents are integrated into society in a broadly beneficial and equitable way. While many of these drivers apply to all forms of AI, they are amplified for agents due to their unique potential for deep societal and economic integration. Rooted in both technical realities and the broader socio-political landscape, these forces and decision points are likely to bear decidedly on the future of agentic inequality.

4.1. Supply-Side and Ecosystem Drivers

4.1.1 Compute Costs and Capital Barriers

The political economy of AI is fundamentally shaped by the economics of computation, creating a tension between centralised capital and democratised application. Two cost structures are critical. The first is the immense capital expenditure required to train a foundation model, which concentrates this capability within a few large technology firms (Sastry et al., 2024; Besiroglu et al., 2024). This directly influences the availability of the most powerful agents and their underlying core intelligence. Second is the ongoing operational cost of inference (using the model). While many orders of magnitude lower, inference costs can scale with task complexity. This creates a "pay-to-perform" dynamic where premium tiers frequently offer improved capabilities such as superior agentic reasoning or faster responses. A countervailing force is the potential for diminishing marginal returns on performance for many tasks, which, combined with falling inference costs, could enable widely accessible "good enough" agents to democratise baseline autonomous capabilities, even as a *quality* gap with frontier models persists.

4.1.2 Agent Architecture and Platform Governance

The governance of foundation models – proprietary or open-weight – creates distinct pathways for diffusion and control. Proprietary models, governed by a central entity and distributed to the public through user-friendly chat interfaces and to developers via APIs, lower the user adoption barrier by abstracting away the immense costs of infrastructure and maintenance. However, this convenience establishes platform dependency, which can subtly create *quality* gaps in an agent's ability to reliably execute complex instructions with the help of user data and past experience. Conversely, open-weight models lower the innovation barrier for well-resourced actors by providing direct access to the model's parameters. This does not eliminate barriers but rather shifts them to the significant computational and technical expertise required for effective deployment, creating a divide based on implementation capacity. Neither model is inherently more equitable; they trade off different forms of access barriers and *quality* gaps (cf. Tiwana, 2014).

4.1.3 Integration and Deployment

The value of an agent is contingent on its integration with the broader digital ecosystem, making control over this "agent infrastructure" a critical source of power. An agent's tool use capabilities – a cornerstone of its ability to act in the world – is mediated by its access to APIs, proprietary databases, and software environments. Actors who control this infrastructure – such as cloud providers, financial institutions, and major software platforms – are positioned as gatekeepers at strategic chokepoints. By setting the rules of engagement for their systems (e.g., through API access policies, data sharing rules, or platform regulations), these actors can enable or disable the benefits of AI agents for different users or developers, directly affecting agents' *quality* through tool use (Chan et al., 2025; Rothschild et al., 2025). While this creates risks of new dependencies, it also offers an opportunity: developing open standards and promoting interoperability as a core technical principle of this infrastructure could prevent gatekeeping and ensure that a wider range of agents can function effectively within the broader digital ecosystem.

4.2. Socio-Political and Institutional Forces

4.2.1 Economic Incentives and Market Dynamics

Prevailing market incentives have the potential to simultaneously drive diffusion and stratification. The commercial imperative for market penetration will likely ensure broad, low-cost access to basic agentic systems for a large user base. However, standard business strategies based on product differentiation and price discrimination may also lead to significant disparities in agent *qualities* and *quantity*. Premium tiers would then offer superior autonomous capabilities, such as more sophisticated planning or the ability to use a wider range of software tools. Pricing models may also render the deployment of large agent "swarms" – essential for scalable delegation – economically infeasible for individuals and small organisations, creating a significant competitive disadvantage. Policies such as robust antitrust enforcement and the promotion of open standards may become important economic levers to disrupt monopolistic incentives and ensure benefits are broadly shared (Mazzucato, 2017; Tiwana, 2014).

4.2.2 Digital Literacy and Human-Agent Interaction

The effective use of AI agents interacts directly with existing distributions of human capital. Initially, harnessing an agent's full autonomous potential may depend on a user's skill in specifying complex, multi-step goals, favouring those with higher digital literacy. The long-term potential, however, is for advanced agents to reduce the need for specialised technical capital, allowing users to delegate complex tasks through intuitive natural language (Gabriel et al., 2024). The relatively high usage of LLMs in certain middle-income countries, such Nigeria and India,[3] suggests that intuitive, language-based interfaces can indeed broaden technological adoption in unexpected ways.

4.2.3 Geopolitics and Jurisdictional Fragmentation

At the global level, a tension exists between competitive techno-nationalism and the need for international cooperation on shared risks. National interests are driving protectionist policies, such as export controls on semiconductors, that limit global access to the foundational hardware for AI (Leicht, 2025; Larsen, 2022). This strategy, which forces manufacturers to create less powerful, export-grade chips for certain markets, sets a clear precedent for future differentiation. As agentic systems mature, this logic could extend from hardware to software, leading to mercantilist strategies where firms offer degraded "export versions" of their AI agents. Such models could come with restricted planning capabilities or prohibitions on certain autonomous actions, deliberately creating international asymmetries in agentic power. This is compounded by jurisdictional fragmentation, as diverging regulations on issues like data privacy create complex compliance burdens that favour large, multinational firms over their smaller competitors (Mueller, 2021). Countervailing these nationalist pressures, however, is a growing international consensus on the need for cooperation around AI safety and responsible deployment (Concordia AI, 2024; The French Center for AI Safety (CeSIA) et al., 2025), which could create a foundation for future efforts to ensure more equitable global access to agentic power.

4.2.4 Governance, Regulation, and Societal Adaptation

The efficacy of any response to agentic inequality is constrained by the "pacing problem": the gap between the rapidly accelerating pace of technological change and the incremental pace of institutional adaptation (Marchant, 2011). This inertia can create a window where inequalities

[3] Web-analytics data indicate outsized uptake in several MICs. Similarweb reports chatgpt.com among the top websites in India (ranked #4 nationally, August 2025) and lists India as one of the largest traffic sources to the site; Semrush's Top Websites in Nigeria ranks chatgpt.com among the country's most-visited sites with 36.35 million visits in August 2025 (accessed 4 Oct 2025).

in the ability to deploy autonomous systems can become embedded by default, a risk amplified by the potential for regulatory capture. The opportunity, however, lies in anticipatory governance for this specific technological transition, designing frameworks to manage the societal deployment of systems capable of autonomous action in complex social environments (Floridi et al., 2018).

5. Governing Agentic Inequality: Complexity and Future Directions

5.1. The Governance Challenge

Effectively governing agentic inequality is not simply about preventing the harms of disparity; it is also about creating the conditions for agents to serve as a powerful equalising tool. The challenge is complex, involving fundamental disagreements about what constitutes a "fair" distribution of agentic power. Indeed, the initial obstacle is a normative one: deciding which kinds of agent-driven inequalities are socially acceptable (Gabriel, 2022). For instance, societies currently tolerate significant disparities in access to expert human services like legal or financial advice. This forces a societal decision about whether disparities in agentic capability should be governed by the same norms that apply to disparities in human capability, or whether their unique power and scalability demand a new, more stringent standard of justice.

Even if a normative consensus is reached, significant practical obstacles remain. Existing legal frameworks are ill-equipped to handle certain harms arising from agentic inequality. Concepts like product liability struggle to assign accountability when a superior agent outcompetes or disadvantages another, creating a complex problem of legal attribution (Kolt, 2025; Calvo et al., 2020). Furthermore, attempts to level the playing field through direct intervention are fraught with coordination problems and face the "Collingridge dilemma", where acting early is difficult due to uncertainty, while acting late is difficult due to the technology's entrenchment (Collingridge, 1980). For example, attempts to cap the maximum power of agents could inadvertently lower the ceiling on innovation, while mandates to distribute computational resources could draw them away from large-scale projects that may offer significant societal benefits. Overly burdensome regulatory requirements for developers could lead them to withdraw services from smaller or more complex jurisdictions, leaving those populations with no agent access at all, which may be worse than access to imperfect or costly agents. Even government efforts to provide universal access by championing a single provider could lead to vendor lock-in and a lack of competition, thereby reinforcing market concentrations instead of alleviating them (Sastry et al., 2024; Stigler, 1971). The challenge, therefore, lies not just in intervening, but in designing policies that navigate these complex trade-offs effectively.

5.2. A Forward-Looking Research Agenda

The path forward lies in developing a coherent research agenda that uses the dimensions of agentic inequality as a direct framework for action. The following research programmes are structured to build the knowledge required to govern disparities in agent *availability*, *quality*, and *quantity*.

5.2.1 Empirical Foundations: Measuring the Dimensions of Agentic Inequality

Before interventions can be designed, a primary task is to develop robust metrics for tracking disparities across agent *availability*, *quality*, and *quantity* in real-world settings. The real-world technologies that may drive this inequality – such as automated negotiation or scaled task delegation – are novel and largely unmeasured. Key questions include: What are the most effective methods for empirically tracking the deployment of AI agents and their distributional impacts? How can we conduct direct technical comparisons to assess the extent to which disparities in agent *quality* allow more capable agents to persuade, manipulate, or out-negotiate weaker ones?

5.2.2 Normative Foundations: Defining Fairness Across the Dimensions

Building on empirical data, this programme addresses the fundamental challenge of deciding which disparities in agent *availability*, *quality*, and *quantity* are socially or ethically unacceptable. This question has a fresh urgency because agents' autonomous actions could directly produce outcomes with significant socioeconomic consequences, forcing a societal reckoning. Key questions include: What participatory methods (e.g., citizen assemblies) are most effective for eliciting public preferences on unacceptable capability gaps? How can these preferences be translated into concrete design principles or regulatory guardrails governing disparities in agent *quality*?

5.2.3 Technical and Infrastructural Levers for Equality

This programme focuses on building equity-enabling features directly into the agentic ecosystem to directly mitigate harmful gaps in agents' capabilities and prevent the excessive concentration of large quantities of agents. It is crucial to understand that neither open-source nor proprietary release strategies on their own guarantee equitable outcomes. Key questions include: What models for governing critical resources, such as compute, can best balance innovation with equitable access? What technical standards for interoperability might be needed to prevent platform lock-in, which entrenches disparities in agent *quality* (Chan et al., 2025; Kapoor et al., 2025)?

5.2.4 Public Service Models for Agentic AI

A more direct approach to ensuring equity could involve exploring non-market models. This research should investigate policies designed to guarantee universal access and a sufficient baseline of agent *quality* for all citizens. A key concept is "universal basic agency" – analogous to universal basic income – which suggests every citizen could be entitled to a baseline level of autonomous AI assistance. Key questions include: Should a government provide a "public option" agent to guarantee access, or would subsidies for private service be more effective? What level of agent *quality* would be needed to provide a meaningful baseline of empowerment? How could a public agent (or public infrastructure for private agents) be designed to be trustworthy and aligned with public, rather than commercial, interests? A key avenue for this is to explore instilling such agents with a legally enforceable fiduciary responsibility – a duty to act exclusively in their user's best interest (Kapoor et al., 2025).

5.2.5 Regulatory Frameworks for Agentic Interactions

Finally, this programme must explore regulatory frameworks designed to manage the social consequences of competitive interactions between agents with vastly different *qualities* and *quantities*. Existing legal doctrines are strained by these new asymmetries, and without clear rules, the advantages conferred by superior agents could become entrenched. Key questions include: How can legal frameworks be adapted to provide recourse for individuals harmed by disparities in agent *quality* (Calvo et al., 2020)? What "agile" or outcome-based regulatory models can keep pace with change while protecting individuals from exploitation by actors deploying vast *quantities* of agents (Marchant, 2011)?

5.3. Conclusion

The analysis presented in this paper suggests that agentic inequality represents a novel and significant societal challenge, driven by a confluence of powerful technical, economic, and political forces. The governance of this issue is fraught with complexity, and simplistic solutions are unlikely to succeed. However, the very drivers that create these risks – from the design of agent architectures to the structure of market incentives – are also the levers that can be used to steer development toward more equitable outcomes. A proactive and sustained research agenda, focused explicitly on the questions of equity, access, and power outlined above, is therefore essential. By prioritising these research programmes, the

scientific, policy, and technology communities can build the knowledge and tools required to steer the development of autonomous AI agents towards a more just, inclusive, and beneficial future for all.

Acknowledgements

We thank Alan Chan, Noam Kolt, Seth Lazar, and Sam Manning for helpful feedback. This work was supported by the Future Impact Group and the Cooperative AI Foundation.

References

Acemoglu, D. and Restrepo, P. (2020), 'Robots and Jobs: Evidence from US Labor Markets', *Journal of Political Economy* **128**(6), 2188–2244. **URL:** *https://www.journals.uchicago.edu/doi/10.1086/705716*

Agrawal, A., Gans, J. and Goldfarb, A. (2022), *Prediction Machines, Updated and Expanded: The Simple Economics of Artificial Intelligence*, Harvard Business Review Press.

Bergman, A. S., Shelby, R. and Gabriel, I. (2024), Chapter 15: Access and Opportunity, *in* 'The Ethics of Advanced AI Assistants', arXiv, pp. 146–156.

Besiroglu, T., Bergerson, S. A., Michael, A., Heim, L., Luo, X. and Thompson, N. (2024), 'The Compute Divide in Machine Learning: A Threat to Academic Contribution and Scrutiny?'. **URL:** *http://arxiv.org/abs/2401.02452*

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Arx, S. v., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L., Goel, K., Goodman, N., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D. E., Hong, J., Hsu, K., Huang, J., Icard, T., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Koh, P. W., Krass, M., Krishna, R., Kuditipudi, R., Kumar, A., Ladhak, F., Lee, M., Lee, T., Leskovec, J., Levent, I., Li, X. L., Li, X., Ma, T., Malik, A., Manning, C. D., Mirchandani, S., Mitchell, E., Munyikwa, Z., Nair, S., Narayan, A., Narayanan, D., Newman, B., Nie, A., Niebles, J. C., Nilforoshan, H., Nyarko, J., Ogut, G., Orr, L., Papadimitriou, I., Park, J. S., Piech, C., Portelance, E., Potts, C., Raghunathan, A., Reich, R., Ren, H., Rong, F., Roohani, Y., Ruiz, C., Ryan, J., Ré, C., Sadigh, D., Sagawa, S., Santhanam, K., Shih, A., Srinivasan, K., Tamkin, A., Taori, R., Thomas, A. W., Tramèr, F., Wang, R. E., Wang, W., Wu, B., Wu, J., Wu, Y., Xie, S. M., Yasunaga, M., You, J., Zaharia, M., Zhang, M., Zhang, T., Zhang, X., Zhang, Y., Zheng, L., Zhou, K. and Liang, P. (2022), 'On the Opportunities and Risks of Foundation Models'. **URL:** *http://arxiv.org/abs/2108.07258*

Bovens, M. and Zouridis, S. (2002), 'From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control', *Public Administration Review* **62**(2), 174–184. **URL:** *https://onlinelibrary.wiley.com/doi/10.1111/0033-3352.00168*

Brynjolfsson, E., Chandar, B. and Chen, R. (2025), Canaries in the Coal Mine? Six Facts About the Recent Employment Effects of Artificial Intelligence, Technical report, Stanford Digital Economy Lab, Stanford Institute for Human-Centered Artificial Intelligence. **URL:** *https://digitaleconomy.stanford.edu/wp-content/uploads/2025/08/Canaries_BrynjolfssonChandarChen.pdf*

Brynjolfsson, E., Li, D. and Raymond, L. (2025), 'Generative AI at Work', *The Quarterly Journal of Economics* **140**(2), 889–942. **URL:** *https://academic.oup.com/qje/article/140/2/889/7990658*

Brynjolfsson, E. and McAfee, A. (2014), *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, W. W. Norton & Company, New York, NY, US.

Calvo, R. A., Peters, D., Vold, K. and Ryan, R. M. (2020), Supporting Human Autonomy in AI Systems: A Framework for Ethical Enquiry, *in* C. Burr and L. Floridi, eds, 'Ethics of Digital Well-Being', Vol. 140, Springer International Publishing, Cham, pp. 31–54. **URL:** *http://link.springer.com/10.1007/978-3-030-50585-1_2*

Chan, A., Wei, K., Huang, S., Rajkumar, N., Perrier, E., Lazar, S., Hadfield, G. K. and Anderljung, M.

(2025), 'Infrastructure for AI Agents'.
URL: *http://arxiv.org/abs/2501.10114*

Chen, C. (2025), 'When AIs Bargain, a Less Advanced Agent Could Cost You'.
URL: *https://www.technologyreview.com/2025/06/17/1118910/ai-price-negotiation/*

Collingridge, D. (1980), *The Social Control of Technology*, Frances Pinter.

Concordia AI (2024), 'The State of China-Western Track 1.5 and 2 Dialogues on AI'.
URL: *https://aisafetychina.substack.com/p/the-state-of-china-western-track*

Cui, Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S. and Salz, T. (2024), 'The Effects of Generative AI on High Skilled Work: Evidence from Three Field Experiments with Software Developers'.
URL:
https://www.ssrn.com/abstract=4945566

DiPrete, T. A. and Eirich, G. M. (2006), 'Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments', *Annual Review of Sociology* **32**, 271–297.
URL:
https://www.annualreviews.org/content/journals/10.1146/annurev.soc.32.061604.123127

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. and Vayena, E. (2018), 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations', *Minds and Machines* **28**(4), 689–707.
URL: *http://link.springer.com/10.1007/s11023-018-9482-5*

Gabriel, I. (2022), 'Toward a Theory of Justice for Artificial Intelligence'.
URL: *http://arxiv.org/abs/2110.14419*

Gabriel, I., Manzini, A., Keeling, G., Hendricks, L. A., Rieser, V., Iqbal, H., Tomašev, N., Ktena, I., Kenton, Z., Rodriguez, M., El-Sayed, S., Brown, S., Akbulut, C., Trask, A., Hughes, E., Bergman, A. S., Shelby, R., Marchal, N., Griffin, C., Mateos-Garcia, J., Weidinger, L., Street, W., Lange, B., Ingerman, A., Lentz, A., Enger, R., Barakat, A., Krakovna, V., Siy, J. O., Kurth-Nelson, Z., McCroskery, A., Bolina, V., Law, H., Shanahan, M., Alberts, L., Balle, B., Haas, S. d., Ibitoye, Y., Dafoe, A., Goldberg, B., Krier, S., Reese, A., Witherspoon, S., Hawkins, W., Rauh, M., Wallace, D., Franklin, M., Goldstein, J. A., Lehman, J., Klenk, M., Vallor, S., Biles, C., Morris, M. R., King, H., Arcas, B. A. y., Isaac, W. and Manyika, J. (2024), 'The Ethics of Advanced AI Assistants'.
URL: *http://arxiv.org/abs/2404.16244*

Gimbel, M., Kinder, M., Kendall, J. and Lee, M. (2025), Evaluating the Impact of AI on the Labor Market: Current State of Affairs, Technical report, The Budget Lab at Yale University.
URL:
https://budgetlab.yale.edu/research/evaluating-impact-ai-labor-market-current-state-affairs

Jasanoff, S. (2004), *States of Knowledge: The Co-Production of Science and the Social Order*, Routledge, London.

Jimenez-Romero, C., Yegenoglu, A. and Blum, C. (2025), 'Multi-agent Systems Powered by Large Language Models: Applications in Swarm Intelligence', *Frontiers in Artificial Intelligence* **8**, 1593017.
URL: *https://www.frontiersin.org/articles/10.3389/frai.2025.1593017/full*

Kapoor, S., Kolt, N. and Lazar, S. (2025), 'Build Agent Advocates, Not Platform Agents'.
URL: *http://arxiv.org/abs/2505.04345*

Kasirzadeh, A. and Gabriel, I. (2025), 'Characterizing AI Agents for Alignment and Governance'.
URL: *http://arxiv.org/abs/2504.21848*

Kolt, N. (2025), 'Governing AI Agents'.
URL: *http://arxiv.org/abs/2501.07913*

Larsen, B. C. (2022), 'Geopolitics of AI and the Rise of Digital Sovereignty', *Brookings Institution* .
URL:
https://coilink.org/20.500.12592/swc5mh

Lazar, S. (2025), 'Governing the Algorithmic City', *Philosophy & Public Affairs* **53**(2), 102–168.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/papa.12279*

Leicht, A. (2025), 'Mercantilist AI Policy: Reconstructing the Doctrine Behind US AI Foreign Policy'.
URL: *https://writing.antonleicht.me/p/mercantilist-ai-policy*

Lichtinger, G. and Hosseini Maasoum, S. M. (2025), 'Generative AI as Seniority-Biased Technological Change: Evidence from U.S. Résumé and Job Posting Data'.
URL: *https://papers.ssrn.com/abstract=5425555*

Liu, C., Wang, C., Cao, J., Ge, J., Wang, K., Zhang, L., Cheng, M.-M., Zhao, P., Li, T., Jia, X., Li, X., Li, X., Liu, Y., Feng, Y., Huang, Y., Xu, Y., Sun, Y., Zhou, Z. and Xu, Z. (2025), 'A Vision for Auto Research with LLM Agents'.
URL: *http://arxiv.org/abs/2504.18765*

Mamie, N. and Rao, S. X. (2025), 'The Society of HiveMind: Multi-Agent Optimization of Foundation Model Swarms to Unlock the Potential of Collective Intelligence'.
URL: *http://arxiv.org/abs/2503.05473*

Marchant, G. E. (2011), The Growing Gap Between Emerging Technologies and the Law, *in* G. E. Marchant, B. R. Allenby and J. R. Herkert, eds, 'The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight', Vol. 7, Springer Netherlands, Dordrecht, pp. 19–33.
URL: *https://link.springer.com/10.1007/978-94-007-1356-7_2*

Mazzucato, M. (2017), *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*, Penguin Books, London.

Merton, R. K. (1968), 'The Matthew Effect in Science', *Science* **159**(3810), 56–63.
URL: *https://www.science.org/doi/10.1126/science.159.3810.56*

Mueller, B. (2021), How Much Will the Artificial Intelligence Act Cost Europe?, Technical report, Centre for Data Innovation.
URL: *https://www2.datainnovation.org/2021-aia-costs.pdf*

Mumford, L. (1970), *The Myth of the Machine. 1: Technics and Human Development*, Harcourt, Brace, Jovanovich, New York, NY.

Noy, S. and Zhang, W. (2023), 'Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence', *Science* **381**(6654), 187–192.
URL: *https://www.science.org/doi/10.1126/science.adh2586*

Peng, S., Kalliamvakou, E., Cihon, P. and Demirer, M. (2023), 'The Impact of AI on Developer Productivity: Evidence from GitHub Copilot'.
URL: *https://arxiv.org/abs/2302.06590*

Persily, N. and Tucker, J. A. (2020), *Social Media and Democracy: The State of the Field, Prospects for Reform*, Cambridge University Press.
URL: *https://books.google.com/books?hl=en&lr=&id=NEjzDwAAQBAJ&oi=fnd&pg=PR11&dq=.+Social+Media+and+Democracy:+The+State+of+the+Field,+Prospects+for+Reform.+Cambridge+University+Press.&ots=5k2QmsFi8F&sig=ZmxLt4z0bNXzvPY1IZefwiUInsU*

Plantin, J.-C., Lagoze, C., Edwards, P. N. and Sandvig, C. (2018), 'Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook', *New Media & Society* **20**(1), 293–310.
URL: *https://journals.sagepub.com/doi/10.1177/1461444816661553*

Prunkl, C. (2024), 'Human Autonomy at Risk? An Analysis of the Challenges from AI', *Minds and Machines* **34**(26).
URL: *https://link.springer.com/10.1007/s11023-024-09665-1*

Rawls, J. (2001), *Justice as Fairness: A Restatement*, Harvard University Press.

Rothschild, D. M., Mobius, M., Hofman, J. M., Dillon, E. W., Goldstein, D. G., Immorlica, N., Jaffe, S., Lucier, B., Slivkins, A. and Vogel, M. (2025), 'The Agentic Economy'.
URL: *http://arxiv.org/abs/2505.15799*

Sastry, G., Heim, L., Belfield, H., Anderljung, M., Brundage, M., Hazell, J., O'Keefe, C., Hadfield, G. K., Ngo, R., Pilz, K., Gor, G., Bluemke, E., Shoker, S., Egan, J., Trager, R. F., Avin, S., Weller, A., Bengio, Y. and Coyle, D. (2024), 'Computing Power and the Governance of Artificial Intelligence'.
URL: *http://arxiv.org/abs/2402.08797*

Sen, A. (1999), *Development as Freedom*, Oxford University Press.

Song, K., Trotter, A. and Chen, J. Y. (2025), 'LLM Agent Swarm for Hypothesis-Driven Drug Discovery'.
URL: *http://arxiv.org/abs/2504.17967*

Stigler, G. J. (1971), 'The theory of economic regulation', *The Bell Journal of Economics and Management Science* **2**(1), 3–21.
URL: *http://www.jstor.org/stable/3003160*

The French Center for AI Safety (CeSIA), The Future Society and UC Berkeley's Center for Human-Compatible Artificial Intelligence (2025), 'Global Call for AI Red Lines'.
URL: *https://red-lines.ai/*

Tiwana, A. (2014), *Platform Ecosystems: Aligning Architecture, Governance, and Strategy*, Elsevier.
URL: *https://linkinghub.elsevier.com/retrieve/pii/C20120066252*

Van Loo, R. (2019), 'Digital Market Perfection', *Michigan Law Review* **117**(5), 815–884.
URL: *https://repository.law.umich.edu/mlr/vol117/iss5/2/*

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z. and Wen, J.-R. (2024), 'A Survey on Large Language Model Based Autonomous Agents', *Frontiers of Computer Science* **18**(6), 186345.
URL: *http://arxiv.org/abs/2308.11432*

Warschauer, M. (2003), *Technology and Social Inclusion: Rethinking the Digital Divide*, MIT Press.
URL: *https://books.google.com/books?hl=en&lr=&id=nU4zz1O88mAC&oi=fnd&pg=PP11&dq=Warschauer,+M.+(2004).+Technology+and+social+inclusion:+Rethinking+the+digital+divide.+MIT+press.&ots=rXWxXXo34q&sig=zcrhPMUB7JD9KvXzOY1qqK1jyx4*

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., Isaac, W., Legassick, S., Irving, G. and Gabriel, I. (2021), 'Ethical and Social Risks of Harm from Language Models'.
URL: *https://arxiv.org/abs/2112.04359*

Zhu, S., Sun, J., Nian, Y., South, T., Pentland, A. and Pei, J. (2025), 'The Automated but Risky Game: Modeling Agent-to-Agent Negotiations and Transactions in Consumer Markets'.
URL: *https://arxiv.org/abs/2506.00073*

Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Profile Books.