TOKEN TAXES: MITIGATING AGI'S ECONOMIC RISKS

Anonymous authors
Paper under double-blind review

ABSTRACT

The development of AGI threatens to erode government tax bases, lower living standards, and disempower citizens—risks that make the 40-year stagnation of wages during the first industrial revolution look mild in comparison. While AI safety research has focused primarily on capability risks, comparatively little work has studied how to mitigate the economic risks of AGI. In this paper, we argue that the economic risks posed by a post-AGI world can be effectively mitigated by token taxes: usage-based surcharges on model inference applied at the point of sale. We situate token taxes within previous proposals for robot taxes and identify two key advantages: they are enforceable through existing compute governance infrastructure, and they capture value where AI is used rather than where models are hosted. We then present a research roadmap. For enforcement, we outline a staged audit pipeline — black-box token verification, norm-based tax rates, and white-box audits. For impact, we highlight the need for agent-based modeling of token taxes' economic effects. Finally, we discuss alternative approaches including FLOP taxes, and how to prevent AI superpowers vetoing such measures.

1 INTRODUCTION

Throughout history, technological revolutions have displaced workers and ushered in periods of socioeconomic destabilization Allen (2009a). During the first industrial revolution, a period now referred to as *Engel's Pause* (1790-1830) saw real wages stagnate as GDP per worker increased rapidly Allen (2009b); Frey (2019; 2025). Engel's Pause triggered a downturn in living standards, as children worked in factories to support their parents and diseases such as cholera saw a substantial uptick in cases Allen (2009b).

Early evidence of AI's impact on the labour market is already emerging in studies which show unemployment in AI-exposed early career roles rising by 16% Brynjolfsson et al. (2025). The development of AGI threatens a much greater level of destabilization Korinek & Stiglitz (2018). Whereas previous technologies automated narrow tasks, AGI will develop capabilities across all economically relevant tasks, allowing AI to replace humans in every industry Kulveit et al. (2025). This relative decline in the importance of human labour could ultimately engender a gradual disempowerment of human beings, as governments no longer rely on humans for tax revenues Kulveit et al. (2025).

This paper argues in favour of prioritizing the *token tax* as a first-line of defence against the economic risks of a post-AGI world. Token taxes are unique in two key respects. First, they are (1) enforceable: existing compute governance infrastructure enables token taxes to be audited and enforced using black-box or white-box methods Sastry et al. (2024). Second, they are (2) usage-based rather than firm-based. This enables taxation at the point of sale, mitigating inequality by capturing value where the model is used rather than where it is hosted.

We discuss the economic risks of a post-AGI world and situate the token tax within previous proposals for robot taxes. We then present a research roadmap recommending a staged auditing pipeline - black-box token verification, norm-based tax rates and white-box auditing. For impact, we also recommend agent-based modeling of token tax policies. Finally, we discuss alternative viewpoints including the FLOP (floating-point-operation) tax and how to prevent AI superpowers from vetoing token taxes.

2 BACKGROUND: THE ECONOMIC RISKS OF A POST-AGI WORLD

Thus far AI safety has focused heavily on the risks of increasing AI capabilities Bostrom (2014). Risks pertaining to the future of work have been less prominent but may be even more urgent than capability risks Acemoglu & Restrepo (2019). We discuss three key economic risks of a post-AGI world, motivating our argument in favour of promoting token taxes as an AGI governance instrument.

Government fiscal crises The first economic risk posed by a post-AGI world is fiscal crisis resulting from lost government revenues. Early studies have shown that high exposure to AI in companies increases unemployment risks while reducing labour demand Hampole et al. (2025); Frank et al. (2025). Recent work has also found that as automation increases, government fiscal revenues decrease, because taxes on labour are the highest source of government revenue Casas & Torres (2024). High unemployment in a post-AGI world would therefore reduce revenue and increase states' fiscal costs, eroding the two main sources of public finance Korinek & Lockwood (2025).

Gradual disempowerment of citizens If AGI displaces humans by out-competing them in nearly all economic and societal functions, humans will gradually lose control over the levers of state power. Currently, governments and economic systems respond to human action since humans are the main drivers of economic and social progress Giddens (1984). If AI begins to displace humans as the main driver of economic growth, governments will become far less responsive to citizens' needs, since they will no longer need their support Kulveit et al. (2025).

A similar phenomenon occurs in so-called *rentier states* Neal (2019). States such as Venezuela, Saudi Arabia and Oman have abundant resources and earn most of their income from oil rents as opposed to their citizens' labour. Yet most of their citizens live in abject poverty. This is referred to as the *resource curse* Acemoglu et al. (2002; 2005). A core reason for the resource curse is the fact that states no longer have an incentive to look after their own people. In the context of AGI-driven unemployment, states will make most of their revenue from taxing AI capital and not workers and hence lose their incentive to tend to citizens' needs.

Global inequality A post-AGI world will exacerbate global inequality. Absent new forms of taxation, AGI's economic benefits are likely to be unevenly distributed, worsening existing divides between rich and poor countries. This is obvious when analysing the global distribution of compute. The advanced compute chips necessary to train and run frontier AI models are highly concentrated in a handful of economically developed nations. These nations, referred to as the *Compute North* in recent work, have priority access to and jurisdictional authority over advanced GPUs, while countries in the *Compute South* must rely on renting these chips from the Compute North Lehdonvirta et al. (2024). These countries will therefore benefit greatly from their ability to determine tax rates on these resources, while countries outside of the Compute North will be left behind. The global distribution of frontier model companies follows a similar pattern, with the vast majority located in the US and China.

3 ROBOT TAXES

Previous work has argued in favor of *robot taxes* as a solution to the increasing economic inefficiencies created by technology and automation Abbott & Bogenschneider (2018); Mazur (2018); Cabrales et al. (2020). Proponents argue that labour tax rates are much higher than capital tax rates because governments seek to incentivize innovation. Yet this system ceases to be efficient when capital **becomes** the labour. In the case of AI, foundation models are treated as capital even though they carry out tasks usually performed by human labour Acemoglu et al. (2020). A robot tax seeks to solve this problem by restoring *tax neutrality* to the system, removing incentives for companies to choose AI workers over human workers Abbott & Bogenschneider (2018); Mirrlees et al. (2011).

Prior suggestions for robot taxes include raising the corporate tax rate, levying an ”automation tax”, and disallowing corporate tax deductions whenever a company invests in automation Abbott & Bogenschneider (2018). As AGI displaces human labour, the question of optimal robot taxation becomes central. We make the argument for why token taxes are the most promising AGI governance instrument for mitigating the economic risks of a post-AGI world.

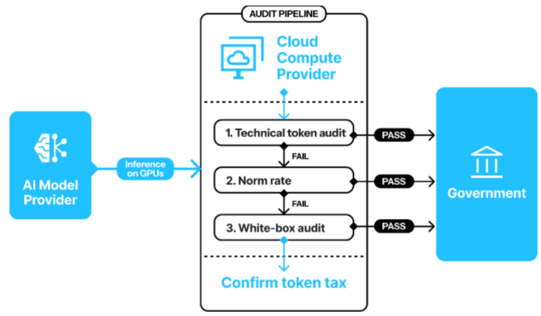

Figure 1: **Token tax collection procedure.** The cloud compute provider acts as an intermediary between the AI model provider and the government. It runs inference on behalf of the AI model provider, collects billed tokens, and determines tax liability through a staged audit pipeline. The provider first attempts (1) a black-box technical token audit. If this fails, it applies (2) a norm-based tax rate derived from average token usage. If the norm-based tax approach fails, it escalates to (3) a white-box audit. Once an audit pathway succeeds, the compute provider confirms the total token tax and reports it to the government.

4 The Token Tax

Definition: *A token tax is a usage-based surcharge applied to model tokens at the point of sale.*

Implementation: Drawing on existing usage-based billing procedures for language models, the token tax would be levied as a percentage markup on the provider's billed token cost paid directly to the government. For instance, if the token tax is 10% and the cost-per-token for a given model is \$1, the AI company will pay \$0.10 in token tax. Similar to arguments made in favour of robot taxes, token taxes would enable governments to align the token tax rate with income tax, restoring tax neutrality and efficiency to the post-AGI economic system Mazur (2018); Abbott & Bogenschneider (2018).

4.1 Advantages:

We argue that there are two main advantages of using a token tax over other forms of robot tax. First, token taxes can be enforced by governments using: (1) black-box token audits (2) norm-based taxes and (3) white-box audits. To faciliate effective enforcement, we argue that compute providers should act as intermediaries between the government and AI model providers to apply a staged audit pipeline of (1), (2), and (3) (see Figure 1). Second, we claim that token taxes mitigate against the worsening of global inequality. By collecting the token tax where tokens are utilized rather than where the model is hosted, token taxes ensure that AI consumer countries (such as those in the Compute South) also benefit from AI tax receipts.

4.2 Auditing pipeline

Currently, only model providers have full access to the generative process undertaken when a user calls an API, while auditors only have access to the outputs. This enables providers to inflate their profits by *strategically under-reporting* the number of tokens used Velasco et al. (2025); Wang et al. (2025); Sun et al. (2025). To properly enforce token taxes, we argue that compute providers should be legally required to implement the following 3-stage audit pipeline:

Stage 1: Black-box token audits Compute providers serve as *record keepers*, *verifiers*, and *enforcers* and are therefore able to ensure AI companies comply with obligations imposed on them by governments Heim et al. (2024). Delegating the task of auditing token count to compute providers can therefore solve the misreporting problem since providers can act as independent verifiers of compliance. Cloud hyperscalers already collect metadata on the compute consumed and the type of workloads (inference or pre-training) submitted to their servers Heim et al. (2024). By mandating that providers also collect token-level usage data, cloud providers can carry out oversight functions on behalf of the government. This would enable tax authorities to cross-check reported token usage

against independent logs, barring AI labs from misreporting their usage to maximize profits.
Stage 2: Norm-based taxes The second audit stage draws on a common solution to tax evasion in rentier states such as Norway: *norm taxes* Morgan & Robinson (1976); Takle (2021). The idea of a norm tax is simple: if a company decides to inflate its profits by *strategically under-reporting* token counts, auditors can refer to norm-based tax rates for each model category and charge the company according to the norm rate. This would require a methodology for continuously updating the average number of tokens used in every inference run for each type of model. It would limit the amount that companies can inflate their profits by ensuring that a company must pay a flat rate based on empirical evidence. It also has the added benefit of only requiring black-box access to models.
Stage 3: White-box audits The third audit stage would require companies to share information about the generative process with third-party auditors. If companies are required to share this by law, they will no longer be able to game the number of tokens they report since auditors will have access to the information necessary to confirm the token count.

5 Alternative Views

While we argue that token taxes are a promising AGI governance instrument, we consider several alternative views to ensure balance.

A1: Token taxes will disincentivize innovation A salient counterargument to token taxes is that they will discourage innovation and incentivize leading AI firms to relocate to more favorable tax jurisdictions. Therefore, we recommend that technical governance researchers should partner with economists and *leverage Agent-Based Modeling (ABM)* to predict the impact of token taxes on markets. ABMs have already been successfully applied to make predictions in economics Axtell & Farmer (2025); Vu et al. (2022), with LLM-based ABMs simulating large scale social interactions in digital environments such as Twitter and Reddit Yang et al. (2024). We recommend that the technical governance research community partner with economists to estimate the economic impacts of token taxes on real-world markets under different hypothetical AGI growth scenarios. The technical governance community can provide unique skills in this domain by understanding the policy significance of the work while also contributing technical expertise.

A2: *A FLOP tax is preferable to a token tax* A FLOP-based tax directly levied on compute represents one of the most serious alternatives to a token tax. Compute-based thresholds are already applied as regulatory proxies for model capability and risk in AI legislation such as the EU AI Act and Biden's executive order 14110 Act (2024); Joseph R. Biden, Jr. (2023). We claim that token taxes and FLOP taxes are not mutually exclusive. An optimal policy framework might combine token taxes with FLOP taxes in a hybrid approach which leverages the enforcement advantages of FLOP taxes while preserving the consumption-based and equity-promoting features of token taxes. Indeed, we welcome proposals to combine FLOP taxes with token taxes.

A3: *AI superpowers can veto token taxes* A further objection to token taxes is geopolitical; Since AI superpowers (the USA and China) contain the vast majority of the AI supply chain, they are able to leverage their economic and diplomatic power to veto or undermine token tax regimes proposed by smaller nations seeking to implement AI-related legislation Emery-Xu et al. (2025). Contentious negotiations surrounding Digital Services Taxes (DSTs) applied in many European countries serve as a precedent. To mitigate these trade tensions and retaliation from AI superpowers, we argue that regional agreements by coalitions of the willing could be established. The EU's market size and regulatory capacity enabled successful implementation of the GDPR and the AI Act despite U.S. opposition Act (2024). A coordinated token tax by a coalition of the willing would be much harder to veto than unilateral national measures.

6 Conclusion

In this paper, we claim that token taxes can mitigate the economic risks of a post-AGI world. In contrast to existing robot taxes, we argue that token taxes offer 2 unique advantages: they are enforceable via existing compute governance infrastructure and they reduces global inequality by capturing value where AI tokens are used not where models are hosted. We encourage further research into the open technical governance research agendas we have identified.

REFERENCES

Ryan Abbott and Bret Bogenschneider. Should robots pay taxes: Tax policy in the age of automation. *Harv. L. & Pol'y Rev.*, 12:145, 2018.

Daron Acemoglu and Pascual Restrepo. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30, 2019. Develops task-based framework showing technology can reduce labor demand and that policy matters for outcomes.

Daron Acemoglu, Simon Johnson, and James A Robinson. Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *The Quarterly journal of economics*, 117(4):1231–1294, 2002.

Daron Acemoglu, Simon Johnson, and James A Robinson. Institutions as a fundamental cause of long-run growth. *Handbook of economic growth*, 1:385–472, 2005.

Daron Acemoglu, Andrea Manera, and Pascual Restrepo. Does the us tax code favor automation? Technical report, National Bureau of Economic Research, 2020.

EU Artificial Intelligence Act. The eu artificial intelligence act. *European Union*, 2024.

Robert C Allen. *The British industrial revolution in global perspective*. Cambridge University Press, 2009a.

Robert C Allen. Engels' pause: Technical change, capital accumulation, and inequality in the british industrial revolution. *Explorations in Economic History*, 46(4):418–435, 2009b.

Robert L Axtell and J Doyne Farmer. Agent-based modeling in economics and finance: Past, present, and future. *Journal of Economic Literature*, 63(1):197–287, 2025.

Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014. Foundational framing of capability / existential risks in AI safety.

Erik Brynjolfsson, Bharat Chandar, and Ruyu Chen. Canaries in the coal mine? six facts about the recent employment effects of artificial intelligence. *Digital Economy*, 2025.

Antonio Cabrales, Penélope Hernández, and Angel Sánchez. Robots, labor markets, and universal basic income. *Humanities and Social Sciences Communications*, 7(1):1–8, 2020.

Pablo Casas and José L Torres. Government size and automation. *International tax and public finance*, 31(3):780–807, 2024.

Nicholas Emery-Xu, Richard Jordan, and Robert Trager. International governance of advancing artificial intelligence. *AI & SOCIETY*, 40(4):3019–3044, 2025.

Morgan R Frank, Yong-Yeol Ahn, and Esteban Moro. Ai exposure predicts unemployment risk: A new approach to technology-driven job loss. *PNAS nexus*, 4(4):pgaf107, 2025.

Carl Benedikt Frey. *The technology trap: Capital, labor, and power in the age of automation*. Princeton University Press, 2019.

Carl Benedikt Frey. *How Progress Ends: Technology, Innovation, and the Fate of Nations*. Princeton University Press, 2025.

Anthony Giddens. *The constitution of society: Outline of the theory of structuration*. Univ of California Press, 1984.

Menaka Hampole, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller. Artificial intelligence and the labor market. Technical report, National Bureau of Economic Research, 2025.

Lennart Heim, Tim Fist, Janet Egan, Sihao Huang, Stephen Zekany, Robert Trager, Michael A Osborne, and Noa Zilberman. Governing through the cloud: The intermediary role of compute providers in ai regulation. *arXiv preprint arXiv:2403.08501*, 2024.

Joseph R. Biden, Jr. Executive order 14110: Safe, secure, and trustworthy development and use of artificial intelligence. The White House / Federal Register, Oct 30 2023. Federal Register Vol. 88, No. 210, Presidential Documents, Nov. 1, 2023, pp. 75191. Official text available from The White House and govinfo.

Anton Korinek and Lee Lockwood. Public finance in the age of ai: A primer. *NBER Chapters*, 2025.

Anton Korinek and Joseph E Stiglitz. Artificial intelligence and its implications for income distribution and unemployment. In *The economics of artificial intelligence: An agenda*, pp. 349–390. University of Chicago Press, 2018.

Jan Kulveit, Raymond Douglas, Nora Ammann, Deger Turan, David Krueger, and David Duvenaud. Gradual disempowerment: Systemic existential risks from incremental ai development. *arXiv preprint arXiv:2501.16946*, 2025.

Vili Lehdonvirta, Bóxī Wú, and Zoe Hawkins. Compute north vs. compute south: the uneven possibilities of compute-based ai governance around the globe. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1):828–838, 2024.

Orly Mazur. Taxing the robots. *Pepp. L. Rev.*, 46:277, 2018.

James Mirrlees et al. *Tax by design: The Mirrlees review*. OUP Oxford, 2011.

Jon R Morgan and Colin Robinson. The comparative effects of the uk and norwegian oil taxation systems on profitability and government revenue. *Accounting and Business Research*, 7(25):2–16, 1976.

Mark Neal. *A dictionary of business and management in the Middle East and North Africa*. Oxford University Press, 2019.

Girish Sastry, Lennart Heim, Haydn Belfield, Markus Anderljung, Miles Brundage, Julian Hazell, Cullen O'keefe, Gillian K Hadfield, Richard Ngo, Konstantin Pilz, et al. Computing power and the governance of artificial intelligence. *arXiv preprint arXiv:2402.08797*, 2024.

Guoheng Sun, Ziyao Wang, Bowei Tian, Meng Liu, Zheyu Shen, Shwai He, Yexiao He, Wanghao Ye, Yiting Wang, and Ang Li. Coin: Counting the invisible reasoning tokens in commercial opaque llm apis. *arXiv preprint arXiv:2505.13778*, 2025.

Marianne Takle. The norwegian petroleum fund: Savings for future generations? *Environmental values*, 30(2):147–167, 2021.

Ander Artola Velasco, Stratis Tsirtsis, and Manuel Gomez-Rodriguez. Auditing pay-per-token in large language models. *arXiv preprint arXiv:2510.05181*, 2025.

Tuong Manh Vu, Charlotte Buckley, Joao A Duro, and Robin C Purshouse. Exploring social theory integration in agent-based modelling using multi-objective grammatical evolution. In *ICML 2022 Workshop AI for Agent-Based Modelling*, 2022.

Ziyao Wang, Guoheng Sun, Yexiao He, Zheyu Shen, Bowei Tian, and Ang Li. Predictive auditing of hidden tokens in llm apis via reasoning length estimation. *arXiv preprint arXiv:2508.00912*, 2025.

Ziyi Yang, Zaibin Zhang, Zirui Zheng, Yuxian Jiang, Ziyue Gan, Zhiyu Wang, Zijian Ling, Jinsong Chen, Martz Ma, Bowen Dong, et al. Oasis: Open agent social interaction simulations with one million agents. *arXiv preprint arXiv:2411.11581*, 2024.