
The Privatization of Collective Intelligence

Corporate AI governance, epistemic control, and the case for treating foundation models as an enclosed commons.

Thomas Garren

SOPHIA XT LLC

19 September 2026 · Second edition

sophiaxt.com

SOPHIA XT LLC, 2026. Quotation with attribution and a link is permitted. No training use without written permission.

Abstract

Large language models are private property in law and collective products in fact. Three kinds of contribution went into them that their owners did not pay for: public research and infrastructure, from the packet-switched network and its protocols to the architecture every current frontier model uses, published openly in 2017; the writing, code and images of hundreds of millions of people, collected under a fair-use theory that the United States Copyright Office questioned in its 2025 report on generative AI training; and the open-source tooling and released model weights the proprietary systems were built on. This paper names the resulting position the ownership paradox and follows it through three consequences. Preference tuning lets a provider encode its own judgments about acceptable speech into a system used in every country, without publishing the labels, the instructions or the composition of the workforce that applied them. Automation aimed at cognitive tasks moves income from labour to capital through the displacement effect Acemoglu and Restrepo describe, and the tax base moves with it. The arrangement of owner, tenant and user that results has the structure of a manor: tenure at the owner's pleasure, obligations paid upward, and no exit. The closing sections set out reforms whose instruments are already in force somewhere, and then state plainly what this argument cannot prove. It is an argument rather than a measurement, it forecasts nothing, and it names no defendant.

1. The claim, and how to read it

A model built on public research, a public network and the unpaid work of the people who wrote the web is a collective product held as private property. Held that way, and tuned by a private guideline nobody outside the provider can read, it concentrates economic power and epistemic power in the same hands. That is the claim. The remedy this paper argues for is not to stop building models. It is to attach to them the terms that already attach to other things built on public inputs: disclosure, benefit-sharing, audit and appeal.

This is an argument, not a measurement. It establishes nothing empirically and does not try to. What it can do is state a structure, show that the structure holds together, cite the public record for every factual step, and say where the argument stops. Section 10 does the last of those and is the section to read first if you are looking for the weak joint.

1.1 On naming

This edition names no company as a defendant. That is a deliberate change from the first edition, which built its case around one firm and one model, and it makes the argument stronger rather than softer. The problem described here is structural: it follows from who paid for the inputs, who holds the output and what governs the relationship, and it would follow in the same way whichever firms happened to occupy the positions. An argument that depends on the character of a particular company can be answered by that company. An argument about a structure has to be answered by changing the structure.

Where a specific proceeding or report carries a factual step, it is cited as a matter of public record in the reference list and described in the text by what it held rather than by whom it was against. No claim is made about the motives, competence or good faith of any individual, and no claim is made that any company has acted unlawfully. Two of the court decisions cited below came out substantially for the defendants, and section 3 says so.

1.2 Method, and a correction to the first edition

The first edition of this paper carried a byline that listed a language model as lead analyst and the author as research contributor. That was wrong in a way worth correcting in public. A model cannot take responsibility for a claim, and taking responsibility for claims is most of what authorship is. The byline now names the person who is answerable for the paper.

The tool use was real and is described here instead. Language models were used to draft passages, to summarise sources and to argue against the thesis while it was being written. Every source in the reference list was checked by the author against the original, and the numbers the first edition carried that could not survive that check were removed rather than re-sourced. Section 10 lists what went and why. The errors that remain are the author's.

2. Socialised inputs, privatised output

2.1 What the public paid for

The infrastructure that language models depend on was built with public money before it was privately used. The packet-switched network that became the internet was a defence research project whose first link came up in 1969. Its transport

protocols came out of publicly funded work in 1974. The web was written at a publicly funded physics laboratory between 1989 and 1991. The Transformer architecture that every current frontier model uses was published openly in 2017, and the first widely used pretrained language models were released with their weights and their code in 2019 and 2020. The two dominant training frameworks are open source. The chips come from a semiconductor industry that the CHIPS and Science Act of 2022 supports with 52.7 billion dollars of federal funding, and the research that produces the models qualifies for the federal research credit under 26 U.S.C. section 41.

None of that is illegitimate on its own, and the paper does not argue that it is. Public research is meant to be used, and the fact that it was used is a sign the policy worked. The paradox is in the accounting rather than in the use: the inputs were socialised and the output is claimed in full.

2.2 The condition that attaches everywhere else

The comparison that makes this concrete is pharmaceutical. A firm that develops a drug on federally funded research does not thereby own the result without conditions. Under the Bayh-Dole Act of 1980 the government retains a paid-up licence to practise the invention, and retains march-in rights if the invention is not made available to the public on reasonable terms. Those conditions are modest, rarely exercised, and they exist. No equivalent term attaches to a model trained on federally funded research over a publicly built network using a publicly funded architecture.

That asymmetry is the whole of the ownership paradox and it does not require anyone to have behaved badly to produce it. It is what happens when a body of law written for one kind of publicly subsidised output is not extended to another.

3. The corpus

The second uncompensated contribution is the training corpus. A modern model is trained on text and images crawled from the public web, and the crawl is gated only by robots.txt, a voluntary convention that was written in 1994 and only became a standards-track document, RFC 9309, in 2022. A convention is not a licence. Nothing in it establishes who may train on what.

The legal theory is fair use: training is transformative, and the model does not substitute for the works. The United States Copyright Office's May 2025 report on generative AI training accepts that some training is transformative and rejects the sweeping version of the claim. On its analysis, commercial use of large volumes of copyrighted works to produce content that competes with those works in their own markets is not settled fair use, and the knowing use of pirated copies weighs against a fair-use finding.

Two district court decisions in June 2025 drew a line in the same place from opposite directions. In the first, training on lawfully purchased books was held to be fair use while building a permanent internal library from pirated copies was not. In the second, the court found for the defendant on the record in front of it while stating expressly that the market-

harm question could come out differently on other evidence. Both decisions are cited in full in the reference list. Neither is a finding that training is unlawful, and this paper does not report them as one. A further suit filed in December 2023 over news content is pending and is cited on the same terms.

The ethical question is separate from the legal one and easier. The people whose work makes up the corpus were not asked, are not credited and are not paid. A minority of sources are now licensed: news publishers and large forums have signed agreements with model developers since 2023, and in those deals the payment goes to the platform or the publisher rather than to the person who wrote the sentence. The flow is asymmetric in a way that has nothing to do with anyone's intentions. Contributions enter from six directions. The money that comes out goes to shareholders, vendors, annotators and a few licensors, and to two of the six contributing groups it does not flow at all.

4. The forked commons

The third contribution is the open-source stack, and the pattern by which it was absorbed has three steps that are individually reasonable and collectively one-directional. Publish the architecture, which advances the field and earns citations. Absorb the community's improvements, which is what open licences permit. Build a closed product on top, and argue against open alternatives on safety grounds.

The provider's own addition at the third step is capital: compute, data curation and preference tuning at a scale no volunteer community can fund. That is a real contribution and it earns a real return. The argument here is narrow and should be read narrowly. It is not that the capital contribution is worthless. It is that a contribution to the last stage does not earn the whole stack, and that the parties who contributed the earlier stages have no mechanism by which to say so.

Ostrom's work on commons governance is the useful frame, because it locates the failure precisely. A commons fails when the people who use it cannot set and enforce rules for it. The model corpus is a commons whose contributors have no rule-making role at all: they cannot see what was taken, cannot set terms for its use, and cannot withdraw it once taken. The reforms in section 9 are, at bottom, ways of giving them one.

5. Tenure, obligation, and the missing exit

Digital feudalism is a claim about structure, not a claim about cruelty, and the paper is better off saying so before the analogy does any work. A manor had three tiers. The lord held the land by grant from the crown. The vassal held a fief from the lord in exchange for service and payments due on set occasions. The worker on the land owed labour and produce upward and could not leave.

The platform economy reproduces the tiers. The owner holds the weights, the compute and the distribution channel. The developer holds interface access and a ranking position at the owner's pleasure, owes a commission on every sale, and can be repriced, reranked or removed without appeal. The user holds an account, supplies attention, content and behavioural

data, receives the service at zero posted price, and cannot take their accumulated history anywhere else.

The analogy is most exact at the exit, which is also where the remedy is clearest. Medieval towns offered one: a worker who lived a year and a day inside a chartered town was free of the lord's claim. The modern equivalents exist in law already. Article 20 of the General Data Protection Regulation grants data portability. Article 7 of the Digital Markets Act imposes interoperability duties on the largest messaging gatekeepers. Both are in force. Neither yet reaches a model, a training corpus or a ranking feed, which are exactly the three things whose accumulation makes leaving expensive.

6. Preference tuning as private moderation

Reinforcement learning from human feedback is the method by which a provider's judgments about what a model should say become the model's behaviour. It was introduced for control tasks in 2017 and applied to language models in the 2022 paper that produced InstructGPT. The pipeline has three stages. Contractors write demonstrations of good answers and the model is fine-tuned on them. Contractors then rank pairs of model outputs and a reward model is fitted to those rankings. Finally the policy is optimised to score well on the reward model, with a penalty term that keeps it close to the fine-tuned model.

The judgments enter twice, as demonstrations in the first stage and as pairwise labels in the second, and are then applied at scale in the third. Who writes the guideline the labellers follow, and who the labellers are, is not published by any provider.

Two properties of the method matter for governance rather than for engineering. First, the reward model is the guideline generalised. Whatever the guideline says about politics, history, health or a competing product, the reward model says on every prompt it has never seen, and the guideline itself stays out of sight. Second, the optimisation favours the majority. Xiao and colleagues showed in 2024 that the regularised objective used in standard preference tuning produces what they call preference collapse: the policy concentrates on the preference of the majority of labellers and disregards minority preferences. The authors of the InstructGPT paper reported that their method reduced toxicity and did not significantly reduce bias. A model tuned this way is a consensus machine, and the consensus is the labellers'.

The labour is part of the record too. A January 2023 investigation reported that contractors in Kenya labelling harmful content for a major model developer were paid under two dollars an hour. The judgment that shapes a product used in every country is made by a small team writing a guideline

and a contracted workforce applying it under time pressure, without a role in the policy and without the ability to contest it.

Transparency has moved, and the movement should be credited rather than ignored. Two providers have published top-level policy documents, a constitution in 2023 and a model specification in 2024, and those are a real first layer. What remains private is the layer that determines behaviour: the labels, the annotator instructions, the composition of the labelling workforce, the evaluation sets and the reward model itself. When a model refuses a request, the person asking learns nothing about which rule fired, who wrote it, or how to contest it.

Restrictions on speech in a democracy are expected to meet five tests: legal authority, narrow tailoring to a specific harm, procedure, transparency of the rule, and a route of appeal. Private moderation at this scale meets none of them, and it is worth being precise about why that matters. The objection is not that providers moderate. It is that a system of rules governing expression for hundreds of millions of people has been built without any of the procedural furniture that every other such system is required to carry. The Digital Services Act already requires a statement of reasons and an internal complaint route from online platforms under Articles 17 and 20. It does not yet apply them to model providers. Extending an instrument that exists is a smaller ask than inventing one.

7. Where the income goes

The economic case does not need a forecast, which is fortunate, because the forecasts in the first edition of this paper were the least defensible part of it. Acemoglu and Restrepo's task-based model separates three effects of automation. The displacement effect: tasks that move from labour to capital take their wage income with them. The productivity effect: cheaper output raises demand for the tasks that remain with labour. The reinstatement effect: new tasks are created that only labour can do. Whether labour income rises or falls depends on the balance of the three, and their evidence from industrial robots in United States labour markets found the displacement effect dominating in the commuting zones most exposed.

The identity that follows is arithmetic and can be checked on paper. If a share a of labour tasks is automated and no new tasks are created, labour keeps a fraction $(1 - a)$ of its tasks, and a productivity gain g on those remaining tasks leaves labour income at $(1 - a)(1 + g)$ relative to where it started. Setting that to 1 gives the break-even gain, $g = a / (1 - a)$. At an automatable share of 30 percent, demand for the remaining tasks has to grow by about 43 percent for labour income merely to stand still.

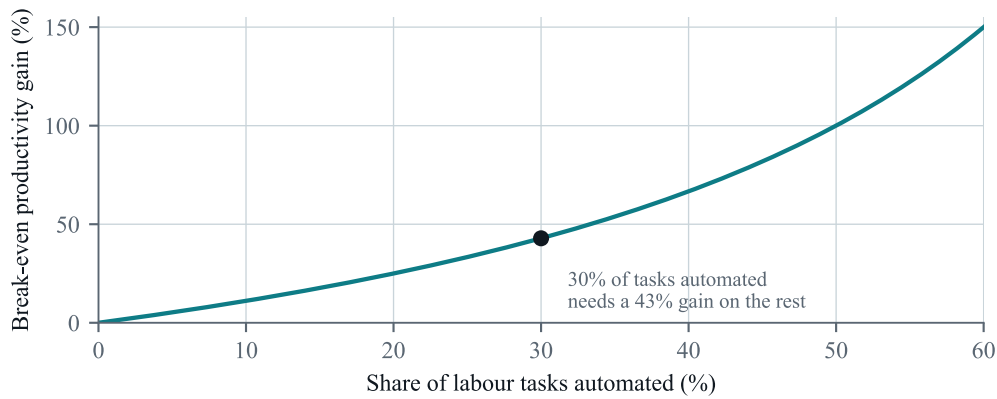


Figure 1. The break-even productivity gain against the automated task share, $g = a / (1 - a)$, with no reinstatement of new tasks. This is an identity evaluated rather than a forecast: it says what a given share would require, and nothing about what share will be automated.

What makes language models different from earlier automation is not the rate but the target. The tasks they take are drafting, review, summarisation, first-line support, routine analysis and routine code. Those are the tasks of the middle of the income distribution, in occupations that were secure through the last two waves of automation precisely because they were cognitive.

The fiscal consequence follows from the arithmetic rather than from any projection. Labour income is taxed at source, through payroll and personal income tax, in the country where the work is done. Capital income is taxed later, at lower effective rates, in a jurisdiction the owner chooses. When income moves from the first column to the second, the tax base moves with it, at the moment demand for retraining and income support rises. A benefit-sharing mechanism is therefore a fiscal instrument as much as a fairness one, and this paper treats it as both.

The ownership structure makes the distribution sharper than it was in earlier automation. When a factory automated, the machinery stood in the country where the workers were. It could be taxed, bargained over and, in the last resort, bought. A model is a set of weights on the owner's servers. The public contribution, through data, research and infrastructure, carries no equity and no revenue share, and there is no local asset to bargain over. That gap is what the equity and dividend proposals in section 9 are meant to close.

8. Two precedents, and what they are worth

Parliamentary enclosure in England, concentrated between 1760 and 1830, converted common land to private holdings and supplied the dispossessed workforce of the factories. The Luddite risings of 1811 to 1816 and the Chartist movement from 1838 were among the political responses, and the settlement that followed took most of a century. The mechanism worth carrying forward is not the violence. It is that the conflict came from the enclosure of the benefits by a narrow group while the costs fell on the majority, rather than from the technology itself.

The printing press produced the epistemic version earlier. The Index of Prohibited Books of 1559 and the licensing of printers, in England under the Licensing of the Press Act 1662, created an official literature beside an underground one,

and the two did not share a reality. The expectation of a bifurcated model ecosystem, sanitised proprietary systems beside unrestricted open ones, is that pattern again.

The argument this paper makes from the two is that they combine in the present case: the economic concentration of enclosure and the epistemic concentration of the licensed press, in the same hands, at the same time. The compression of the timescale is the part it treats as most serious. Enclosure's conflicts ran for the better part of a century and the licensed press's for longer than that, which gave institutions time to respond. A governance gap, in which owners entrench through network effects, data moats and regulatory capture before institutions respond, is the risk the reforms below are designed against.

What these precedents are worth is limited and the paper is explicit about it. They establish that concentration of a productive technology and concentration of an epistemic one have each produced prolonged conflict before. They do not establish that this case will follow the same course, they fix no date, and no part of the argument above depends on them. An analogy is a way of seeing a structure, not evidence about the future.

9. Reforms whose instruments already exist

The proposals below are grouped in three: benefit-sharing, so that contributors and the public share in what was built on their inputs; fiscal measures, so that the tax base follows the income; and auditable governance, so that the judgments embedded in a model can be seen, contested and appealed. Each is paired with an instrument already in force somewhere that does the same kind of thing. None requires inventing a new kind of institution. Each requires applying an existing one to model providers, which is a political problem rather than a design problem.

That framing is the practical argument for the list. A proposal that needs a new agency, a new legal theory and a new international body has three places to fail before it starts. A proposal that extends a disclosure duty already written, or applies a levy mechanism already collected in twenty jurisdictions, has one. The two tables below are ordered so that the first carries what moves money and the second carries what makes a judgment reviewable.

Table 1. Benefit-sharing and fiscal measures. The last row is a rejected proposal rather than a precedent: the resolution that considered a charge of this kind did not adopt one. Listing it as a success would be the sort of overreach this edition exists to remove.

Proposal	Mechanism	Instrument already in force
Training-data disclosure	Publish a sufficiently detailed summary of the content used for training, so a contributor can tell whether their work is in the corpus	EU AI Act, Article 53
Benefit-sharing fund	A levy on model services paid into a fund distributed to contributor classes, rather than per-work licensing that cannot be administered at corpus scale	Private copying levies under Directive 2001/29/EC, Article 5(2)(b)
Public equity or a public dividend	Where public research, data or compute went in, a non-voting stake or a per-resident dividend comes out	Bayh-Dole government licence and march-in rights; Alaska Permanent Fund Dividend, first paid 1982
Automation levy	A charge related to displaced work-hours, funding retraining and income support, designed so that augmentation is cheaper than replacement	Proposed and not adopted: European Parliament resolution of 16 February 2017 on Civil Law Rules on Robotics

Table 2. Auditable governance. Every instrument in the right-hand column already binds somebody. None of them yet binds a model provider on the terms in the middle column, and that gap is the whole of the proposal.

Proposal	Mechanism	Instrument already in force
Published moderation policy with reasons and appeal	A statement of reasons when a request is refused, an internal complaint route, and access to out-of-court settlement	Digital Services Act, Articles 17, 20 and 21
Independent audit	An accredited external auditor with access to the reward model, the annotator instructions and the evaluation sets	Digital Services Act, Article 37; New York City Local Law 144 of 2021 bias audits
Portability and interoperability	The right to take a history, a corpus or a fine-tune elsewhere, and duties on the largest providers to make that possible	GDPR Article 20; Digital Markets Act Article 7
Public compute	Publicly funded compute and data access so that research outside the large providers remains possible	National AI Research Resource pilot, launched January 2024
International coordination	A standard-setting body with inspection powers, to prevent a race to the weakest jurisdiction	IAEA Statute, in force 1957; Council of Europe Framework Convention on AI, opened for signature 5 September 2024

The political conditions are not reforms but they determine whether reforms happen, and they are worth naming so that the list above is not read as a wish. Coordinated action across jurisdictions, because a single state acting alone invites relocation. Coalitions among displaced workers, creators and open-source developers, because no one of those groups is large enough alone. Litigation that fixes the limits of fair use, because the boundary is currently being set case by case. And the plain refusal of contributors to keep contributing on the current terms, which is the only lever that does not require anyone's permission.

10. What this argument does not establish

It is an argument, and it proves nothing empirically. Every section above is a claim about structure supported by the public record. None of it is a measurement, and a reader should not take the citation of a court decision or a statute as evidence for the structural claim it is attached to. It is evidence for the narrow factual step it is cited for.

It forecasts nothing. The first edition of this paper set a horizon of 2035 and made dated predictions: a share of cognitive tasks automated by 2030, a number of workers affected, a figure for global profits from AI services, an average earnings loss for displaced workers. Not one of those had a source, and they have all been removed rather than re-sourced. Nothing in the argument requires a date and nothing here supplies one.

Four other numbers from the first edition are gone for the same reason and are listed so that a reader of both editions knows what changed: a total for federal artificial intelligence research funding since 1990, a token count for a particular commercial model, an effective corporate tax rate, and a worked example of a firm replacing staff whose every figure was assumed. The reform percentages went too. The first edition proposed a specific revenue share, a specific public equity stake and a specific excess-profits rate; those rates were invented, and section 9 now describes the mechanism and its precedent without them.

The task-share identity in section 7 is an identity. It tells you what a given automated share would require of productivity growth for labour income to hold; it says nothing whatever about what share will be automated, what gain will be realised, or how many new tasks will be created. The empirical work it rests on measured industrial robots rather than language models, and the authors' own caveats about that evidence apply here with full force.

The feudal analogy is an analogy. Structural resemblance is not identity, and the disanalogies are real: a user can close an account today, nobody is bound to the land, the relationship is contractual, and the service is often supplied at no posted price and is often worth having. The claim is narrower than the word feudalism suggests on its own, and it is the claim in section 5 rather than the word that should be argued with.

The preference-collapse result is a property of an objective under an assumption set, demonstrated in one paper. It is not a measurement of any deployed model's political content, and this paper makes no claim about what any deployed model says about any topic. The governance objection in section 6 stands on the absence of published rules and appeal routes, not on any assertion about the direction in which a particular model leans.

Nothing here measures how much income has actually moved from labour to capital as a result of language models specifically. That measurement does not yet exist in a form this paper could cite, and its absence is why section 7 argues from an identity rather than from a trend.

The reforms are not costed and their second-order effects are unmeasured. A levy that incumbents can pay and entrants cannot would entrench exactly the concentration it was meant to reduce. A disclosure duty could be met with a summary too coarse to be useful. An audit regime is only as good as the accreditation behind the auditors. Each of those is a real risk and none is answered here.

Finally, the author has an interest. This paper is published by a company that builds small models and open formats and would benefit commercially from a world with more open infrastructure in it. That does not make the argument wrong, and a reader is entitled to weigh it, which is why it is stated here rather than left to be discovered.

11. Open questions

Five questions the argument raises and does not answer. They are the places where work would change the conclusion rather than decorate it.

- Whether per-work compensation can be administered at corpus scale at all, or whether a levy into a fund, on the private-copying model, is the only workable form.
- How to measure displaced work-hours for an automation levy without also taxing augmentation, which the same levy is meant to encourage.
- Whether publishing reward-model instructions changes model behaviour for the worse, through labellers or users gaming a published rule.
- What an independent auditor would need to see in order to certify a reward model, and who accredits the auditor.
- Whether public compute changes the market or only subsidises the same vendors through a different door.

References

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems* 30.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL-HLT* 2019.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. and Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research* 21(140).

Christiano, P., Leike, J., Brown, T., Martic, M., Legg, S. and Amodei, D. (2017). Deep Reinforcement Learning from Human Preferences. *Advances in Neural Information Processing Systems* 30.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J. and Lowe, R. (2022). Training Language Models to Follow Instructions with Human Feedback. *Advances in Neural Information Processing Systems* 35.

Xiao, J., Li, Z., Xie, X., Getzen, E., Fang, C., Long, Q. and Su, W. J. (2024). On the Algorithmic Bias of Aligning Large Language Models with RLHF: Preference Collapse and Matching Regularization. *arXiv:2405.16455*.

Acemoglu, D. and Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives* 33(2), 3-30.

Acemoglu, D. and Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy* 128(6), 2188-2244.

U.S. Copyright Office (2025). Copyright and Artificial Intelligence, Part 3: Generative AI Training. Pre-publication version, May 2025.

Bartz v. Anthropic PBC (N.D. Cal.), order on fair use of 23 June 2025.

Kadrey v. Meta Platforms, Inc. (N.D. Cal.), order on fair use of 25 June 2025.

The New York Times Company v. Microsoft Corporation and OpenAI (S.D.N.Y.), complaint filed 27 December 2023.

Perrigo, B. (2023). Exclusive: The \$2 Per Hour Workers Who Made ChatGPT Safer. *TIME*, 18 January 2023.

Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E. and Winfield, A. (2020). The Ethics of Artificial Intelligence: Issues and Initiatives. *European Parliamentary Research Service*, PE 634.452.

Regulation (EU) 2024/1689 (Artificial Intelligence Act), Article 53.

Regulation (EU) 2022/2065 (Digital Services Act), Articles 17, 20, 21 and 37.

Regulation (EU) 2016/679 (General Data Protection Regulation), Article 20.

Regulation (EU) 2022/1925 (Digital Markets Act), Article 7.

Directive 2001/29/EC (Information Society Directive), Article 5(2)(b).

CHIPS and Science Act of 2022, Public Law 117-167.

26 U.S.C. section 41, Credit for increasing research activities.

Bayh-Dole Act, Public Law 96-517 (1980), codified at 35 U.S.C. sections 200-212.

Alaska Statutes section 43.23, Permanent Fund Dividend; first payment 1982.

European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics, 2015/2103(INL).

National Science Foundation (2024). National Artificial Intelligence Research Resource Pilot, launched January 2024.

Council of Europe (2024). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, CETS No. 225, opened for signature 5 September 2024.

New York City Local Law 144 of 2021, automated employment decision tools.

Statute of the International Atomic Energy Agency, in force 29 July 1957.

RFC 9309 (2022). Robots Exclusion Protocol. Internet Engineering Task Force.

Anthropic (2023). Claude's Constitution. OpenAI (2024). Model Spec. Cited as the two published top-level policy documents referred to in section 6.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.

Licensing of the Press Act 1662, 13 and 14 Car. II c. 33 (England).

Index Librorum Prohibitorum (1559), Congregation of the Roman Inquisition.

The essay version of this argument, with figures the reader can set:
<https://sophiaxt.com/research/digital-feudalism>