Practical Principles for AI Cost and Compute Accounting

Stephen Casper [1] **Luke Bailey** [2] **Tim Schreier** [3]

Abstract

Policymakers increasingly use development cost and compute as proxies for AI capabilities and risks. Recent laws have introduced regulatory requirements that are contingent on specific thresholds. However, technical ambiguities in how to perform this accounting create loopholes that can undermine regulatory effectiveness. We propose seven principles for designing AI cost and compute accounting standards that (1) reduce opportunities for strategic gaming, (2) avoid disincentivizing responsible risk mitigation, and (3) enable consistent implementation across companies and jurisdictions.

1. Introduction

As AI systems become more capable, policymakers face a mounting challenge: identifying which AI systems warrant heightened oversight. Recent laws and governance frameworks have approached this challenge by proposing requirements contingent on development costs and computational resources (e.g., European Parliament and Council, 2024; California SB1047, 2024; Li et al., 2025; U.S. Bureau of Industry and Security, 2025). For example, the now rescinded 2025 *Framework for Artificial Intelligence Diffusion*, issued by the U.S. Bureau of Industry and Security, sought to control the export of AI model weights for systems whose development exceeded 10^{26} floating point operations (FLOP) in training compute (U.S. Bureau of Industry and Security, 2025).

Development costs and compute are compelling metrics for use in AI governance (Li et al., 2025) because they "correlate with capabilities and risks, are quantifiable, can be measured early in the AI lifecycle, and can be verified by external actors" (Heim & Koessler, 2024). Cost and compute thresholds also allow regulators to focus oversight on the most advanced AI models while avoiding unnecessary burdens on smaller developers. However, their practicality as regulatory tools depends on establishing clear accounting standards. This requires resolving several technical ambiguities about what should be counted (Hooker, 2024; Heim & Koessler, 2024; Reuel et al., 2024).

This paper asks the question: ***How can the cost and compute used during model development be counted in a way that is practical, limits gameability, and avoids disincentivizing responsible risk management?*** Key challenges include determining which activities to count, establishing reporting requirements, and allowing for standards to adapt as technology evolves.

To address these challenges, we propose seven principles for designing practical AI cost and compute accounting standards. We argue that these principles can resolve technical ambiguities while aligning with public interest and enabling consistent implementation across companies and jurisdictions. While we do not take positions on specific thresholds or requirements, our framework provides a foundation for developing robust standards for AI cost and compute accounting.

2. Background

Related work. AI research has studied "scaling laws" demonstrating how AI model performance improves predictably with increased computational resources and data (Kaplan et al., 2020; Villalobos, 2023; Sevilla & Roldán, 2024). While these relationships provide a scientific basis for the theoretical value of cost and compute thresholds in AI regulation (Sastry et al., 2024), researchers have identified important limitations and highlighted unsolved implementation challenges. Without a standardized methodology, complex technical questions about which activities to count and how to report such counts remain unresolved (Hooker, 2024; Heim & Koessler, 2024; Yew et al., 2025).

To our knowledge, there are only two sets of open, pre-existing guidelines for cost and compute accounting. The first was published as an issue brief by the Frontier Model Forum (2024), a collaboration between major tech companies that represents industry interests (Wei et al., 2024), while the second appears in the draft guidelines on general-

[1]MIT CSAIL (this work was done partly at MIT CSAIL and partly independently) [2]Stanford University [3]Future of Life Institute. Correspondence to: Stephen Casper <scasper@mit.edu>.

Accepted to the ICML 2025 workshop on Technical AI Governance

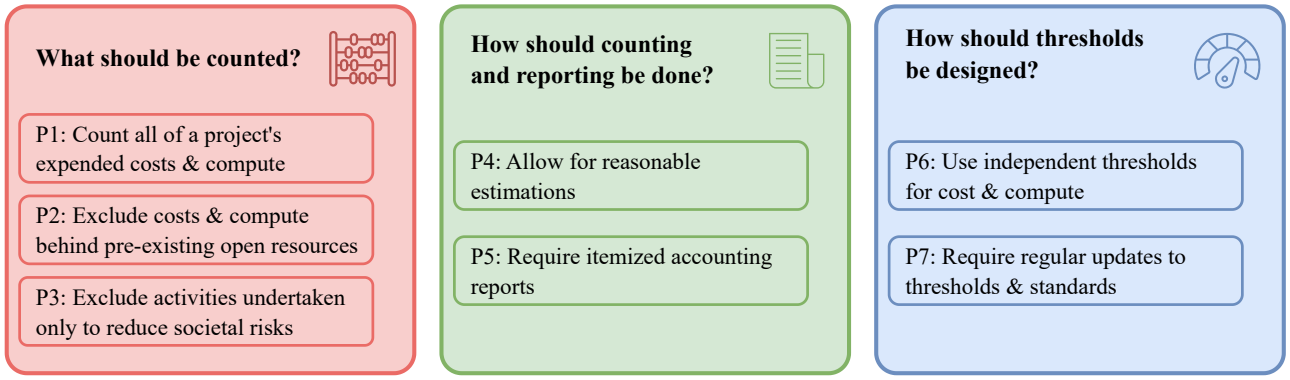

Figure 1. **Seven principles for cost and compute accounting discussed in Section 3.**

purpose AI issued by the European Commission (2025) in support of the EU AI Act. In Figure 2, we compare the principles we recommend to those recommended by each of these prior reports. Here, we echo the (Frontier Model Forum, 2024) in support of excluding open resources from accounting, the European Commission (2025) in support of counting data curation and teacher models used for distillation, and both prior reports in their support of allowing for reasonable estimates. However, we argue that previously recommended practices, such as context-dependent calculations, limiting counts to end-to-end training only, and excluding recomputations and discarded branches, would enable developers to game the system by omitting substantial portions of their development process from oversight (Section 3.1).

Key terms. In the context of this paper:

- **Developer** refers to the entity undertaking the process of creating an AI model. A single developer can encompass multiple legal entities in formal partnership.[1]
- **Development** refers to the process of curating data, training models, creating scaffolding, and testing AI systems. It does not encompass human labor, operations, or procuring hardware.
- An **AI model** refers to a neural information processing structure trained using machine learning. An **AI system** refers to a set of one or more AI models combined with other software components to accomplish a specific task. For example, GPT-4o (Hurst et al., 2024) is an AI model while ChatGPT-GPT-4o is an AI system.

What if multiple models are very similar? It is possible to develop two distinct but very closely related models. For example, two models may only differ by a small amount of fine-tuning if they are different derivatives of the same 'base' model. This poses a challenge to regulators because two such models will often, but not always, have similar behaviors. It may also often be impractical to subject multiple models to potentially redundant requirements. For this reason, policymakers may want to make developers or 'model families' the object of regulation as opposed to individual models. However, recommendations for how to practically handle these cases are beyond the scope of this paper.

[1]We use this definition to preclude loopholes involving multiple legal entities formally collaborating to develop a single model. Detailed standards will need to account for collaborative developments, including crowdsourced or federated approaches.

3. Principles

3.1. Count all of a project's expended costs and compute

Principle: Count all technical costs and compute that the developer expends in the process of developing an AI model, not simply theoretical, proximal, or upstream ones.

Purpose: Closing loopholes (especially involving distillation), and limiting the gameability of accounting standards.

Developers undertake a variety of activities during frontier models development. However, a narrow view of what counts could be used to exclude certain activities integral to the model development process. For example:

- Some activities are not theoretically needed for the final model to have been produced. For example, in the process of training models, there are often many multiplications or additions by zero due to the use of dropout (Srivastava et al., 2014) and sparsity (e.g., Correia et al., 2019). Even though they are not theoretically needed, they are carried out by hardware nonetheless and are often used to improve performance.
- Some activities are not proximal to the model's training process. For example, dataset creation/curation/compression (e.g., Kaddour, 2023; Solaiman & Dennison, 2021; Chen & Mueller, 2024) or training a teacher model for distillation (Yang et al.,

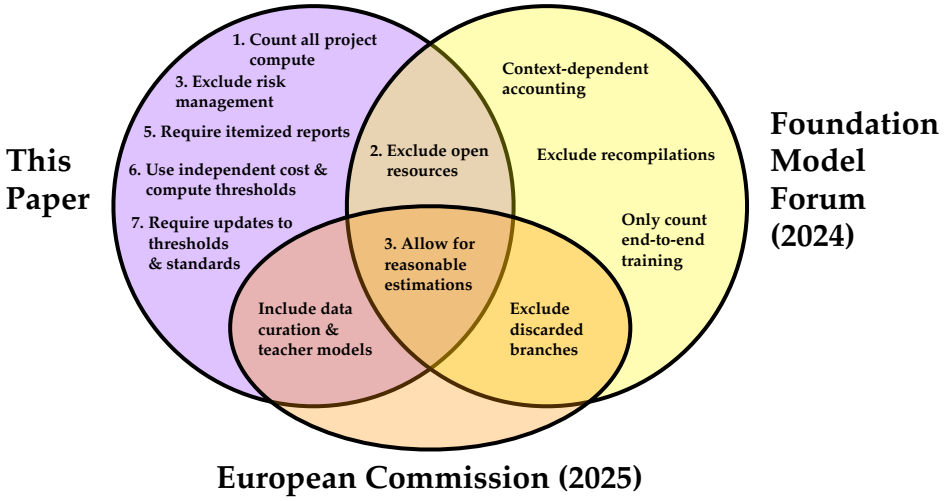

Figure 2. **Comparing principles proposed here, in Frontier Model Forum (2024), and in European Commission (2025).** In Section 3, we argue that some exclusions recommended by Frontier Model Forum (2024) and European Commission (2025) could allow for loopholes. Most notably, the recommendations from Frontier Model Forum (2024) allow for the "distillation loophole" (see Section 3.1) in which developers sample-efficiently train a model on the outputs of a 'teacher' model without accounting for the teacher's training.

2024) are not directly involved in the final model's training. However, these activities are nonetheless integral to the model's development and capabilities.

- Some activities are not upstream of the model training process. For example, developers often iteratively train, evaluate, retrain, re-evaluate, etc. until they have a system that meets their desired specifications. Other times, they will train multiple models using different setups and simply select the one which performs best (Singh et al., 2025). Although some evaluations and branches are not directly upstream of the final model, they are nonetheless an integral part of the development process.

When there exists an incentive to make a model seem very cheap using narrow accounting standards, developers can find creative ways to game the count (Yew et al., 2025). One of the most pertinent ways that narrow accounting can be used to obscure a model's total development costs involves model distillation. For instance, DeepSeek-V3 (Liu et al., 2024) was produced, in part, by distilling the more powerful DeepSeek-R1 (Guo et al., 2025), yet its widely quoted $6 million budget excludes the compute spent training DeepSeek-R1—making the total project appear far less costly than it was in actuality, demonstrating how narrow accounting can mask large hidden costs. For this reason, accounting standards that do not close the distillation loophole in particular may be highly limited in their effectiveness.

> *"Any statistical relationship will break down when used for policy purposes."*
> – Jón Danielsson

Finally, sometimes, there will exist genuine grey-areas related to whether a certain activity was a meaningful part of a model's development process. For example, if a developer conducts basic research to develop techniques that they will use downstream for the development of a model, should that be counted? Since company R&D processes are never completely isolated, some ambiguity in attributing activities to specific model development is inevitable. However, Section 3.5 will discuss reporting requirements as a mechanism for fostering transparency and accountability for cost and compute accounting practices.

3.2. Exclude costs and compute behind pre-existing, openly-available resources

Principle: Count costs and compute that the developer directly incurs through their activities, purchases, and partnerships. Exempt the costs and compute used to produce open resources that developers obtain for free.

Purpose: Practicality and focus on proprietary resources.

Developers can produce capable models through multiple sources of cost and compute. They often curate their own data and train their own models in-house. However, they can also purchase resources, query systems from external providers, and outsource parts of the development process to partners. For the reasons outlined in Section 3.1, these are generally needed for thorough accounting.

We argue that the costs and compute behind pre-existing, openly-available resources should be excluded, both because their model- and data-provenance records are often inconsistent (Mitchell et al., 2019; Longpre et al., 2023)

and because such resources already provide a zero-effort capability floor.

However, one modification to this exemption may be necessary to close a loophole. If resources that were recently (e.g., within 6 months prior to when a model's development begins) and openly released by the developer[2] in question, regulators may wish to require that this resource is still counted. Without this exception, developers could openly release partially-developed model components (e.g., a pretrained base model) to exclude them from accounting.

As a final note, regulators may wish to uniquely handle cases in which a developer begins with an open model whose development already passed thresholds and further develops it. For example, a failed 2024 California bill (California SB1047, 2024) defined a "covered" model in terms of either a primary threshold or a secondary threshold for when additional development is applied to an existing "covered" model. This type of strategy may be appealing to regulators because of how modest amounts of further development of highly capable models can substantially alter their capabilities. However, recommendations on how regulators should handle these cases are beyond the scope of this paper.

3.3. Exclude activities undertaken only to reduce societal risks

Principle: Allow developers to exempt activities undertaken only for the purpose of reducing risks to society which do not have side effects of enhancing model capabilities.

Purpose: Incentivizing societal risk-reduction practices.

Over the course of developing an AI model, key activities are undertaken either partially or entirely to improve its capabilities. For example, pretraining and fine-tuning are principally meant to make models more capable. However, some activities are undertaken strictly to reduce risks. Examples include filtering child sexual abuse material (CSAM) from training data (Thiel, 2023)), fine-tuning models to refuse criminal requests (Yuan et al., 2024), and testing for national security risks (Shevlane et al., 2023). To avoid disincentivizing such measures, developers must be allowed to exclude these types of activities from their accounting.

How should it be determined when an activity is undertaken only for the purpose of mitigating societal risks? This can be difficult due to the lack of a clean dichotomy and the prevalence of "safetywashing" (Ren et al., 2024). To mitigate this challenge, developers can be required to produce a rigorous, auditable justification for why an activity only reduces risks without simultaneously increasing capabilities in an accounting report (see Section 3.5).

[2]Recall in Section 2 that we define "developer" to include formal collaborations between multiple legal entities. This type of definition would prevent multiple legal entities from spliting the development process via open-weight checkpoints to avoid passing a threshold so long as they had a formal agreement to do so.

3.4. Allow for reasonable estimates

Principle: When counting costs and compute used for a model's development, developers should be permitted (and often expected) to use reasonable estimations when precise information is not practically attainable.

Purpose: Practicality.

Information about costs and compute expended during a model's development is not always precisely quantifiable. For example, developers will often not know exactly how much compute has been expended when they query a closed-source system from some outside provider. However, in a case like this, reasonable estimates can be made based on contextual knowledge and the market value of compute (Sevilla et al., 2022; Cottier et al., 2024). Such estimates parallel 'fair value' asset estimations in financial accounting (IFRSF, 2022). Some imprecision is inevitable. However, to reduce the risk of estimations being gamed or resulting in unreliable counts, developers can be required to provide a report on their approach to accounting that documents estimates and justifications (see Section 3.5). Meanwhile, regulators or standards bodies could publish guidance on appropriate estimation methodologies, tolerable error margins, and suitable documentation templates.

3.5. Require itemized accounting reports

Principle: Require developers to produce an auditable, itemized accounting report detailing their approach to accounting, including justifications for estimates and exemptions.

Purpose: Transparency and accountability.

In financial accounting, companies are regularly required to send records and reports to governing bodies (e.g., U.S. SEC, Division of Corporation Finance, 2017). This has both the direct effect of helping government oversight offices spot issues and the indirect effect of incentivizing due diligence from companies. The same applies to AI cost and compute accounting. These reports would also be key for developers to provide explanations and justifications for technical exemptions (Section 3.1), open resource exemptions (Section 3.2), risk mitigation exemptions (Section 3.3), and estimations (Section 3.4). Such reports would improve accountability around accounting practices and inform regulators about industry trends in development expenditures.

For accounting reports to be effective, they must contain sufficient detail to enable meaningful review. For example, regulators may wish to offer standardized guidance on reporting including, for each distinct activity involving data curation, pretraining, fine-tuning, or testing:

- A clear description of the activity and its purpose.
- An explanation (and, if necessary, evidence) for if and how the activity relied on open resources.
- An explanation (and, if necessary, evidence) for whether the activity was undertaken *solely* for risk reduction.
- An explanation (and, if necessary, evidence) for estimations involved in accounting for that activity.
- The calculation used to quantify cost and compute for the activity.
- The activity's final accounted cost and compute.

3.6. Use independent thresholds for costs and compute

Principle: Regulatory requirements should be independently triggered by separate thresholds for cost and compute.

Purpose: Closing loopholes, and limiting the gameability of accounting standards.

Cost and compute are correlated, but they can still be decoupled, especially when developers have an incentive to game standards. For example, machine-generated data is cheap but computationally intensive while human-generated data is expensive but computationally free. A developer could design a project to be low-cost/high-compute or vice versa by adjusting the extent to which they use machine- versus human-generated data. As a result, having both cost and compute triggers would reduce gameability.

Separate cost and compute thresholds can also serve as failsafes for each other in case of error or fraud. For example, if a developer purchases queries or other services from an external provider, precise information on the amount of compute used might not be available, but the costs are unambiguous and auditable. Meanwhile, different computing devices can use different amounts of power to perform the same computations, but the compute is unambiguous.

3.7. Require regular updates to thresholds and standards

Principle: Require that thresholds and accounting standards are regularly updated to reflect technological developments.

Purpose: Ensuring standards remain effective by adapting to technological advances and evolving societal needs.

Rapid developments in AI technology create uncertainty about how scaling trends and efficiency gains will affect developers' costs and compute requirements (Pilz et al., 2023). Accordingly, governance frameworks will need to be adaptive to ensure they remain relevant over time. To regulate incisively, government offices and/or standards bodies will need to revisit and curate standards on a regular (e.g., quarterly or semiannual) basis in response to new developments in the state-of-the-art.

4. Discussion

Significance: Regulatory thresholds involving cost and compute are a uniquely practical (Heim & Koessler, 2024; Li et al., 2025) yet technically challenging (Hooker, 2024) strategy for designing regulations that target frontier AI models. To make cost and compute thresholds more tenable as a regulatory strategy, standards for accounting must be clear, consistent, and aligned with public interest. To support the development of such standards, we have proposed a principles-first framework to resolve ambiguities and introduced seven principles designed to reduce gameability, avoid disincentivizing societal risk mitigation practices, and enable consistent implementation across companies and jurisdictions.

Drawbacks: Despite the benefits discussed in this paper, rigorous accounting standards can also come with drawbacks. These may including undesired regulatory burden and information security risks that come from sharing of potentially sensitive information with regulators. Regulators should weigh these potential drawbacks against the benefits discussed above when designing standards.

Limitations: This work was not written in the context of any specific law. We make no recommendations about what kinds of regulatory requirements should be triggered and how. Key questions about how high to set thresholds, what they should trigger, and how they should be incorporated into legal frameworks are all beyond the scope of this paper. Furthermore, as we have discussed, while the principles presented here can greatly reduce ambiguity, grey areas will be inevitable. However, this underscores the role that itemized reporting can play in regulatory awareness. Due to the constantly evolving nature of AI technology, we recommend a paradigm of maintaining continuous awareness and oversight while allowing regulatory regimes to evolve rather than pursuing a single fixed regime.

Future work: Whereas this paper has sought to outline principles for designing standards, future work will be needed to produce concrete standards. Implementing these principles in practice will require specific attention to the purpose and scope of any individual law and may require compromises to ensure logistical and/or political feasibility.

Acknowledgments

We would like to thank Anthony Aguirre, Alexander Erben, James Petrie, Joe Kwon, Lennart Heim, Mark Brakel, Nandi Schoots, and Richard Mallah for discussions and feedback. Stephen Casper and Luke Bailey are funded by a Future of Life Institute Vitalik Buterin Fellowship. Luke Bailey is also funded by the SAP Stanford Graduate Fellowship.

References

California SB1047. California senate bill 1047: Safe and secure innovation for frontier artificial intelligence models act, 2024. URL https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB1047. Vetoed by Governor on September 29, 2024.

Chen, J. and Mueller, J. Automated data curation for robust language model fine-tuning. *arXiv preprint arXiv:2403.12776*, 2024.

Correia, G. M., Niculae, V., and Martins, A. F. Adaptively sparse transformers. *arXiv preprint arXiv:1909.00015*, 2019.

Cottier, B., Rahman, R., Fattorini, L., Maslej, N., and Owen, D. The rising costs of training frontier ai models. *arXiv preprint arXiv:2405.21015*, 2024.

European Commission. Commission seeks input to clarify rules for general-purpose ai models. https://digital-strategy.ec.europa.eu/en/news/commission-seeks-input-clarify-rules-general-purpose-ai-models, June 2025. https://digital-strategy.ec.europa.eu/en/news/commission-seeks-input-clarify-rules-general-purpose-ai-models.

European Parliament and Council. Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence (artificial intelligence act), article 51: Classification of general-purpose ai models as general-purpose ai models with systemic risk. *Official Journal of the European Union*, 2024. URL https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

Frontier Model Forum. Issue brief: Measuring training compute, May 2024. URL https://www.frontiermodelforum.org/updates/issue-brief-measuring-training-compute/.

Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Heim, L. and Koessler, L. Training compute thresholds: Features and functions in ai regulation. *arXiv preprint arXiv:2405.10799*, 2024.

Hooker, S. On the limitations of compute thresholds as a governance strategy. *arXiv preprint arXiv:2407.05694*, 2024.

Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A., Welihinda, A., Hayes, A., Radford, A., et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024.

Kaddour, J. The minipile challenge for data-efficient language models. *arXiv preprint arXiv:2304.08442*, 2023.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

Li, F.-F., Cuéllar, M.-F., and Chayes, J. T. Draft report of the joint california policy working group on ai frontier models. Technical report, Joint California Policy Working Group on AI Frontier Models, March 2025. URL https://www.cafrontieraigov.org/. Draft report released for public feedback.

Liu, A., Feng, B., Xue, B., Wang, B., Wu, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.

Longpre, S., Mahari, R., Chen, A., Obeng-Marnu, N., Sileo, D., Brannon, W., Muennighoff, N., Khazam, N., Kabbara, J., Perisetla, K., et al. The data provenance initiative: A large scale audit of dataset licensing & attribution in ai. *arXiv preprint arXiv:2310.16787*, 2023.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., and Gebru, T. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pp. 220–229, 2019.

Pilz, K., Heim, L., and Brown, N. Increased compute efficiency and the diffusion of ai capabilities. *arXiv preprint arXiv:2311.15377*, 2023.

Ren, R., Basart, S., Khoja, A., Gatti, A., Phan, L., Yin, X., Mazeika, M., Pan, A., Mukobi, G., Kim, R. H., et al. Safetywashing: Do ai safety benchmarks actually measure safety progress? *arXiv preprint arXiv:2407.21792*, 2024.

Reuel, A., Bucknall, B., Casper, S., Fist, T., Soder, L., Aarne, O., Hammond, L., Ibrahim, L., Chan, A., Wills, P., et al. Open problems in technical ai governance. *arXiv preprint arXiv:2407.14981*, 2024.

Sastry, G., Heim, L., Belfield, H., Anderljung, M., Brundage, M., Hazell, J., O'Keefe, C., Hadfield, G. K., Ngo, R., Pilz, K., et al. Computing power and the governance of artificial intelligence. *arXiv preprint arXiv:2402.08797*, 2024.

Sevilla, J. and Roldán, E. Training compute of frontier ai models grows by 4–5x per year. *Epoch AI, May*, 28, 2024.

Sevilla, J., Heim, L., Hobbhahn, M., Besiroglu, T., Ho, A., and Villalobos, P. Estimating training compute of deep learning models. *Epoch, January*, 20, 2022.

Shevlane, T., Farquhar, S., Garfinkel, B., Phuong, M., Whittlestone, J., Leung, J., Kokotajlo, D., Marchal, N., Anderljung, M., Kolt, N., et al. Model evaluation for extreme risks. *arXiv preprint arXiv:2305.15324*, 2023.

Singh, S., Nan, Y., Wang, A., D'Souza, D., Kapoor, S., Üstün, A., Koyejo, S., Deng, Y., Longpre, S., Smith, N., et al. The leaderboard illusion. *arXiv preprint arXiv:2504.20879*, 2025.

Solaiman, I. and Dennison, C. Process for adapting language models to society (palms) with values-targeted datasets. *Advances in Neural Information Processing Systems*, 34: 5861–5873, 2021.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

Thiel, D. Identifying and eliminating csam in generative ml training data and models. Technical report, Technical Report. Stanford University, Palo Alto, CA. https://purl. stanford . . . , 2023.

U.S. Bureau of Industry and Security. Framework for artificial intelligence diffusion. https://www.federalregister.gov/documents/2025/01/15/2025-00636/framework-for-artificial-intelligence-diffusion, 2025. Federal Register Document No. 2025-00636, 90 FR 4544, Department of Commerce, Bureau of Industry and Security. Published January 15, 2025.

U.S. SEC, Division of Corporation Finance. Financial reporting manual: Topic 2 - other financial statements required, 2017. URL https://www.sec.gov/corpfin/cf-manual/topic-2. Accessed: 2025-02-15.

Villalobos, P. Scaling laws literature review. *Epoch AI*, 2023. URL https://epoch.ai/blog/scaling-laws-literature-review.

Wei, K., Ezell, C., Gabrieli, N., and Deshpande, C. How do ai companies "fine-tune" policy? examining regulatory capture in ai governance. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, pp. 1539–1555, 2024.

Yang, C., Zhu, Y., Lu, W., Wang, Y., Chen, Q., Gao, C., Yan, B., and Chen, Y. Survey on knowledge distillation for large language models: methods, evaluation, and application. *ACM Transactions on Intelligent Systems and Technology*, 2024.

Yew, R.-J., Marino, B., and Venkatasubramanian, S. Red teaming ai policy: A taxonomy of avoision and the eu ai act. *arXiv preprint arXiv:2506.01931*, 2025.

Yuan, Y., Jiao, W., Wang, W., Huang, J.-t., Xu, J., Liang, T., He, P., and Tu, Z. Refuse whenever you feel unsafe: Improving safety in llms via decoupled refusal training. *arXiv preprint arXiv:2407.09121*, 2024.