An Alternative to Regulation: The Case for Public AI

Nicholas Vincent
Simon Fraser University
Computing Science
nvincent@sfu.ca

David Bau
Northeastern University
Khoury College of Computer Sciences
davidbau@northeastern.edu

Sarah Schwettmann
Massachusetts Institute of Technology
Computer Science and Artificial Intelligence Lab
schwett@mit.edu

Joshua Tan
Oxford University and Metagov
joshua.tan@magd.ox.ac.uk

Abstract

Can governments build AI? In this paper, we describe an ongoing effort to develop "public AI"—publicly accessible AI models funded, provisioned, and governed by governments or other public bodies. Public AI presents both an alternative and a complement to standard regulatory approaches to AI, but it also suggests new technical and policy challenges. We present a roadmap for how the ML research community can help shape this initiative and support its implementation, and how public AI can complement other responsible AI initiatives.

1   Introduction

In light of continued progress in AI, there is growing concern about the status quo in which powerful consumer-facing AI systems are operated almost exclusively by a small number of private firms, who primarily build and deploy AI systems using privately owned computing infrastructure (see e.g. discussion such as [34, 36, 27]). In response, a number of voices have called for public bodies to directly support the development and deployment of public interest AI models [19, 29, 28]. In the UK and Sweden, there are early efforts to organize country-level ChatGPT-like systems [3, 25, 31]. In the US, there are efforts to build a National AI Research Resource (NAIRR) [21], and a recent executive order signaled support for both government use of and investment in AI that serves the public good [16].

These efforts are early steps towards *public AI*—publicly accessible AI models funded, provisioned, and governed by governments or other public bodies. Public AI might take the form of a "public option" service for large language models (like a public option for banking), a national AI agency (like a national health service), or a set of library-like organizations that offer self-serve access to and support for a variety of models. In any of its instantiations, public AI offers a realistic institutional alternative to an ecosystem where AI models, especially foundation models [4], are primarily maintained by private corporate actors (with or without regulation). As an approach for provisioning access to AI, public AI also offers an alternative to open-source AI and decentralized AI.

Public AI projects share a goal of creating AI systems that are built, governed, and operated in accordance with shared values and for the public's benefit. They also provide a mechanism to ensure public funding for AI research results into public benefit rather than private capture. Ethically, developing public AI reflects the shared, open, and public nature of the internet and of the cultural data upon which AI foundation models are built [17, 20]. For a longer rationale, we defer to [2].

In this paper we describe the public AI concept, comparing it to other regulatory approaches. We then propose ways for the research community to contribute to and make use of public AI, organized in terms of open challenges and new opportunities for research. In other words, we explain how public AI motivates exciting new research questions, and why public AI can directly benefit researchers in academic and non-academic settings.

2 Background

There is a long-running tradition of public bodies funding and otherwise supporting emerging technologies. In this sense, public AI is an extension of the logic that has engaged governments in high-stakes technological development [8].

A variety of ongoing initiatives that range from nascent ideas to fleshed-out policy proposals aim to build something close to the definition of *public* we use here—AI technologies that are publicly accessible, publicly funded and/or provisioned, and governed by public bodies accountable to the public. This definition permits a wide range of approaches that span different processes for provisioning AI outputs, auditing and updating models, and funding and handling operational infrastructure challenges in collaboration with private and nonprofit entities. "AI models" can refer to a broad class of data-dependent systems, but our discussion specifically centers foundation models [4]. However, the positive outcomes from investing in public AI (especially with respect to building state capacity) will be synergistic with public use of simpler "AI" systems.

Support for public AI has appeared in a range of venues. In 2019, Gansky et al. called for a public research consortium in a New York Times op-ed [10]. More recently, pieces in Politico [19], Slate [29], and more [28] have argued for public AI. In the US, the proposed National Artificial Intelligence Research Resource Task Force (NAIRR) serves as an example of shared computing and data infrastructure [21]. Furthermore, an executive order on AI issued in November 2023 and the release of the "ai.gov" website signaled further support for public AI-type initiatives, including efforts by the government to hire AI talent, use AI, and support democratic inputs in AI policy. In the UK, Parliament has expressed serious interest in supporting AI-related initiatives [25].

In the coming months we expect to see discussion of public AI advance in both public forums (e.g., high profile op-eds) and in government bodies (e.g., the U.S. and U.K. lawmaking bodies). We also expect these discussions identify weaknesses of public AI, including both contexts where public systems might struggle to compete with private alternatives and contexts in which public bodies may fail to meet standards of transparency and participation. We hope that the ML community can help lead these discussions, for example by organizing concrete evidence on potential harms or by working with governments to build and govern initial pilots.

2.1 Comparison to other regulatory approaches

There are a variety of regulatory modalities that governments can already apply to AI: direct regulation that restricts the behavior of AI companies and developers, trade policy [18, 32], competition policy, tax incentives, and direct grants and contracts. Within this space of regulatory interventions, public AI has a number of distinct advantages and drawbacks.

First and foremost, the act of building and provisioning AI provides institutions with indispensable practical experience—experience that does not arrive from hands-off regulation. In this view, public AI is a direct complement to other regulatory actions insofar as it builds the expertise and capacity to regulate within public bodies. Even compared to a "heavy" regulatory regime, where a given jurisdiction regulates AI companies akin to (privately-owned) public utilities, there is no replacement for actually building AI.

Second, public AI centers a particular expertise of governments: vendor selection. Governments contract significant amounts of work to other bodies, whether private entities, nonprofits, or academic institutions (cf. research grants). While not all forms of public AI emphasize contracting (e.g. opting for a national AI agency), governments are uniquely positioned to bring together organizations and stakeholders from a wide range of industries within a shared enterprise, e.g. the COVID-19 vaccines or the human genome project. Public AI thus represents one channel for deploying resources arranged for AI development.

Third, deploying a model or service comes with its own costs and risks. Perhaps the biggest risk: public AI has to be *good*, or at least competitive with private options, or risk irrelevance—though it may still serve to develop government expertise, provide a platform for science (see Section 4.2), or help to develop standards (see Section 4.3). Note that the riskiness of adoption depends on the availability of alternatives (private, open-source, and other) within each particular jurisdiction as well as the particular form of public AI, e.g. some versions of public AI may use privately-developed or open-source models. By offering a service, public bodies may open themselves up to liability and other costs and claims (though how liability applies to some generative AI use cases remains to be decided [35]) . While governments are no strangers to service provision, some may face the cost of learning to grapple with these concerns in a new domain for the first time. Direct regulation does not directly incur these costs.

Lastly, public AI can be more participatory than direct regulation (see e.g. Delgado et al. [6] for a discussion of participatory AI). The drafting of industry regulations is typically a technical, elite activity. Public AI, by virtue of its accessibility, offers a direct line by which a citizenry can engage with and see the effects of government action. Many parties have called for "democratic input" to AI, but most proposals assume such input to the development of otherwise undemocratic private models [38]. Public AI has an enforceable form of democratic input built in, though the nature of that input depends on the governance model of the jurisdiction and on practical decisions around implementation—e.g. whether public hearings, learning from citizen interactions, or just a plain vote.

3 Open challenges for the research community

Implementing public AI poses unique technical and policy challenges; it also offers a new perspective on existing research questions within AI safety, interpretable ML, and technology governance more broadly. The research community has a critical role to play in posing and solving these challenges.

3.1 Data practices

There are many open questions that will influence how public AI institutions obtain and use data. For example: how will the privacy of information in training data be managed? Must personally identifiable data be removed, as faces were blurred in ImageNet [37]? What limits are there on objectionable content in training data? Stable Diffusion 2 removed nudity from its training data [26]. Must the intellectual property used to create public AI also be public? Efforts such as SpawningAI [1] are creating systems to allow copyright holders to remove their content from training data.

These questions may be answered based on research about data valuation and people's preferences regarding data use. Should public AI be subject to the same rules as private AI, or might it have broader rights, for instance better arguments around fair use claims that hinge on serving the public interest?

Technical methods also need to be developed to enable attribution and editing of data in response to policy choices and regulation. For example, when an improper piece of training data is identified, it may be impractical for public AI systems to be fully retrained with that data item removed, potentially at massive expense. Public AI will be a natural application for efficient methods for handling responsible data curation [22, 23, 15, 9].

Finally, there is the question of how these decisions should be made. In private organizations, there is no guarantee of consistency, transparency, or fairness; public institutions have stronger norms around this but must engage in decision-making appropriate for the fast-changing environment.

3.2 Models for public-private partnership

One of the ongoing challenges that will face public AI will be to ensure that public models employ state-of-the-art methods, even as the state-of-the-art advances quickly. Private companies are making tremendous investments in AI technology, so a public AI will be most useful if it is able to benefit from private-sector innovations.

Yet if private companies participate in the creation of public AI, we will need to ask: to what extent can they contribute while protecting their private intellectual property? There is a possible distinction between the methods and data used to train an AI and the design, parameters, and knowledge contained

within the AI itself. And participants in a public AI project may choose to contribute some methods while holding others back. Large-scale open-source projects provide one possible inspiration, where many private companies make individual decisions about which intellectual property to contribute and which to hold back as proprietary. A framework for similar collaborative engineering in public machine learning will be a key ingredient for public AI.

3.3 Engineering challenges

The same model with the same weights may be public, private, or open-source. But public AI is not just a form of private, corporate AI where we swap all the 'Anthropic' labels in the Terms of Service with 'Department of Commerce' or 'State of Pennsylvania'—nor is it just open-source AI with a public body as the controlling entity of the repository rather than a foundation or private company. For example, in the interest of accessibility, public operators may put more emphasis on offering *some* level of service rather than a consistent product offering—but how do we offer a lower-cost experience that degrades gracefully while preserving safety? Or, suppose that public bodies want to publish summary statistics in the interest of accountability, but not all model weights—what statistics should they publish, and how can we help the public interpret this technical information?

For now, many of the engineering challenges are still to be discovered. To understand the scope of these challenges, consider a comparison: open-source AI in its current incarnation requires a wide range of new technologies and services to be practically workable compared to private AI—licenses, model hubs, infrastructure for running and deploying models, community management, and more. The engineering challenges faced by a community of open-source AI contributors are not the same as those faced by a group of engineers operating within a private company because of disparities in engineering resources (data, compute, know-how, management, etc.). Nor will they be the same engineering challenges faced by the builders of public AI.

3.4 Preventing misuse

As large AI models become more capable, the risk of misuse will also grow. Highly-capable AI models could be used to create persuasive misinformation [33, 24, 39, 11, 12], create dangerous substances [30, 5, 13], or to find vulnerabilities in other systems [7, 14]. Public AI must have a mechanism to limit these types of applications in a way that honors the principle of public accessibility.

There are two natural approaches to ensure that a system is used in ways that advance the public interest. One is to regulate and monitor *usage* of the system; and the other is govern the *creation* of the system. Both present policy and research challenges; public AI emphasises the latter. This is particularly salient given the ongoing challenges of monitoring private models operating on private devices.

4 Opportunities for the research community

4.1 Research

First and foremost, support for public AI can help allocate funding and other important kinds of capital to a variety of public-interest AI research projects. A successful national implementation of public AI would operate at a far greater scale—and afford many more opportunities for research—than even the largest public funding proposals being discussed right now.

Public AI would likely be developed in partnership with academic institutions and offer teaching and research opportunities, akin to the role of CERN for experimental physics. Public AI could especially complement "talent surges" for government hiring in AI [16]. Furthermore, public AI offers an additional path to help research conducted in academic settings reach the front lines and directly contribute to public good.

Finally, concrete steps to support public AI can help lend credence to public interest research. Even for researchers that don't directly receive resources as a result of public AI programs, by legitimizing an alternative future, it may be easier to publish and share work. Creating a sense of a shared mission – with concrete progress – can motivate public interest research and help those conducting it justify their work.

4.2 Transparency

Public AI will support transparent studies of mechanisms of very large models, beyond what might be achieved with a focus on open-source (but privately steered) development [36]. Open code and open weights do not guarantee transparent access to model outputs, but publicly provisioned models can.

Currently, researchers with expertise in scientifically evaluating model behavior may not have access to the internals of proprietary models that may be crucial for developing understanding. Furthermore, there are major challenges in using traditional empirical methods (e.g. audits) to study private large models. In the wake of increased concerns around competition and safety, firms now reveal extremely minimal details about model design (does it use an ensemble? what layers of filtering and classification are applied to prompts? etc.). Carefully collected audit-style data may be misinterpreted. Public AI enables better study of core model capabilities, which is crucial for downstream studies of societal impact. Another possible advantage of public AI may be working with census bodies to develop truly representative ecologically valid evaluation procedures.

Finally, it is important to note that transparency for some models may be in direct tension with safety. For these cases, public auditing can help, but it may be the case that maximal transparency is not always desirable. Some kind of gradation of transparency could be achieved via the accountability component of public AI.

4.3 Accelerating standardization

Public AI can accelerate standardization in AI—a major focus of the recent US executive order [16]—by providing exemplar systems. Perhaps more importantly, insofar as public AI defines a floor for AI services and a "base level" of AI services in a given jurisdiction, it can prevent both common abuses of market power as well as the selective inattention experienced by Global South countries that arise when a single company controls a widely-used platform.

Successful public AI projects will also advance the development of a set of standard technological components for AI. In a world with standardized components (defined and maintained by publicly accountable standards bodies), both public institutions aiming to make scientific progress and private institutions aiming to build capable consumer-facing AI products will benefit.

Standardization can be decisive in increasing the stability of downstream product development. If an industry becomes heavily reliant on AI model outputs (e.g., suppose the medical text processing industry comes to rely on a specific AI model), they may be vulnerable to massive disruption if one or two private AI operators change their policies.

5 Conclusion

Here, we have described the public AI concept as an alternative and complement to regulation. We have focused highlighted both open challenges and reasons the research community might be excited about this approach. We refer readers to `https://publicai.network/` for more information about public AI-related initiatives.

Acknowledgements

References

[1] Spawning AI. Spawning ai faq, Oct 2023. URL `https://spawning.ai/FAQ-1`.

[2] Multiple authors. Public ai: Collaborative working document, Oct 2023. URL `https://publicai.network`.

[3] Haydn Belfield. Great British Cloud and BritGPT: the UK's AI Industrial Strategy Must Play to Our Strengths. Technical report, Labour for the Long Term,

2023. URL `https://www.labourlongterm.org/briefings/great-british-cloud-and-britgpt-the-uks-ai-industrial-strategy-must-play-to-our-strengths`.

[4] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

[5] Andres M Bran, Sam Cox, Andrew D White, and Philippe Schwaller. Chemcrow: Augmenting large-language models with chemistry tools, 2023.

[6] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice, October 2023. URL `http://arxiv.org/abs/2310.00907`. arXiv:2310.00907 [cs].

[7] Gelei Deng, Yi Liu, Víctor Mayoral-Vilches, Peng Liu, Yuekang Li, Yuan Xu, Tianwei Zhang, Yang Liu, Martin Pinzger, and Stefan Rass. Pentestgpt: An llm-empowered automatic penetration testing tool. *arXiv preprint arXiv:2308.06782*, 2023.

[8] Wilfred Dolfsma and DongBack Seo. Government policy and technological innovation—a suggested typology. *Technovation*, 33(6):173–179, June 2013. ISSN 0166-4972. doi: 10.1016/j.technovation.2013.03.011. URL `https://www.sciencedirect.com/science/article/pii/S0166497213000473`.

[9] Rohit Gandikota, Hadas Orgad, Yonatan Belinkov, Joanna Materzyńska, and David Bau. Unified concept editing in diffusion models. *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024.

[10] Ben Gansky, Michael Martin, and Ganesh Sitaraman. Opinion | Artificial Intelligence Is Too Important to Leave to Google and Facebook Alone. *The New York Times*, November 2019. ISSN 0362-4331. URL `https://www.nytimes.com/2019/11/10/opinion/artificial-intelligence-facebook-google.html`.

[11] Catherine A Gao, Frederick M Howard, Nikolay S Markov, Emma C Dyer, Siddhi Ramesh, Yuan Luo, and Alexander T Pearson. Comparing scientific abstracts generated by chatgpt to real abstracts with detectors and blinded human reviewers. *NPJ Digital Medicine*, 6(1):75, 2023.

[12] Jocelyn Gravel, Madeleine D'Amours-Gravel, and Esli Osmanlliu. Learning to fake it: limited responses and fabricated references provided by chatgpt for medical questions. *Mayo Clinic Proceedings: Digital Health*, 1(3):226–234, 2023.

[13] Taicheng Guo, Kehan Guo, Zhengwen Liang, Zhichun Guo, Nitesh V Chawla, Olaf Wiest, Xiangliang Zhang, et al. What indeed can gpt models do in chemistry? a comprehensive benchmark on eight tasks. *arXiv preprint arXiv:2305.18365*, 2023.

[14] Maanak Gupta, CharanKumar Akiri, Kshitiz Aryal, Eli Parker, and Lopamudra Praharaj. From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access*, 2023.

[15] Thomas Hartvigsen, Swami Sankaranarayanan, Hamid Palangi, Yoon Kim, and Marzyeh Ghassemi. Aging with grace: Lifelong model editing with discrete key-value adaptors. In *NeurIPS 2022 Workshop on Robustness in Sequence Modeling*, 2022.

[16] The White House. FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, October 2023. URL `https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/`.

[17] Saffron Huang and Divya Siddarth. Generative AI and the Digital Commons, March 2023. URL `https://arxiv.org/abs/2303.11074v1`.

[18] Kristina Irion. AI Regulation in the European Union and Trade Law: How Can Accountability of AI and a High Level of Consumer Protection Prevail over a Trade Discipline on Source Code? *SSRN Electronic Journal*, 2021. ISSN 1556-5068. doi: 10.2139/ssrn.3786567. URL `https://www.ssrn.com/abstract=3786567`.

[19] Charles Jennings. Opinion | There's Only One Way to Control AI: Nationalization, August 2023. URL https://www.politico.com/news/magazine/2023/08/20/its-time-to-nationalize-ai-00111862.

[20] Hanlin Li, Nicholas Vincent, Stevie Chancellor, and Brent Hecht. The Dimensions of Data Labor: A Road Map for Researchers, Activists, and Policymakers to Empower Data Producers. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, pages 1151–1161, New York, NY, USA, June 2023. Association for Computing Machinery. ISBN 9798400701924. doi: 10.1145/3593013.3594070. URL https://dl.acm.org/doi/10.1145/3593013.3594070.

[21] Shawna Lynch. New Report Details Costs and Structure of a National AI Research Resource, 2023. URL https://hai.stanford.edu/news/new-report-details-costs-and-structure-national-ai-research-resource.

[22] Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. Locating and editing factual associations in gpt. *Advances in Neural Information Processing Systems*, 35:17359–17372, 2022.

[23] Kevin Meng, Arnab Sen Sharma, Alex Andonian, Yonatan Belinkov, and David Bau. Mass editing memory in a transformer. *The Eleventh International Conference on Learning Representations (ICLR)*, 2023.

[24] Yisroel Mirsky and Wenke Lee. The creation and detection of deepfakes: A survey. *ACM Computing Surveys (CSUR)*, 54(1):1–41, 2021.

[25] UK Parliament. Governance of artificial intelligence (AI) - Committees - UK Parliament, 2023. URL https://committees.parliament.uk/work/6986/governance-of-artificial-intelligence-ai.

[26] Robin Rombach. Stable diffusion 2.0 release, Nov 2022. URL https://stability.ai/blog/stable-diffusion-v2-release.

[27] Jathan Sadowski, Salomé Viljoen, and Meredith Whittaker. Everyone should decide how their digital data are used — not just tech companies. *Nature*, 595(7866):169–171, July 2021. doi: 10.1038/d41586-021-01812-3. URL https://www.nature.com/articles/d41586-021-01812-3. tex.copyright: 2021 Nature.

[28] Bruce Schneier Sanders, Nathan E. Build AI by the People, for the People, June 2023. URL https://foreignpolicy.com/2023/06/12/ai-regulation-technology-us-china-eu-governance/.

[29] Bruce Schneier, Henry Farrell, and Nathan E. Sanders. How Artificial Intelligence Can Aid Democracy. *Slate*, April 2023. ISSN 1091-2339. URL https://slate.com/technology/2023/04/ai-public-option.html.

[30] Emily H Soice, Rafael Rocha, Kimberlee Cordova, Michael Specter, and Kevin M Esvelt. Can large language models democratize access to dual-use biotechnology? *arXiv preprint arXiv:2306.03809*, 2023.

[31] Martin Svensson and Daniel Gillblad. AI Sweden | Advancing AI in Sweden, 2023. URL https://web.archive.org/web/20230621173329/https://www.ai.se/en.

[32] Joshua Tan. How code is used. Technical report, Public Citizen, April 2023. URL https://dtalliance.org/2023/04/07/ai-trade-technical-briefing/.

[33] torzdf and andenixa. Faceswap, a tool that utilizes deep learning to recognize and swap faces in pictures and videos, 2023. URL https://github.com/deepfakes/faceswap.

[34] James Vincent. AI is entering an era of corporate control, April 2023. URL https://www.theverge.com/23667752/ai-progress-2023-report-stanford-corporate-control.

[35] Dylan Walsh. The legal issues presented by generative AI | MIT Sloan, November 2023. URL `https://mitsloan.mit.edu/ideas-made-to-matter/legal-issues-presented-generative-ai`.

[36] David Gray Widder, Sarah West, and Meredith Whittaker. Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI, August 2023. URL `https://papers.ssrn.com/abstract=4543807`.

[37] Kaiyu Yang, Jacqueline H Yau, Li Fei-Fei, Jia Deng, and Olga Russakovsky. A study of face obfuscation in imagenet. In *International Conference on Machine Learning*, pages 25313–25330. PMLR, 2022.

[38] Wojciech Zaremba, Arka Dhar, Lama Ahmad, Tyna Eloundou, Shibani Santurkar, Sandhini Agarwal, and Jade Leung. Democratic inputs to AI, 2023. URL `https://openai.com/blog/democratic-inputs-to-ai`.

[39] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. Synthetic lies: Understanding ai-generated misinformation and evaluating algorithmic and human solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2023.