**About the author**

Dean W. Ball is a Senior Program Manager at Stanford University’s Hoover Institution. He previously served as Executive Director of the Calvin Coolidge Presidential Foundation and before that worked at the Manhattan Institute for Policy Research. He has written extensively about artificial intelligence as well public sector pensions, prisoner re-entry, urban transportation infrastructure. His writing has appeared in *National Affairs*, *The Hill*, *National Review*, and *The Dispatch*, as well as on his Substack *Hyperdimensional*. He has published in academic journals on comparative political theory (*Collingwood and British Idealism* Studies, 2016) and an essay of his on AI and political theory will be published in a forthcoming volume tentatively titled *Liberalism Revisited*, to be published by Palgrave.He serves on the Board of Directors of the Alexander Hamilton Institute for the Study of Western Civilization and on the Advisory Council of Purdue University’s Krach Institute for Tech Diplomacy. He will be joining George Mason University’s Mercatus Center as a Research Fellow focused on AI in April 2024. He graduated *magna cum laude* with a B.A. in History and Government from Hamilton College in 2014. This comment expresses the author’s personal beliefs and is not written on behalf of, or represent the views of, any of the organizations mentioned here.

The National Telecommunications and Information Administration (NTIA) has raised important questions about the risks and benefits of open-source artificial intelligence (AI) foundation models, also referred to as “foundation models with widely available weights.”

Considering these questions, it is critical to recognize the vital role that open-source software has played and continues to play in the development of technology, the digital economy, and society at large. Open-source software has been instrumental in the creation of the Internet and the digital economy as we know it today because it enables faster innovation, more robust and secure code through community input, and the flexibility to adapt tools to specific needs without vendor lock-in.

Open-source AI is also important for advancing research into AI safety, alignment, and interpretability. Open source enables deep access to the inner workings of models, which is essential for interpretability and safety research. For example, [Representation Engineering](https://arxiv.org/pdf/2310.01405.pdf), arguably the most important mechanistic interpretability breakthrough of 2023, would not have been possible if the researchers did not have access to model weights. Further breakthroughs are necessary for the National Institute of Standards and Technology (and other organizations) to develop robust technical standards for AI models, and such standards are critical for crafting reasonable law and regulation related to AI models. In this sense, open-source AI is a key part of how we will collectively get AI regulation right.

**Defining Open Source**

The traditional definition of open-source software breaks down somewhat for AI models, because unlike traditional software, an AI model includes weights and training data in addition to code. It is useful, therefore, to articulate different categories of open models: “fully open-source models,” such as the Allen Institute for AI’s OLMo or Eleuther AI’s models, are models with data, code, and weights fully available, and “open weight” or “permissible usage models,” which only make weights publicly available or impose license restrictions. In general, fully open-source models offer the greatest benefit for researchers because the access to the underlying code and data allows for full auditing of the system. Open-weight models are still useful to researchers, however, and offer roughly same benefits to the private sector as open-source models. An emerging best practice, therefore, might be to release less capable models as fully open source, and models closer to the frontier (or at the frontier) as open weight.

**Timeframe for Open-Source Models to Match the Frontier**

Based on improvements in computing power and its reduced cost over time, as well as improvements to dataset curation and model architecture, it is reasonable to expect that open-source models will reach the level of the current frontier with a lag of approximately18-36 months. However, it is difficult to predict exact timelines with any certainty, but generally reasonable to expect that open-source developers will eventually match the frontier with a significant lag time.

**Benefits and Risks of Open-Source Models**

Open-source AI will be critical for broad diffusion of the technology's benefits, democratizing access and enabling more people and organizations globally to take advantage of AI productively. For companies, open-source AI offers the ability to exercise fine-grained control over their models, ensure consistent model performance, integrate models with an existing set of technologies, and avoid vendor lock-in. AI researchers have used open-source models to advance the field, creating more interpretable, safe and aligned systems. Closed models with API access are an important part of the AI ecosystem, but they do not offer the same economic or scientific benefits.

Open-source AI is also crucial to ensure competition and dynamism in the AI industry. The AI field is replete with examples of single researchers or small teams making radical improvements and innovations. If open-source AI is to be disfavored or banned, we will lose the benefit of those innovations, particularly if they are disruptive to the incumbent firms’ business models.

Some argue open-source AI introduces unacceptable risks by allowing bad actors to remove safeguards created by the model maker. While this is a valid concern, determined malicious users will find ways to abuse AI systems regardless of whether they are open or closed source. In addition, removing safeguards is non-trivial to perform, and many malicious uses could be achieved with less technical work using a closed model. Though open access to weights does create slightly higher risk, the risk is worth it for the many benefits open-source AI can provide. Indeed, even today many cyberattacks are perpetrated using (non-AI) open-source software, but the benefits of open-source significantly outweigh the downsides. The same dynamic will apply with open-source AI.

The solution to the risks is not to restrict openness in AI development, but to have strong laws and enforcement governing illicit applications of AI and other digital technologies. Emerging best practices like "know your customer" requirements for AI model hosting services can help in this regard, as can clear technical standards for foundation models.

Openness and transparency in AI development will be essential for properly governing the technology and ensuring accountability. If AI development happens entirely behind closed doors, it will be much harder for policymakers and the public to assess systems' true capabilities, limitations, biases and failure modes. Insight gleaned from open models can inform evidence-based AI regulations.

Open-source software exemplifies key values that have made the U.S. a leader in technology: openness, ingenuity, dynamism and productive exchange of knowledge. As we enter an "AI arms race" period, open-source AI will be a key strategic advantage–it will allow the U.S. to harness the creativity of far more developers and organizations in advancing the technology responsibly. While we must remain vigilant and proactive in mitigating risks, adopting an overly restrictive approach to open-source AI would not only undermine the significant benefits, but could put the U.S. at a disadvantage globally as other countries with more open models race ahead. Sensible AI policy should preserve the ability to develop open models, while governing downstream uses and applications.