COMMENTARY ON THE RFC NTIA–2023–0009

RFC COMMENT

A U.S. Citizen

March 26, 2024

ABSTRACT

While there are detailed technical responses to be made to the questions posed in RFC NTIA-2023-0009, the core issues are political and economic. Who do we want to be as Americans, and how do we want to encourage and reward economic competition in the marketplace? Though I lead AI developments as an executive and have products, publications, and patents in AI and neural nets stretching over 25 years back to my Ph.D. thesis, in this response I focus on these core issues rather than the technical ones. Regulation of openly-available weights for foundation models is directly antithetical to core American values.

1 Directed Response to the Questions

1. Lawmakers should make no attempt to define "open" or "widely available" for foundation weights. Any such definition will lead to a morass of ambiguity and conflict, both technically and with respect to First Amendment rights. Any such definition will also rapidly be overtaken by events and innovation.
2. The risks of making model weights widely available compared to keeping them non-public are the risks associated with language and speech generation; namely, they are identical to the risks in accepting that "Congress shall make no law...abridging the freedom of speech."
3. The benefits of having foundation models with widely available model weights are to encourage innovation by non-established players, to support Americans in their Fourth Amendment rights to be "secure in their persons, houses, papers, and effects" by giving them private options for their language generation activities, to support Americans in their First Amendment rights, and to establish and preserve options for on-premesis security for sensitive materials, including national security materials.
4. An open foundation model that demonstrated physical capability to withstand human control (i.e. physical protection of its own power supplies, physical interference with human-directed input) would significantly increase risks associated with foundation models. It is my judgment that this risk is vastly higher in closed-source developments than open-source ones. Should the U.S. Federal Government see this risk being realized, the best mechanism is open democratic discussion of the real issue around autonomous physically-instantiated agents–not the red herring of foundational language models.
5. "Dual-use" is a laughable term for a foundation model. Is mathematics dual-use? Windows 11? The Encyclopedia Brittanica? All enable dual-use threats more readily than a foundation model. Foundation models' dual-use threats are highly diffuse and nonspecific.
6. The primary business risks to date are around potential security flaws in open models that might allow an attacker to leak data. These risks are probably lesser than the same risks for closed models served from remote servers. Many businesses also have concerns around potential IP litigation around the training data used in the foundation models. (This is where I urge clarity from the U.S. Federal Government–clarifying the role of copyright and fair use, as has been done in Japan.)
7. The mechanisms for managing risks around foundation models are the same as the mechanisms for managing risks around free speech: if you are concerned about misinformation, then provide more credible, more reliable, more coherent, better true information.

8. Thinking about managing the future is sometimes best tested against the past. What if we as a nation had regulated RNN's (state of the art 10 years ago) in 2014? The fact that that regulation would, at best, be useless today, and at worst, would have prevented the innovations of the past 10 years, should be borne in mind. The attention methods powering today's transformers were invented and published by [Bahdanau et al 2014], an open source implementation and description of the attention mechanism (before it was called that) from university research groups.

9. When considering open-source weights, these additional important things should be borne in mind.

   - The open-source community powers all flavors of Linux, which run much of the world's most important code already. We as a nation have already learned how to live with such risks.
   - The information available to, and made available by, publicly available foundation models is, almost by definition, already publicly available–this is how it made it into the training data. We as a nation have already learned to live with these risks.
   - The RFC mentions "risks to security, equity, civil rights, or other harms due to, for instance,affirmative misuse, failures of effective oversight, or lack of clear accountability mechanisms". In every one of these cases, there is a simple principle for accountability and regulation from a U.S. Federal Government perspective: "The Cossacks work for the Czar." Those humans employing foundation models to carry out harmful actions are responsible for the consequences of those actions, just as if they had carried out those actions with any other tool. It is the responsibility of those employing algorithms of any sort to mitigate their potential harms–there must be no "get out of jail free" card by pointing at an algorithm.
   - The most important closed-source foundation models today–from Meta, OpenAI, Google, Mistral, and Anthropic–are the domain of, at most, a couple of thousand researchers. To the extent that great-power competition is a part of any thinking in this arena, unilaterally disbarring all but a few thousand Americans from contributing to developments sounds like a mistake.
   - Closed-source actors in the economy naturally wish to foreclose competition, and are notably those most vocal in calling for regulation.