



Thank you Leader Schumer, Senator Rounds, Senator Heinrich, and Senator Young for inviting me to join this AI Insight Forum on Transparency, Explainability, Intellectual Property, & Copyright.

My name is Ali Farhadi, and I am the CEO of the [Allen Institute for AI](#)- also known as AI2. We are one of the world’s largest nonprofit AI research institutes founded in 2014 by the Allen Family. Our mission at AI2 is to conduct high-impact artificial intelligence research and engineering in service of the common good, and to develop applications of AI that contribute to humanity and the preservation of the environment. Our organization is overseen by an experienced, diverse Board of Directors and expert Science Advisory Board who support our goals of advancing AI for humanity. We are driven by impact, not profit.

AI has myriad applications that will benefit society, including accelerating research into medical diagnosis, treatments, and cures; assistive technologies for people with disabilities; intelligent tutoring systems for personalized and more equitable education; climate modeling to predict impacts in specific regions; and conservation to preserve our wildlife and natural resources. We also recognize the inherent and potential challenges that exist with this technology. Our focus at AI2 is to work not only on cutting edge AI research and applied science, but also at the intersection of AI ethics, AI policy, and AI literacy to create solutions that enable a future where AI is designed, developed, and deployed safely and responsibly.

In addition to serving as the CEO of AI2, I have been a Professor of Computer Science at the University of Washington for over a decade, and I co-founded xnor.ai, an AI startup acquired by Apple (where I went on to lead the Machine Intelligence team) in 2019. It is with this wide-ranging insight as a nonprofit, academic, startup, and industry AI executive that I offer the following comments on transparency, explainability, intellectual property, and copyright on behalf of AI2.

It’s important to ground our perspective by clarifying what we mean by a handful of key terms that permeate AI conversations today¹. A broad debate is raging between proponents of “closed

¹ This comment includes reference to AI Models and AI Systems. For the purposes of this comment, we use the definitions from U.S. Copyright Office Notice of Inquiry and Request for Comments Docket No. 2023-6.

AI Model: A combination of computer code and numerical values (or “weights,” which is defined below) that is designed to accomplish a specific task. For example, an AI model may be designed to predict the next word or word fragment in a body of text. Examples of AI models include GPT-4, Stable Diffusion, and LLaMA.

AI System: A software product or service that substantially incorporates one or more AI models and is designed for use by an end-user. An AI system may be created by a developer of an AI model, or it may incorporate one or more AI models developed by third parties.

models” and those who advocate for “open models,” but what these terms mean in the context of AI is murky.

We see open and closed as a spectrum. Along that spectrum, models are more open if you can access the components, inspect and modify the model, and there is transparency around how it was built. We consider a model “open” when it is released in full, including the training data, model weights, training configurations, evaluations, and documentation. When possible, we release open models and datasets to the public under a responsible AI license, the [AI2 ImpACT license](#), which includes use-based restrictions and a risk assessment that determines rights and responsibilities. All of our research at AI2 is published in open scientific forums.

We consider “closed” models to be those released without information that is key to understanding how the models were made or how they work, or that more generally prohibits access to their core components.

What complicates this discussion is that when many people advocate for closed models, they are doing so because they believe closed models are inherently safer than open models. I would like to unequivocally state that **closed and safe are not synonyms**. Transparency is at the core of all research into trust, safety and explainability in AI, and **open models are essential for transparency**. The road to safety is absolutely not paved with models that are closed off from public and expert scrutiny.

Transparency

Transparency is one of AI2’s four core values, along with impact, collaboration, and accountability. Our team has been a role model for transparently conducting AI research since our founding.

A transparent model has two key elements: 1) its training components and process are fully and publicly disclosed, and 2) it can be comprehensively and scientifically evaluated. Only when this is achieved can we spur innovation and establish trust and safety. Closed models fall short of these criteria. Without access to the full stack of a model, the research community is operating in the dark.

Transparency spurs innovation and is a building block for trust and safety

As the [NIST AI Risk Management Framework outlines](#), transparency and safety are characteristics of trustworthy AI. AI evolved in an open and transparent environment with new research building on and improving previous work. Below I highlight some of AI2’s open contributions. It’s important to note that AI2 is one of many advocates for open and transparent model development, and our work is built on other openly published work.

Our release of [ELMo](#) in 2018 is a prime example of this concept. ELMo was one of the first language models to train on large web corpora. It was the first language model trained in an

unsupervised manner to be efficiently adapted to solve many language tasks; large models we know of today build on this foundational idea. The release of this model was radically transparent; it consisted of a detailed research paper along with open software. This included sharing the “recipe” of the model, the evaluation results detailing strengths and weaknesses, model weights, and code. Collectively, this comprehensive and transparent release ultimately allowed people to build substantially better language models by improving upon the limitations of ELMo. This is a prime example of how open and transparent research leads to breakthroughs with [broad benefit to the research community](#).

Another example is the [COVID-19 Open Research Dataset \(CORD-19\)](#). As the world began to learn about a new coronavirus in early 2020, our team at AI2, at the direction of the White House and in partnership with the National Institutes of Health and other leading research groups, created a machine-readable corpus of all available scientific literature directly relevant to COVID-19 so research communities could develop and use cutting-edge techniques to obtain insights on the disease, potential treatments, and paths towards a vaccine. This dataset inspired the creation of countless tools across industries allowing for rapid response to an immediate crisis.

In 2019, AI2 put forth a paper called [Green AI](#) coining the term which sparked great interest in the carbon footprint of AI models. The vision of Green AI raises many exciting research directions that help to overcome the challenges of inclusiveness, efficiency, and climate impacts of training large models. Building on this line of research, we created a partnership with other academic and industry researchers to [measure and report on software carbon intensity](#), a first step in minimizing emissions. Breakthroughs like this cause a ripple effect within the research community leading to important collaborations, meaningful, practical tools, and greater awareness of current challenges and limitations of AI.

Understanding what is in a dataset is an essential part of transparency. Our team has pioneered the creation, release, and analysis of web scale datasets. One of our latest projects, [What's In My Big Data](#), introduces a new platform to reveal and compare the contents of large text corpora and unveils significant, novel findings about the type and quality of training data that powers the large language models we all use today. Our team also recently released [Dolma](#), the largest open dataset for language model training created from a diverse mix of web content, academic publications, code, books, and encyclopedic materials. Dolma is one of the resources AI2 has committed to the National AI Research Resource (NAIRR) pilot in support of their need for responsible and trustworthy AI resources.

Additionally, transparency is at the root of our focus on [AI literacy](#), an often missing but essential part of the current AI conversation. Equipping the public with knowledge about AI- its capabilities, its potential, its strengths, and its weaknesses- is overdue.

Explainability

Transparency is a building block for explainability. This is a nascent, active research area on which our team at AI2 is actively working. Explainability is a field of research that seeks to answer why a model behaves a certain way or generates a certain output, and **we cannot make progress on this research with closed models**. Creating regulation around explainability is premature. We don't yet know why AI systems make certain decisions, but we can accelerate work in this direction by incentivizing openness and transparency.

Copyright and Intellectual Property

Acknowledging that there are other IP issues with AI, such as patentability of AI generated works, our primary focus based on what we do is copyright. [As we shared in our comments in response to the U.S. Copyright Office's Artificial Intelligence Study](#), we share the widely held belief that use of copyrighted training materials in training datasets constitutes, at minimum, fair use. We believe an infringement and fair use analysis should focus only on the outputs of an AI system rather than the training material of the model. The value of training data is not in its creative expression, but rather in its volume and diversity. The most capable AI models require billions of items of material to be effectively trained. Such large scale datasets can only be assembled through large-scale data collection efforts, such as scraping of web content and digitalization of media. After training material is acquired, several steps are required in order to make it suitable for training. Because of the numerous filtering and transformation steps, training datasets do not represent a substitute for human consumption in a competitive way.

We acknowledge the possibility that training on copyrighted data may result in infringing output, and here are some observations:

- Open models provide the foundation to study relationships between input and output in a scientific way; however
- The law as it currently stands disincentivizes openness and transparency because there are no safe harbor or negligence standards, and opening up the AI development lifecycle can create the appearance of a violation (e.g., someone may see their content has been inadvertently used as training data).

Thoughtful regulation is therefore essential to establishing the right incentives to design, develop, and deploy AI in a transparent, safe, and responsible way.

Closing Thoughts

This body has the unique challenge of regulating to both mitigate harm and foster innovation. To accomplish this, we offer the following recommendations:

- 1) **Invest in public infrastructure:** [As NSF Director Sethuraman Panchanathan said](#), "A critical element to the success of the AI research revolution is ensuring that researchers have access to the data and platforms required to continue to drive innovation and

scalability in AI technologies and systems." Building on this, we specifically mean investments in compute, data, and most importantly evaluation. The recent *Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence* is a step in the right direction with its focus on the NAIRR, but the US is falling behind in investment to close the widening gap between research and industry and maintain global leadership in AI.

- 2) **Incentivize open and safe model development:** Open and transparent model development is how the United States has established itself as the leader in AI. Without incentives to keep the field open, such as reasonable protections for model developers, we risk weakening the strong AI ecosystem we have developed over the previous decades, including the foundations for making models safer and more accountable.
- 3) **Promote AI literacy:** As it becomes increasingly common for the public to interact with AI systems in their daily lives, it is imperative that we improve people's understanding of this technology. Similarly, we have a responsibility to develop the next generation of AI researchers and developers. Without a dedicated focus on AI education, we are missing a critical ingredient for a successful AI future.

Together these recommendations support cementing the U.S.'s position as a leader in AI research and development. As you move forward with the difficult task of creating lasting AI policy, AI2 can be helpful in two important ways: 1) surfacing open research questions so policy does not result in unachievable mandates and 2) providing deep technical expertise to support the creation of a truly open and safe AI ecosystem.

I appreciate the opportunity to share these thoughts on behalf of a research-first nonprofit institute that believes in an ethical, interdisciplinary, science-focused approach free from any profit motive, and I extend an offer to be of future assistance, as needed.