Written Testimony of Clement Delangue
Co-founder and CEO
Hugging Face

For a hearing on
"Artificial Intelligence: Advancing Innovation Towards the National Interest"

Before the
Committee on Science, Space, and Technology
U.S. House of Representatives
June 22, 2023

Introduction

Chairman Lucas, Ranking Member Lofgren, and Members of the Committee, thank you for the opportunity to discuss AI innovation with you in a critical time for AI and national interests. I deeply appreciate the work you are doing to advance and guide AI innovation in the U.S. and look forward to supporting this mission. My name is Clement Delangue, and I am the co-founder and CEO of Hugging Face.

Hugging Face is a community-oriented company based in the U.S. with the mission to democratize good machine learning. We conduct our mission primarily through open-source and open-science, with our platform for hosting machine learning models and datasets and an infrastructure that supports research and resources to lower the barrier for all backgrounds to contribute to AI.

In this hearing, I will share the importance of openness in innovation, mechanisms for safe and transparent AI, and the necessary investments across sectors to ensure AI is both developed in the national interest and the U.S. continues its technological leadership.

Foster Safe Innovation via Access and Collaboration

AI innovation especially for popular AI systems today such as ChatGPT has been heavily influenced by open research, from the foundational work on the transformers architecture to open releases of some of the most popular language models today. Making AI more open and accessible, including machine learning models, the datasets used for training, in addition to research breakthroughs, cultivates safe innovation. **Broadening access to artifacts such as models and training datasets allows researchers and users to better understand systems, conduct audits, mitigate risks, and find high value applications.**

The tensions of whether to fully open or fully close an AI system grapples with risks on either end; fully closed systems are often inaccessible to researchers, auditors, and democratic

institutions and can therefore obscure necessary information or illegal and harmful data. A fully open system with broader access can attract malicious actors. All systems regardless of access can be misused and require risk mitigation measures. Our approach to ethical openness acknowledges these tensions and combines institutional policies, such as documentation; technical safeguards, such as gating access to artifacts; and community safeguards, such as community moderation. We hold ourselves accountable to prioritizing and documenting our ethical work throughout all stages of AI research and development.

Open systems foster democratic governance and increased access, especially to researchers, and can help to solve critical security concerns by enabling and empowering safety research. For example, the popular research on watermarking large language models by University of Maryland researchers was conducted using OPT, an open-source language model developed and released by Meta. Watermarking is an increasingly popular safeguard for AI detection and openness enables safety research via access. Open research helps us understand these techniques' robustness and accessible tooling, which we worked on with the University of Maryland researchers, and can encourage other researchers to test and improve safety techniques. Open systems can be more compliant with AI regulation than their closed counterparts; a recent Stanford University study assessed foundation model compliance with the EU AI Act and found while many model providers only score less than 25%, such as AI21 Labs, Aleph Alpha, and Anthropic, Hugging Face's BigScience was the only model provider to score above 75%. Another organization centered on openness, EleutherAI, scored highest on disclosure requirements. **Openness bolsters transparency and enables external scrutiny.**

The AI field is currently dominated by a few high-resource organizations who give limited or no open access to novel AI systems, including those based on open research. **In order to encourage competition and increase AI economic opportunity, we should enable access for many people to contribute to increasing the breadth of AI progress across useful applications, not just allow a select few organizations to improve the depth of more capable models.**

Working with different professions to ensure AI augments workers and improves productivity requires increased access to AI systems and community contributions to determine AI utility through experimentation, adaptation, and fine-tuning. Shared resources can be built and improved over time with multidisciplinary expertise, and leverage the country's public institutions so they reflect democratic values. Hugging Face is particularly invested in the role of machine learning libraries in curating the content of AI systems.

Mandates and Mechanisms for Transparency and Safeguards

All AI systems, regardless of their level of openness, require mechanisms for better understanding and communicating aspects of the system. This includes detailed and easily consumable documentation and better risk evaluations combined with increased research for technical and policy safeguards.

Rigorous documentation practices for AI systems, with transparent reporting that follows well-defined protocols, serves three main goals: incentivizing responsible development; ensuring researchers and developers consider values and priorities that may otherwise be overlooked; and creating a paper trail for review. Documentation highlighting heavy risks disincentivizes developers from further development in that unsafe direction. Guidance for what to document, such as funding sources, can aid scrutiny across multiple dimensions. Accessible documentation that is readable to many audiences and able to be examined by additional parties further down the pipeline can improve understanding of the technology as a whole, including tradeoffs in benefits and risks.

Importantly, documentation should be detailed and applied to AI models, datasets, and all relevant components of an AI system. Hugging Face leads by example in documenting models on our platform with over 250,000 model cards, a vastly adopted structure for documentation that can be a strong template for AI models. Datasheets are the parallel for datasets and are also widely adopted. Documentation should also cover processes and governance mechanisms, which we exemplified in a Governance Card for the BigCode initiative we are co-leading with ServiceNow to create an open-source generative code model. We also lower the barrier for users to create documentation with easy tooling that generates documentation.

In order to document and mitigate risks, we need to understand how to measure them. The current state of AI evaluations, especially for complex social impacts such as biases and environmental costs, requires more resourcing and central fora for testing, as well as transparency from entities about how and where they deploy AI systems to understand what evaluations are most urgently needed. Risk evaluation and mitigation work should be built on technical feasibility and existing research. National Institute for Standards and Technology's (NIST) work on moving toward a standard for AI bias and the AI risk management framework have been excellent examples of technical leadership on safe AI innovation in the U.S. government and guidance for the AI ecosystem. **Increased funding for NIST would both strengthen U.S. government technical leadership and improve the space for researchers across sectors to collaborate on addressing AI risks.**

Safeguards should be developed in tandem with AI systems, and should be a combination of technical, policy, and legal approaches. Technical approaches include detection models, such as the GPT-2 output detector hosted on our Hub, and safety filters, such as the one deployed with Stable Diffusion. Organizational and platform policies should be specific to content and intended system use; we enable researchers to manage access to sensitive systems to promote accountability without facilitating malicious use; in addition to labeling and moderating content that is not for all audiences. And novel legal approaches, such as Open Responsible AI Licenses (RAIL) can provide enforcement mechanisms for abiding by intended uses.

Invest in Infrastructure and Research

AI innovation across many high value applications requires investing in relevant expertises and accounting for resource imbalances among disciplines and sectors. The long-standing and widening resource divides especially between industry and academia limit who is able to contribute to innovative research and applications. **We strongly support the U.S. National AI Research Resource (NAIRR) and resourcing small businesses and startups conducting public interest research.** We recommend incorporating AI system access needs with infrastructure access.

Training and educational resources such as courses and tooling can encourage interdisciplinary research and more engagement from untapped sectors. This can aid professions interested in leveraging AI in conducting complex tasks, such as training an AI model. Our Evaluate library provides tutorials and no-code interfaces to run technical evaluations on our hosted models. AutoTrain empowers anyone regardless of technical skill to train, evaluate, and deploy a natural language processing (NLP) model. Hugging Face Tasks and Spaces enable anyone to build and engage with tasks.

Global leadership in democratic values should include our international allies. Our platform encourages many groups to curate datasets and evaluations in their native or fluent languages, leading to stronger, representative science. **Tooling and resources should be made available across disciplines and accessible to the many languages and perspectives needed to drive safe innovation.**

Furthermore, while high-level guidance is needed to converge the AI community toward similar practices, **government-provided resources are most easily applied when tailored to specific systems**. We recommend increasing NIST capacity to develop AI RMF profiles for popularly used systems, such as language models. We are excited about federal support of data commons and infrastructure, such as current efforts to make NSF funded data more broadly available. **Technical experts with a track record of ethical innovation should be prioritized as advisors for resourcing initiatives.**

Conclusion

The incredible capability and promise for AI to improve American lives and opportunities requires cross-sectoral collaboration, access to systems, and investing in safety. We thank the Committee for the opportunity to discuss this important issue and look forward to continuing to support a safe, innovative U.S. AI ecosystem.