Hugging Face Comments on the UK AI Regulation White Paper

Hugging Face

21 June 2023

Hugging Face congratulates the UK government on its pro-innovation approach to AI regulation that recognizes the many benefits and opportunities of AI while controlling for risks. The following comments are informed by our experiences as an open platform for state-of-the-art (SotA) AI systems, working to make AI accessible and broadly available to researchers for responsible development. Comments are organized by questions listed in Annex C of the White Paper. If a section is not highlighted, we do not have specific, actionable feedback.

About Hugging Face

Hugging Face is a community-oriented company working to democratize good Machine Learning (ML), and has become the most widely used platform for sharing and collaborating on ML systems. We are an open-source and open-science platform hosting machine learning models and datasets within an infrastructure that supports easily processing and analyzing them; conducting novel AI research; and providing educational resources, courses, and tooling to lower the barrier for all backgrounds to contribute to AI. Hugging Face is based in the U.S. and France, with an office in London and a global developer community.

The Revised Cross-Sectoral AI Principles

1. Do you agree that requiring organisations to make it clear when they are using AI would adequately ensure transparency?

Making AI use clear requires guidelines and specificity on how to document and communicate aspects of an AI system and its use. **Transparency and disclosure should address many AI system components throughout its lifecycle in addition to the contexts and applications into which the system is deployed.**

This means not only robust documentation for models and datasets, but also for processes throughout systems development and testing. Guidance on how to document systems, such as through model cards, which we deploy widely at Hugging Face, in addition to tooling, can help lower the barrier for transparency. Approaches to transparency will necessarily differ by system. For large language models, their complex and multi-purpose capabilities in addition to large architecture make them difficult to make fully transparent.

2. What other transparency measures would be appropriate, if any?

Mechanisms for transparency require many skill sets, not all of which are likely present in a developer organization. There is heavy overlap with accountability mechanisms such as audits and certifications. Both require access to AI systems and their components, meaning detailed model and dataset documentation at the very least. Openness can help improve transparency, and organizations dedicated to openness show exceptional transparency compliance. **Pairing transparency requirements, such as documentation, with accountability, such as audits, can bolster trustworthiness by ensuring third-party validation**.

3. Do you agree that current routes to contestability or redress for AI-related harms are adequate? 4. How could routes to contestability or redress for AI-related harms be improved, if at all? 5. Do you agree that, when implemented effectively, the revised cross-sectoral principles will cover the risks posed by AI technologies?

Better evaluations and government-provided research resources are sorely needed for novel AI systems. In order to conduct adequate pre-deployment risk assessments, those evaluating a system need good tools and metrics. Foundation model evaluations, especially outside text and language modalities, face many pitfalls such as lack of standardization, lack of access to systems and necessary computing infrastructure, and lack of existing tools. Approaches by system, such as language models, can provide helpful insights across models and capabilities. Communicating evaluation findings should also be consumable to many audiences. Furthermore, evaluating inherently qualitative and social aspects of a system such as harmful biases, disinformation, and unsafe or violent content is difficult. **More investment in evaluation is needed for models, datasets and other system components.**

A central resource for researchers to access infrastructure such as computing power can increase research on evaluations and safeguards. The many expertises needed to evaluate and mitigate risks may also require computer science and similar technical training in addition to low to no-code tooling. Lessons from the U.S. National AI Research Resource can strengthen global approaches to safe innovation.

6. What, if anything, is missing from the revised principles?

The given principles can be overarching and inclusive to account for social impacts such as privacy and data protection and environmental impacts. Existing AI principles, such as the OECD's AI Principles, EU's requirements for Trustworthy AI, and the U.S. AI Bill of Rights should overlap with each other to bolster a global and allied approach to safe AI. **These principles should be dynamic and updatable as new information about AI risks arise.**

Monitoring and evaluation of the framework

15. Do you agree with our overall approach to monitoring and evaluation? 16. What is the best way to measure the impact of our framework?

Monitoring both the framework and the framework's effectiveness addressing AI risks should also invest in capabilities evaluations, risk taxonomies, and expertise across systems and high risk areas. We commend the given approach and emphasize interoperability with global frameworks to avoid patchwork legislation, while recognizing some tensions may arise in specific approaches such as legal definitions. **Horizon scanning should be agile**; while some risks may be long-standing, such as disinformation from large language models, others may quickly arise, such as AI generations' impact on academic integrity.

Impact should be measured iteratively and in conjunction with our understanding of AI capability and innovation and updating emergent risks.

17. Do you agree that our approach strikes the right balance between supporting AI innovation; addressing known, prioritised risks; and future-proofing the AI regulation framework? 18. Do you agree that regulators are best placed to apply the principles and government is best placed to provide oversight and deliver central functions?

We agree with and applaud highlighting feedback loops, which should be inclusive of all stakeholders, such as regulators, industry, civil society, and academia, and global allies. **Regulators should help guide innovation in a beneficial direction.** The appropriate regulatory body and agency giving guidance will differ based on the type of system, the system's sectoral application and use case, and the urgency of attention/level of risk.

Tools for trustworthy AI

21. Which non-regulatory tools for trustworthy AI would most help organisations to embed the AI regulation principles into existing business processes?

The most effective approaches to trustworthiness are injected throughout system development and address the system's context and application as it is deployed and affects users. For increasingly general-purpose systems, mechanisms can be applied by system and function; novel legal approaches such as Responsible AI Licenses (RAIL) can encourage innovation while preventing harmful uses.

Foundation models and the regulatory framework

F1. What specific challenges will foundation models such as large language models (LLMs) or open-source models pose for regulators trying to determine legal responsibility for AI outcomes?

The continually evolving landscape of foundation model development and deployment makes determining outcomes and ongoing processes. The options for releasing foundational models vary across a spectrum from fully closed to fully open, each option with its own challenges and tradeoffs. Open-source provides many opportunities for broader community research, including empowering researchers to create safeguards by being able to test on an accessible model. Ethical openness requires implementing many types of safeguards. Legal responsibilities will depend on the type of impact and legal precedent. **Existing frameworks and risk considerations work should be examined and expanded.**

F2. Do you agree that measuring compute provides a potential tool that could be considered as part of the governance of foundation models?

Compute needs differ vastly by researcher and type of research. Compute is more tangible than more complex infrastructural needs, such as clean and safe training data. **Necessary research and development infrastructure, and gaps, extend beyond compute** and governance mechanisms should encompass many system components. Furthermore, the development of more compute-efficient models as seen with LLaMA and Alpaca, point to innovation in compute-constrained environments. We also note that most of the risks inherent to AI are tied to the scale of impact on increasingly global populations; broader and more direct reach of a system requires stronger governance requirements than absolute compute needs.

F3. Are there other approaches to governing foundation models that would be more effective?

Better fora for developing community norms must be inclusive of the many developers, providers, researchers across academic disciplines, and stakeholders in AI. Community norms vary from appropriate release methods to shared safety protocols to best evaluations per modality. **Fostering innovation requires supporting small and medium businesses and researchers' needs for access and technical infrastructure. Fostering safe innovation requires investing in technical, policy, and legal safeguards that preempt and protect from emerging risks.**

Conclusion

We thank the UK government for the opportunity to provide feedback and acknowledge each of these proposed questions require continual input as the pace of AI continues to accelerate. We look forward to further discussions and to supporting a regulatory approach that will foster safe innovation.

Respectfully,

Irene Solaiman Policy Director Hugging Face 🤗	Yacine Jernite ML and Society Lead Hugging Face 🤗