



**Hugging Face**  
29 June 2022

20 Jay St  
Suite 620  
New York, NY 11201

## **Hugging Face Comments on Implementing Findings from the National Artificial Intelligence Research Resource Task Force**

Hugging Face commends the National Artificial Intelligence Research Resource (NAIRR) Task Force on its interim report and we offer recommendations to further shape innovation for good, responsible artificial intelligence (AI). 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 Interim Report Chapter, with more granular recommendations italicized below. If a section or chapter is not highlighted, we do not have specific, actionable feedback.

### **About Hugging Face**

Hugging Face is a community-oriented company based in the U.S. and France working to democratize good machine learning. 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.

#### **a. Vision for the NAIRR (Chapter 2)**

##### *Appoint Technical and Ethical Experts as Advisors*

Technical experts with a track record of ethical innovation should be prioritized as advisors. In order for NAIRR to drive innovation in a responsible direction, it must craft a diverse external advisory body with interdisciplinary expertise. As part of Recommendation 2-3, technical and ethical experts can calibrate NAIRR on not only what is technically feasible, implementable, and necessary for SotA systems, but also on how to avoid exacerbating harmful biases and other malicious uses of AI systems. [Dr. Margaret Mitchell](#), one of the most prominent technical experts and ethics practitioners in the AI field and Hugging Face's Chief Ethics Scientist, is a natural example of an external advisor.

#### **c. NAIRR resource elements and capabilities (Chapter 4)**

##### *Resource (Model and Data) Documentation Standards*

NAIRR-provided standards and templates for system and dataset documentation will ease accessibility and function as a checklist. This standardization should ensure readability across audiences and backgrounds; documentation should be robust for researcher and developer information, clearly have examined and reported ethical considerations, and be easily consumable for a nontechnical audience. [Model Cards](#) are a vastly adopted structure for documentation that can be a strong template for AI models. [Datasheets](#) are the strong parallel, also widely adopted but for datasets.

### *Making ML Accessible to Interdisciplinary, Non-Technical Experts*

In addition to being user-friendly, NAIRR should work toward encompassing the many critical expertises in AI that may not have the advanced technical knowledge to leverage the provided resources. Combining recommendations 4-20, 4-23, 4-24, and 4-26, NAIRR should provide education resources as well as easily understandable interfaces and low- or no-code tools for all relevant experts to conduct complex tasks, such as training an AI model. For example, Hugging Face's [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, from prompting the image-generative model [Dall-E Mini](#) to a [Wikipedia Assistant](#) that can answer open-ended questions based on Wikipedia content.

#### **d. System security and user access controls (Chapter 5)**

##### *Guardrails for Open Source Science: Monitor for High Misuse and Malicious Use Potential*

In addition to cybersecurity, a key aspect of preventing harm from dual-use AI systems and resources is to evaluate and monitor for high potential for harm. Datasets and models created with the intent to harm or that overrepresent harmful content should be closely monitored or gated to prevent bad actor access. Harm must be defined by NAIRR and advisors and continually updated, but should encompass egregious and harmful biases, political disinformation, and hate speech. NAIRR should also invest in legal expertise to craft [Responsible AI Licenses](#) to take action should an actor misuse resources.

#### **e. Privacy, civil rights, and civil liberties requirements (Chapter 6)**

##### *Empowering Diverse Researcher Perspectives via Accessible Tooling and Resources*

Tooling and resources must be available and accessible to different disciplines as well as the many languages and perspectives needed to drive responsible innovation. Implementing Finding 6-1 (Engaging diverse stakeholders) is critical for incorporating the expertise and viewpoints of the many stakeholders affected by AI systems. This means at minimum providing resources in multiple languages, which can be based on the most spoken languages in the U.S. The [BigScience Research Workshop](#), a community of over 1000 researchers from different disciplines hosted by Hugging Face and the French government, is a good example of empowering perspectives from over 60 countries to build one of the most powerful open-source multilingual language models. Our platform encourages many groups to curate datasets and evaluations in their native or fluent languages makes for stronger, representative science.

### **Conclusion**

NAIRR aligns closely with Hugging Face's mission to democratize AI in a responsible direction. We look forward to supporting this initiative as the Task Force works toward its final report and are eager to contribute to implementation.

Respectfully,

Irene Solaiman

Hugging Face