AI2 & BigScience Legal Hackathon

March 2024

About

The BigScience workshop (https://bigscience.huggingface.co/en/#!index.md) was an effort started in January 2021 that gathered 1000+ participants from over 40+ countries to research questions surrounding large language models (capabilities, limitations, potential improvements, **bias, ethics,** environmental impact, role in the general AI/cognitive research landscape) as well as the **legal and ethical** challenges around creating and sharing such models and datasets for research purposes and among the AI research community. The core effort resulted in the creation of the multilingual Large Language Model BLOOM and of the Responsible Open-science Open-access Text Sources corpus (ROOTS) in June 2022, and brought together a multidisciplinary community of researchers working on different aspects of AI governance. In the spirit of continuing the project's mission of making AI technology more equitable, this community and the Allen Institute or Artificial Intelligence (AI2) are continuing to organize yearly legal hackathons to help provide users and affected parties with legal resources related specifically to responsible model licensing and data subject rights.

Overview of the research topic(s) and scope

Jurisdictions around the world are taking different approaches to answering the many legal questions raised by the development and adoption of AI systems. New transversal regulations like the EU AI Act provide novel general frameworks for specifying the conditions of use and legal requirements for any tool that relies on AI or Machine Learning model, while documents like the recent White House Executive Order remind readers that AI is subject to existing regulations governing personal data, anti-discrimination, consumer protection, etc. – to be interpreted in the context of this new type of technology. In this context, it can be difficult for the people affected by AI systems to understand where they can turn to for legal recourse when uses of the technology go wrong. As part of this study, the researchers will leverage their legal expertise to help address this issue of fragmentation by helping organizations and individuals identify the jurisdictions and legal frameworks that are most relevant to specific questions about AI systems.

The hackathon is hosted by the Allen Institute for Artificial Intelligence which is a 501c3 non-profit, in collaboration with BigScience, and Hugging Face.

Work Product

*around 50 hours of legal research per student.
- Collaborating on a guide mapping different AI concerns and examples of AI incidents from selected databases (see "Resources for collating questions" below) to:

- ○ the most relevant regulation at the city, state, or federal level in the US and to relevant regional and/or national regulations in the researcher's jurisdictions of interest, and
- ○ the most relevant legal framework (*e.g.* consumer protection, platform governance, intellectual property, labor rights, sectoral regulation in health, education, or finance, etc.) and competent authority (local or federal courts, the competition authority or other agency, an international court, etc.), and outlining gaps where no existing framework applies.
- Collaborating on a research memo outlining the different regulatory approaches to defining rights and responsibilities in the context of AI by mapping the provisions outlined in the proposed EU AI Act and EU updated liability directive to relevant US local/state and federal regulations and agencies.

Relevance of the project(s)

- We aim to create resources to facilitate more responsible development of the technology by actors concerned with its social impact, and to help individuals navigate and exercise any of their rights that may be affected by the use of AI systems in a quickly evolving technical context, while also highlighting gaps for future policy directions.
- These outcomes will also be directly useful to developers and researchers to help them promote responsible practices and help set positive norms for the technology by leveraging legally grounded governance tools

Relevant websites for context

- **Resources for collating questions**:
  - ○ 📗 AI Ethics & Policy News
  - ○ Partnership for AI Incident Database
  - ○ AI, Algorithmic, and Automation Incidents and Controversies Database
  - ○ Awful AI Database
- AI2 x Hugging Face x NYU Law - 2023/24 Edition
- 2022/2023 hackathon research memo:
  📄 Group 1: Regulatory Landscape of AI Use Constraints
- https://bigscience.huggingface.co/

LLM student participants

We seek LLM students with experience or interest in AI, IP, privacy, fundamental rights, open research and data, and standards.