Data & Society and BigScience Legal Hackathon

March 2025

About

The BigScience workshop hosted by Hugging Face 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 continues organizing yearly legal hackathons in collaboration with major non-profits in the field to help provide users and affected parties with legal resources related specifically to responsible model licensing and data subject rights.

This year's hackathon is organized in collaboration with Data & Society. Mission description:

> "*Data & Society is an independent nonprofit research organization. We believe that empirical evidence should directly inform the development and governance of new technologies — and that these technologies can and must be grounded in equity and human dignity. Recognizing that the concentrated, profit-driven power of corporations and tech platforms will not steer us toward a just future, our work foregrounds the power of the people and communities most impacted by technological change. We study the social implications of data, automation, and AI, producing original research to ground informed public debate about emerging technology. This includes research on machine learning, AI systems, algorithmic impact, the future of work, and building trustworthy online infrastructures. Our research offers evidence to counter the notion that technology is the best or only solution to the host of societal challenges we face, and our communications, policy, and engagement work applies and amplifies those findings.*"

You can find the outcomes of previous legal hackathons collated here. For this year's hackathon, we seek applicants with an interest and experience in technology governance. A deep technical understanding of AI is not required as the hosts will be there to discuss specifics, but experience with any of the components of the AI development chain will be particularly valuable (including software compliance, data and privacy regulation, permitting for hardware components, digital labor law, etc.) We also intend to have a cross-jurisdiction analysis

component for all projects, comparing US regulation with any other jurisdictions the participants are familiar with.

Overview: Legislative Recourse against AI Harm

We're proposing three projects that look into how current legislative frameworks can help support the agency of people who are (negatively) affected by the development and deployment of AI systems. In particular, environmental and labor related externalities of AI can be particularly damaging for the very communities and individuals that benefit the least from what the technology has to offer. Identifying what legal tools exist to support these communities in asserting their needs and protected rights in the face of external development decisions can help bring about more balanced outcomes.

Q1 - Legislative Tools to Govern Data Centers: from local collectivities to global regulation

Artificial Intelligence, especially the AI systems that are powered by Large Language Models and similar computation-hungry technical paradigms – such as ChatGPT, Gemini, X's Grok, Anthropic's Claude, etc. – all rely on large data centers[1] to function. These data centers are large physical infrastructure that require massive energy and water to function – putting a strain on the local supply – and can create significant immediate pollution of their environment and harms to residents of the communities where they're implanted. Further, we often see collectivities offering tax incentives to large companies to build these data centers, namely motivated by job creation promises, further draining the communities' resources.

While legal tools do exist from the local to the federal level (in the US) to challenge these practices, or at least ensure that data centers are constructed in less extractive fashions when warranted, they can be fragmented and difficult to navigate for advocacy organizations. The goal of this project will be to map these tools and compare what is available across different cities and states, with a particular focus on the regions that have seen the most AI data center construction in recent years, and how federal regulation and recent Executive Orders impact the conditions for data center construction and governance. We also welcome analysis of and comparison to jurisdictions outside the US that take a different approach to these questions.

Project specific resources:

- North Star Data Center Policy Toolkit: State and Local Policy Interventions
    - We will use this resource as a starting point to see where further work is needed, especially for interventions that rely on transparency and disclosure
- ⚡ Power, Heat, and Intelligence ☁️ - AI Data Centers Explained 🏭

[1] ⚡ Power, Heat, and Intelligence ☁️ - AI Data Centers Explained 🏭

Q2 - Labor Rights and Workplace Legislation Frameworks for AI

AI is becoming increasingly ubiquitous in the workplace, sometimes brought in by workers (especially software engineers) who find utility in its functionalities, but often due to top-down decisions of executives who find it a perfect excuse to require more production for less investment in their workforce. In particular, AI adoption can come at the expense of privacy, job security, and workers' ability to organize and more generally control the conditions of their labor.

Similarly to the case of data center construction, adoption of AI-powered software in the workplace or with application to worker management is subject to specific requirements, whether they come from general privacy rules, domain-specific requirements in e.g. health or finance, or labor protections. However, untested assumptions about the functionality and security of AI systems as well as hard-to-navigate liability regimes can make it difficult for workers and their representation to push back against damaging uses of AI on the basis of these regulations.

The goal of this group will be to identify specific challenges that workers can bring to the introduction of inappropriate AI tools in their workplace. We'll start from complaints expressed by worker representation organizations in the last two years and prior work from previous hackathons. While the primary focus will be on US regulation, we welcome analysis and comparison to other jurisdictions based on the experience of hackathon participants.

Project specific resources:

- Data & Society — (404) Job Not Found
- In Weak Job Market, Middle Managers Increasingly Forced to Feign AI Success | TechPolicy.Press
- NNU on Trump executive order seeking to ban state A.I. regulation: 'Patient lives are at risk' | National Nurses United
- Stakeholder-specific adoption of AI in HRM: workers' representatives' perspective on concerns, requirements, and measures

Q3 - US-China Cross-Analysis of AI Industrial Strategies

Comparative study of regulatory approaches between the US (under the Trump administration) and China, focusing on national-level industrial strategy to pursue national goals. What are stated goals? What tactics are they using to achieve them? What are the similarities and differences between the two?

Project specific resources:

- The mirage of AI deregulation | Science
- Overview of Chinese Regulation and Policy documents on AI