Stanford University

interface

OXFORD MARTIN SCHOOL

Centre for the Governance of AI

Position Paper
Technical Research and Talent is Needed for Effective AI Governance

Anka Reuel*, Lisa Soder*, Ben Bucknall, Trond Arne Undheim

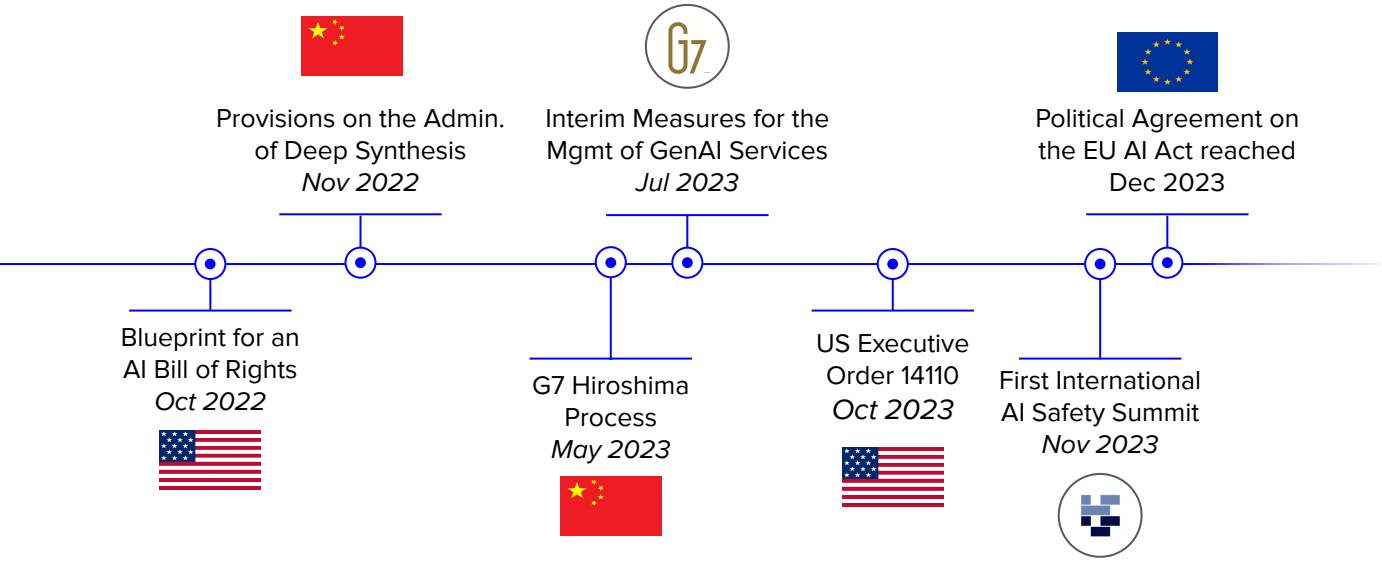
Technical research & tools

Targeted AI/ML research is needed to ensure the effective enactment of current regulatory proposals

Technical expertise & talent

Closer collaboration between technical researchers & policy makers is necessary to ensure informed governance.

AI policy has seen significant developments in recent years



Many of these regulations will require **technical tools** and **expertise** to operationalise and enact them

Providers of GPAI models with systemic risk shall: perform model evaluation in accordance with standardised protocols and tools [..] keep track of, document, and report, without undue delay, to the AI Office

EU AI Act Article 55

Open questions

- How can the **thoroughness** of evaluations be measured?
- How can the **downstream impacts** of AI systems be predicted?
- How can potential **data contamination** be accounted for when conducting evaluations?
- What are the **access & infrastructure** requirements for external testing?
- What information about the testing should be **disclosed**?

The Secretary shall require compliance with these [red teaming] reporting requirements for: (i) any model that was trained using a quantity of computing power greater than 10^26 FLOP

Executive Order 14110

Open questions

- Are compute thresholds a good **proxy for risk**? If so, should they be updated over time?
- What additional **designation criteria** could determine the level for regulatory requirements?
- How to **measure training compute**? What should be counted as compute expenditure of a system's development?

Technical AI Governance:
Technical analysis and tools to support AI governance by

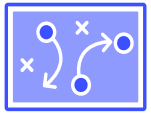
- Identifying need for policy intervention
- Informing governance actions
- Enhancing governance options

And there are many opportunities to get involved as a technical researcher

Roles in AI Governance

- Policy advisory roles, e.g., as part of the OECD AI Expert Groups
- Temporary expert positions, e.g., as seconded expert to the EU AI Office
- Full-time roles, e.g., as technical research at the AI Safety Institutes
- Think tank and research roles, e.g., to translate technical insights into policy recommendations
- Regulatory entrepreneurship, e.g., building AI accountability infrastructure or offering third-party audits

*A lot of this can be done in a **part-time capacity**!*



Research into relevant open technical problems

- Work on technical tools and research that support effective AI governance, for instance
- Assessments, i.e.., ability to evaluate AI systems, involving both technical analyses and consideration of broader societal impacts.
- Operationalization, i.e, the translation of policy objectives into concrete technical strategies, procedures, or standards
- Verification i.e., the ability to verify claims made about AI systems' development, behaviours, capabilities, and safety.

*Check-out our new paper with **100+ research questions**!*

Counterposition | Technical AI governance might risk...

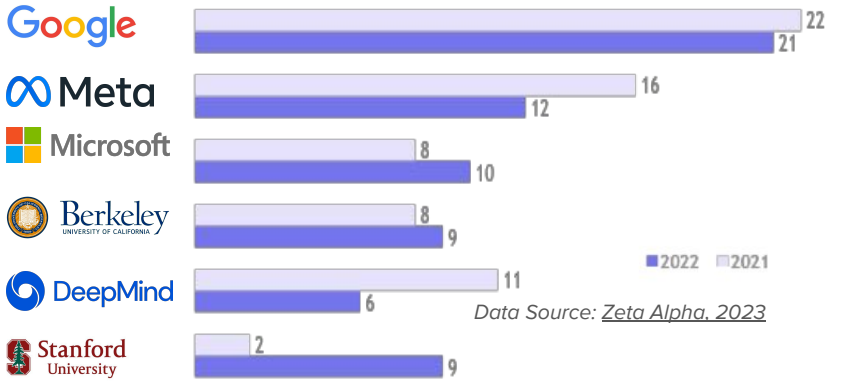
Techno-solutionism. Relying solely on technical fixes for complex social problems can bypass democratic oversight and overshadow other types of approaches and views.

Regulatory capture. Overly focusing on technical experts and fixes can open the door for industry influence (e.g., revolving doors) and create complexity, making scrutiny more difficult and lobbying easier.

Example Cultural Capture

Relying on technical expertise can open the door for cultural capture, in which “regulator begins to **think like the regulated industry**”, as research directions, findings, and ideas in AI are inherently **shaped industry**.

Count of AI papers in the top 100 (measured in citations) by organisation



Data Source: Zeta Alpha, 2023

Thanks!

@AnkaReuel

@lisa_soder_

@ben_s_bucknall

@trondau

Please reach out with questions, feedback, etc. anka@cs.stanford.edu or lsoder@interface-eu.org - we would be excited to hear from you!

Check out our **Position Paper** on the Need for Technical Research & Talent
arXiv: 2406.06987

And our new paper on **Open Problems** in Technical AI Governance
arXiv: 2407.14981