Market Based Compute Permits for Frontier AI Safety

SPAR Spring 2026 Project Proposal

Joel Christoph

1. Summary

This project will design and analyze market based compute permit schemes for frontier AI training. The goal is to make powerful AI systems easier to govern and audit by specifying concrete mechanisms for allocating, trading, and enforcing rights to use high end compute.

Mentees will help survey existing proposals, build simple economic and game theoretic models, and implement small empirical or simulation based case studies. The main outputs will be a short working paper or policy brief and a slide deck suitable for SPAR Demo Day and relevant policy audiences.

2. Motivation and background

Compute has emerged as one of the most promising levers for governing transformative AI. Proposals for compute caps, licensing, and monitoring at the chip and data center level are now on the table in several jurisdictions. However, the design details of market based compute governance mechanisms remain underspecified.

Tradable emissions permits in climate policy offer a natural analogy, but compute differs in several ways. It is highly concentrated in a small number of firms and regions, tightly linked to security and industrial policy, and subject to rapid technological change. The design of compute permit schemes therefore has to balance innovation, competition, security, and safety in a dynamic and strategic environment.

This project builds on recent work on compute governance, hardware enabled oversight, and global AI competition. It aims to provide decision makers with a small set of concrete, analytically grounded options for how market based compute permits could be structured in practice.

3. Research questions

Illustrative questions include:

- At what level should compute permits be defined: per training run, per lab, or per data center, and in what unit (for example FLOP thresholds, chip hours, or verified energy use).
- How should permits be allocated initially (auction, grandfathering, or hybrid), and how quickly should the cap tighten over time under different risk scenarios.
- Under what conditions would international trading of compute permits reduce risk, and when might it instead create new channels for leakage or regulatory arbitrage.

- How should enforcement and monitoring work, given the technical constraints on measuring actual compute usage and the possibility of clandestine training runs.
- How can compute permits be integrated with other policy tools, such as export controls on advanced chips, safety standards for frontier models, or licensing regimes.

The focus will be on simple but transparent models and case studies that clarify tradeoffs rather than on large scale simulations.

4. Methods and work plan

The project is designed for a small team of motivated mentees working about 5 to 10 hours per week over approximately three months. A tentative structure is:

Weeks 1 to 3: Landscape review. Mentees survey and synthesize existing proposals on compute governance, emissions style permit systems, and related areas in mechanism design and industrial organization. The team produces a short internal memo outlining design dimensions and open questions that matter most for AI safety.

Weeks 4 to 8: Modeling and case studies. Mentees develop one or two simple economic or game theoretic models that capture key features of compute permit schemes. Examples include:

- A partial equilibrium model where labs choose compute and safety investments under a binding permit cap.
- A two country model where governments choose enforcement levels and permit trade policies under different security preferences.

Where skills allow, the team can implement these models in Python or a similar language and explore parameter sweeps or simple simulations.

Weeks 9 to 12: Synthesis and public facing outputs. The team distills insights into a concise working paper or policy brief, plus a slide deck aimed at policymakers and funders. We will also prepare a short summary suitable for a blog post or newsletter if time permits.

5. Role of mentees and supervision

Mentees will act as junior coauthors. They will help choose which modeling approaches to pursue, run literature reviews, implement models and scenarios, and contribute directly to writing and visualization.

I anticipate:

- One weekly one hour team meeting for progress updates, troubleshooting, and planning.
- At least one additional hour per week of asynchronous feedback on drafts, code, and ideas.
- Optional co-working or pair research sessions during busier weeks, depending on mentee availability.

I have prior experience mentoring and managing early career researchers through SPAR, AISC, Effective Thesis, and other programs. I enjoy helping mentees sharpen research questions, write clearly for mixed technical and policy audiences, and connect their work to concrete theories of change for AI safety.

6. Theory of change for AI safety

If powerful AI systems are trained on unregulated and opaque compute infrastructure, it will be difficult for any government or international body to enforce meaningful safety standards. Compute permits are one promising way to restrict and monitor large training runs while preserving flexibility for innovation and beneficial uses.

By clarifying the design space for market based compute permits, this project aims to:

- Identify designs that are both safety enhancing and politically feasible.
- Provide concrete options to policymakers in governments and multilateral institutions who are exploring compute caps and licensing.
- Highlight failure modes and unintended consequences that should be avoided in future regulation.

In the medium term, this work could inform white papers, consultations, or regulatory proposals in jurisdictions considering compute based AI governance. In the longer term, it contributes to a broader toolkit for governing transformative AI through the infrastructure that enables it.