

# Frontier AI Regulation

Managing Emerging Risks to Public Safety

By: Connor McDonald, Jonathan Spsychalski, and Ryan Kim



## Regulation of Frontier AI


- **Safety Standards** – Create/enforce standards through expert processes
- **Visibility** – Increase oversight with disclosures, audits, and whistleblower protections
- **Compliance** – Use self-regulation, enforcement, and licensing
- **Balanced Approach** – Avoid stifling innovation; invest in expertise and flexible rules



## The Regulatory Challenges of AI Models

### What are Frontier AI Models?

- Highly capable models with potential for dangerous capabilities
- Potential risks:
  - Biological/chemical weapons
  - Tailored disinformation
  - Cyber threats
  - Evasion of human control
- Regulatory needs
  - Models require evolving definitions and oversight; current definitions are insufficient



## Building Blocks for Frontier AI Regulation



## Institutionalizing Frontier AI Safety Standards

- **Support Multi-Stakeholder Efforts** – Involve experts, researchers, and consumers in developing standards.
- **Pilot and Refine Practices** – Test and improve safety practices; evolve into best practices.
- **Develop Robust Standards** – Create methods to evaluate dangerous capabilities and risks.
- **Government Role** – Invest in testing, auditing, and research; update procurement; provide regulatory guidance.



## Increasing Regulatory Visibility

- **Develop Disclosure Framework** – Facilitate voluntary and mandatory AI disclosures.
- **Implement Reporting Requirements** – Enforce regular reporting on AI models and development processes.
- **Conduct Audits** – Audit AI companies against safety and risk-management frameworks.
- **Establish Whistleblower Protections** – Protect individuals who report safety-critical information.



## Ensure Compliance with Standards

- **Encourage Self-Regulation** – Support voluntary self-regulation and certification frameworks.
- **Mandate Compliance** – Require adherence to standards with penalties for non-compliance.
- **Implement Licensing** – Consider licenses for developing and deploying high-risk AI models.
- **Prepare for Rigorous Enforcement** – Invest in standards, expertise, and adaptable regulation.

## Discussion

- What are the most effective methods for ensuring compliance with AI safety standards?
- How should regulators balance innovation with the need for safety in frontier AI development?
- What role should self-regulation and certification play in the broader regulatory framework for AI?