

## Review

*A link to the preprint is provided below. A peer-reviewed version will appear in the European Journal of Risk Regulation.*

*Kris Carlson prompting ChatGPT4o*

Please provide a summary of Williams et al, On Regulating Downstream Developers, arXiv:2503.11922, itemizing 1) each difficulty of regulating downstream developers, 2) their overall strategies to deal with each difficulty, and 3) each tactic they advocate. Then follow that by a critique of the weak points of their positions, and alternate regulation methods advocated by others. No length limit.

# On Regulating Downstream AI Developers

Sophie Williams, Jonas Schuett, Markus Anderljung

This paper addresses the regulatory challenges posed by downstream developers—those who adapt or fine-tune foundation AI models—and proposes strategies to mitigate associated risks.([arXiv](#))

---

## 1. Difficulties in Regulating Downstream Developers

### a. Diverse and Expanding Developer Base

Downstream developers range from individuals to large corporations, making uniform regulation challenging.([Moonlight](#))

### b. Risk Amplification through Modifications

Modifications can inadvertently or deliberately enhance harmful capabilities or disable safety features.([Moonlight](#))

### c. Limited Oversight and Knowledge

Downstream developers may lack comprehensive understanding of the foundation models, leading to unintended risks.([Moonlight](#))

### d. Practical Challenges in Enforcement

The sheer number and diversity of downstream developers complicate regulatory enforcement and risk stifling innovation.([Moonlight](#))

---

## 2. Strategies to Address These Difficulties

### a. Direct Regulation of Downstream Developers

Impose obligations such as risk assessments, incident reporting, and documentation.([Moonlight](#))

### **b. Upstream Regulation to Mitigate Downstream Risks**

Require foundation model developers to implement safeguards that limit risky modifications.([arXiv](#))

### **c. Non-Regulatory Measures**

Utilize existing legal frameworks (e.g., tort law) and issue voluntary guidelines to encourage responsible behavior.([Moonlight](#))

---

## **3. Recommended Tactics**

### **a. Voluntary Guidance**

Develop best practices for both upstream and downstream developers to mitigate risks.([Moonlight](#))

### **b. Upstream Developer Obligations**

Mandate upstream developers to assess and mitigate potential risks from downstream modifications.([arXiv](#))

### **c. Monitoring and Targeted Regulation**

Continuously monitor the downstream ecosystem and introduce targeted regulations when necessary.([arXiv](#))

---

## **Critique of the Paper's Positions**

### **a. Overreliance on Upstream Developers**

Placing significant responsibility on upstream developers may not be effective, especially when downstream modifications are extensive and unpredictable.

### **b. Insufficient Focus on Enforcement Mechanisms**

The paper lacks detailed discussion on how proposed regulations would be enforced, particularly given the global and decentralized nature of AI development.([Moonlight](#))

### **c. Potential Innovation Stifling**

Even voluntary guidelines, if perceived as de facto standards, could discourage experimentation and innovation among smaller developers.([Moonlight](#))

---

## **Alternative Regulation Methods Advocated by Others**

### **a. Third-Party Compliance Reviews**

Implement independent audits of AI systems to ensure adherence to safety frameworks, providing assurance to both developers and users.([arXiv](#))

**b. Developer Education and Tooling**

Focus on educating downstream developers about AI safety and providing tools to assist in safe model adaptation.([arXiv](#))

**c. Certification Systems**

Establish certification processes for AI models and modifications, signaling compliance with safety standards.([Moonlight](#))

**d. Community-Driven Standards**

Encourage the development of community-driven standards and best practices, fostering a culture of safety and responsibility.

---

In conclusion, while Williams et al. provide a foundational framework for considering the regulation of downstream AI developers, a multifaceted approach that includes enforcement mechanisms, developer support, and community engagement may offer a more balanced and effective path forward.