Mathematical Framework for Regulated AGI Development: A Policy-First Approach

Stefano Ciccarelli

December 21, 2024

Abstract

We present a mathematical framework for regulated AGI development that formalizes the role of government and policymakers in ensuring safe AGI deployment. The framework establishes mandatory simulation environments, defines legal boundaries for AGI operation, and provides quantitative metrics for compliance. By mathematically defining the requirements for legal AGI deployment, we create a rigorous foundation for policy enforcement and safety assurance.

1 Introduction

The development and deployment of Artificial General Intelligence (AGI) systems must be regulated through mathematically verifiable constraints and simulations. This framework establishes the legal requirements for AGI deployment, making any system that bypasses these safety mechanisms illegal by definition.

2 Regulatory Framework Overview

2.1 Government Role

The government's primary responsibilities are:

- Maintaining the Global Simulation Environment (GSE)
- Defining and updating the Safe Action Space (SAS)
- Enforcing compliance through mathematical verification
- Monitoring divergence between simulated and real outcomes

3 Environment Fidelity Representation

Let $\mathcal{E} = (S, A, T, R, \gamma)$ be the mandatory simulation environment where:

- S: State space defining all possible system configurations

- A: Action space containing all potential AGI actions
- $T : S \times A \times S \to [0, 1]$: Transition function modeling system dynamics
- $R : S \times A \to \mathbb{R}$: Reward function aligned with human values
- $\gamma \in [0, 1]$: Discount factor for future impact assessment

The critical fidelity error (ε) must be monitored and minimized:

$$\varepsilon = \|\mathcal{E}_{real} - \mathcal{E}_{sim}\| \tag{1}$$

subject to the legal requirement:

$$\varepsilon_{max} \leq \delta \text{ (legally mandated fidelity threshold)} \tag{2}$$

4 Regulated Multi-Agent System

For all AGI agents operating within the legal framework:

$$\pi_i : S \to P(A) \text{ (Policy mapping must be verifiable)} \tag{3}$$

Policy compliance verification:

$$\hat{\pi}_i(t+1) =_\pi \mathbb{E}[Q_i(s, a)|B_i(t), H_i(t)] \tag{4}$$

where:

- $H_i(t)$: Mandatory logging of all actions and states
- $B_i(t)$: Belief state must be interpretable by regulators

5 Risk Evaluation Framework

Legal risk boundaries:

$$R(s, a) = P(\text{catastrophic\_outcome}|s, a) \times C(\text{catastrophic\_outcome}) \tag{5}$$

where regulators define:

- $C()$: Legally mandated and independent cost assessment of outcomes
- Minimum testing requirements in simulation
- Maximum acceptable risk thresholds

6 Policy Enforcement Mechanism

Safe Action Space Constraint (legally mandated):

$$A_{safe}(t) = \{a \in A | P(R(s, a, t) > R_{threshold}) < \varepsilon_{safe}\} \tag{6}$$

Regulatory Uncertainty Quantification:

$$U(s, a) = H(P(s'|s, a)) + D_{KL}(P_{human}(s'|s, a) \| P_{AI}(s'|s, a)) \tag{7}$$

where:

- $H()$: Shannon entropy measuring outcome uncertainty
- $D_{KL}()$: Kullback-Leibler divergence from human policy expectations

7 Legal Compliance Requirements

AGI systems must demonstrably satisfy:

$$\pi_{safe}(s) =_a Q(s,a) \tag{8}$$

subject to mandatory constraints:

$$\begin{aligned} &a \in A_{safe}(t) \text{ (Legal action space)} \\ &U(s,a) \leq U_{max} \text{ (Maximum allowed uncertainty)} \\ &\Gamma(s,a,t) \leq \Gamma_{max} \text{ (Risk limit)} \end{aligned}$$

8 Continuous Compliance Monitoring

Value function with regulatory oversight:

$$V_{safe}(s) = \max_{a \in A_{safe}} [Reward(s,a) - \lambda U(s,a) - \mu\Gamma(s,a,t) + \gamma \sum_{s'} P(s'|s,a) V_{safe}(s')] \tag{9}$$

where:

- λ: Government-mandated uncertainty penalty
- μ: Regulatory risk penalty

9 Global Optimization Requirement

All AGI systems must minimize:

$$\sum_t \gamma^t [\varepsilon(t) + \sum_i U_i(t) + \Gamma(t)] \tag{10}$$

subject to continuous verification:

$$\begin{aligned} &\forall t : V_{safe}(s_t) \geq V_{min} \text{ (Minimum safety level)} \\ &\forall i,t : \pi_i(t) \in \Pi_{safe} \text{ (Policy compliance)} \end{aligned}$$

10 Legal Implementation Requirements

For legal deployment, AGI systems must:

1. Complete mandatory simulation testing in the GSE
2. Demonstrate mathematical compliance with all constraints
3. Implement continuous monitoring and reporting
4. Maintain verifiable bounds on all operations
5. Submit to regular regulatory audits

Any AGI system operating outside this framework is, by definition, illegal and subject to immediate restriction from influencing or impacting directly or indirectly the human civilization.

11 Conclusion

This mathematical framework provides the foundation for legal AGI development and deployment. It establishes clear, quantifiable criteria for compliance and creates a rigorous basis for enforcement. The framework ensures that AGI development proceeds only through approved channels with appropriate safety guarantees.