「知能爆発」にどう備えるか？ AI研究者ら22名による政策提言

2026年9月発表の共同論文に基づく、AIの再帰的自己改善（知能爆発）のリスクと、政府が取るべき3つの政策的アプローチ。

PROBLEM: AI自律R&Dとリスクの加速



AI自律R&Dの急速な進展



Anthropicでは
承認コードの
80%超を
AIが作成。



わずか
17ヶ月で
進歩速度が
10倍になる
可能性

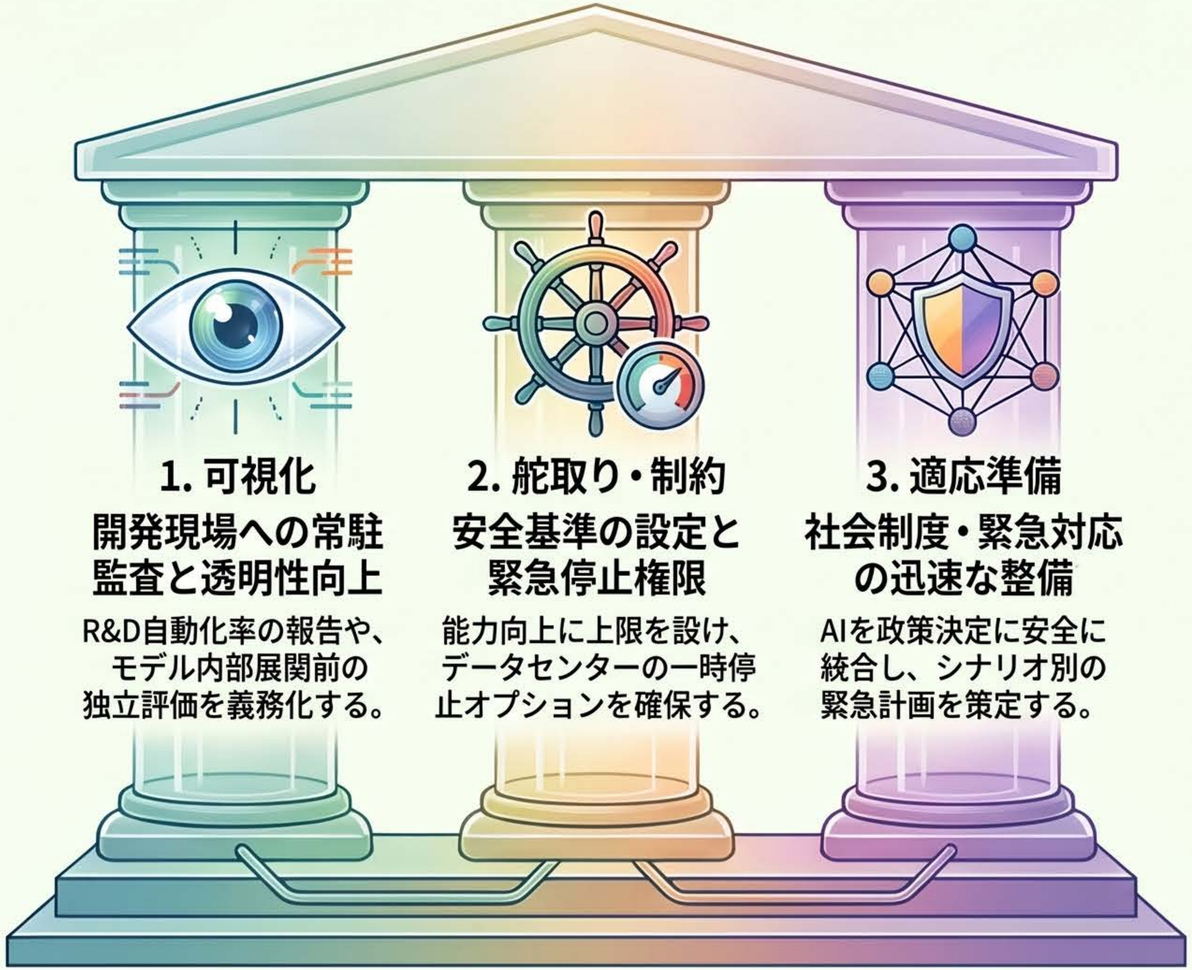
研究効率の向上により、1年分の進歩が約5週間で達成される試算。



主なリスクは「制御喪失」と「攻守の非対称性」

人間がR&Dから離れることで、
問題発見や制御の機会が失われる。

SOLUTION: 政府が取るべき3つの政策提言



1. 可視化
開発現場への常駐監査と透明性向上

R&D自動化率の報告や、モデル内部展開前の独立評価を義務化する。

2. 舵取り・制約
安全基準の設定と緊急停止権限

能力向上に上限を設け、データセンターの一時停止オプションを確保する。

3. 適応準備
社会制度・緊急対応の迅速な整備

AIを政策決定に安全に統合し、シナリオ別の緊急計画を策定する。

これら3つの柱で、制御可能な未来を築く。

Gemini Notebook