

アンソロピック提言「AIが自らを作る時」：自己改善AIと開発一時停止の真実



提言の核心：即時停止ではなく「選択肢」の確保



「即時停止」ではなく「停止できる権利（オプション）」
社会制度や安全研究が技術進歩に追いつくために、開発を減速・停止できる仕組みを事前に構築することを提言しています。



条件付き・仮定法の提言
提言は「もし効果的に減速できるのであれば (if it were possible)」という仮定法で書かれており、一方的な自己規制を呼びかけるものではありません。

実効的な停止に必要な「2つの条件」



(a) 機軸国の主要研究所による同時合意
先頭企業が入れ替わるだけでは意味がなく、資金力のある主要プレイヤーが同一条件で合意する必要があります。



(b) 検証可能性 (Verifiability)
互いに本当に停止したことを外部から検証システムが不可欠であり、これには「compute attestation (計算業務の証明)」等の技術が検討されています。

提言の核心：即時停止ではなく「選択肢」の確保

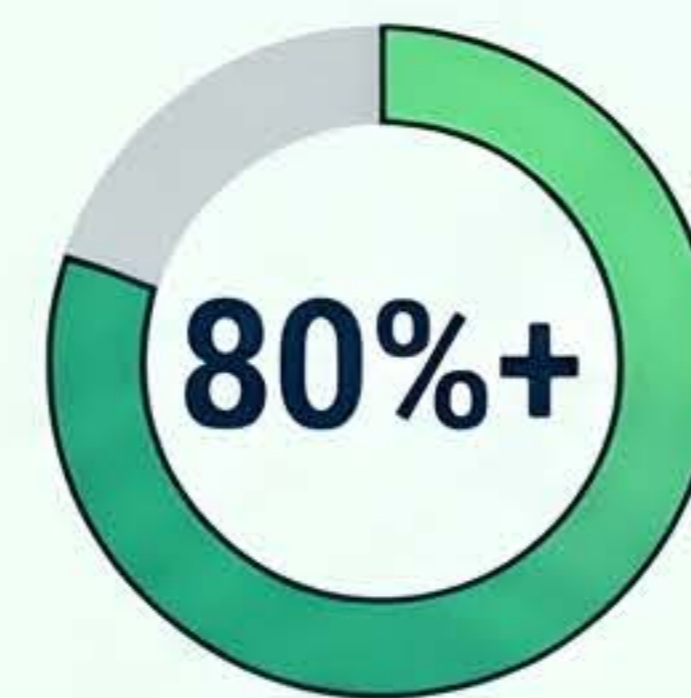


「即時停止」ではなく「停止できる権利（オプション）」
社会制度や安全研究が技術進歩に追いつくために、開発を減速・停止できる仕組みを事前に構築することを提言しています。



条件付き・仮定法の提言
互換は「もし効果的に減速できるのであれば (if it were possible)」という仮定法で書かれており、一方的な自己規制を呼びかけるものではありません。

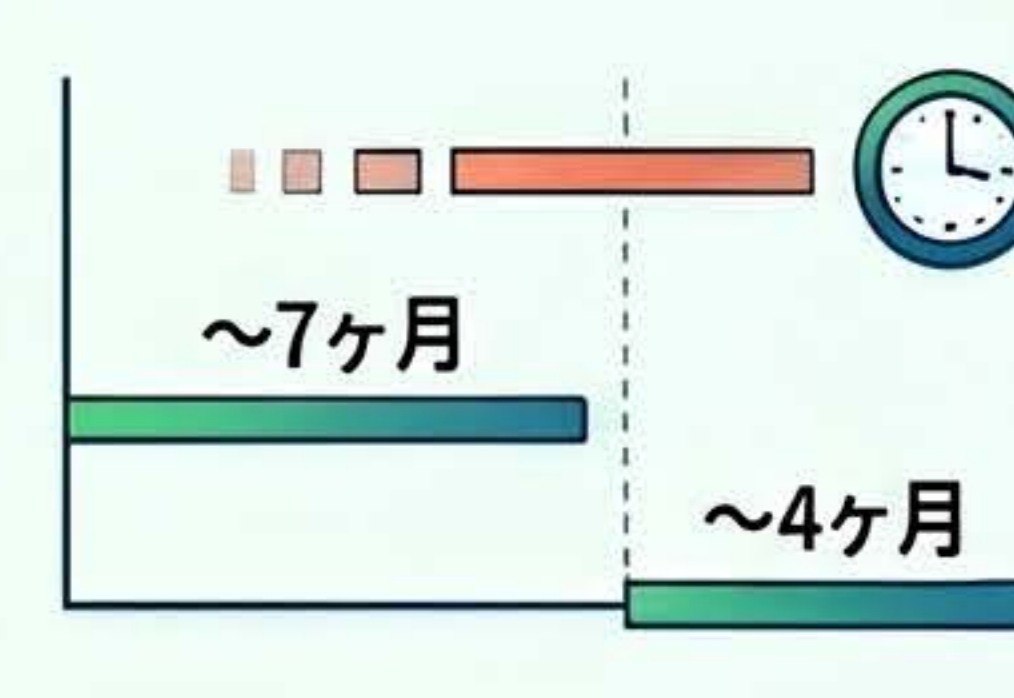
加速する自己改善：公開された内部データ



コード執筆の80%超をClaudeが担当
2026年5月時点で、マージされた社内コードの8割以上をAIが執筆しており、1年前の「約5%」から劇的に増加しました。

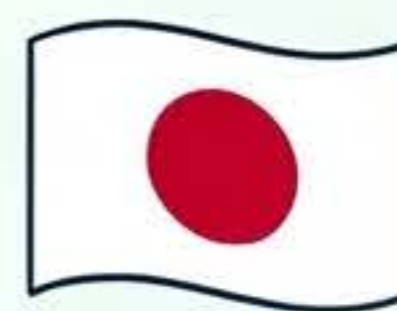


エンジニアの生産性が最大8倍に
2026年第2E5中期の日文コードマージ量は、2024年比で6倍に伸びました（ただし中央値は約4倍との自己評価もあり）。



タスク到達時間の増ペースが加速
AIが自律でこなせるタスクの「時間ホライズン」は、従来の約7か月から、現在は約4か月ごとに増えています。

政策・規制・知財への影響



各国の規制スタンスの温度差
ハードローで規制を強めるEUに対し、日本は「AI推進法」により「世界で最も開発しやすい国」を目指しており、国際的な停止枠組みへの参加は不透明です。

知財高裁「DABUS判決」：発明者は自然人のみ
日本の知財高裁（2025年1月）は、AIを特許法上の発明者とは認めず、現行法下では自然人の署名が必須であると判示しました。



知財実務での「人間の害与」の記録
AI支援発明を保護するためには、極がどのプロンプトを入力し、どう効果確認したかという「人間の創作的害与」の証跡管理が重要になります。

背景にある市場戦略と批判



IPO直前、評価額9,650億ドルの衝撃
2026年6月の他社申請時点で評価額は約150兆円に達し、一時的にOpenAIを上回ったとされ、提言のタイミングに注目が集まっています。



「規制による回り込み」との批判
一部の有識者は、この提言を安全性に基づいた善業ではなく、規則によって後発の競合他社を排除するための戦略的マーケティングであると指摘しています。