

生成AIの次なる戦場：モデル性能から「仕事完遂力」へのパラダイムシフト



新たな競争軸：「知能」から「エージェント実行力」へ

外部ツールを自律操作し、複雑なタスクを最後までやり切る能力が重視されています。単純な知識問答より、Web探索(BrowseComp)やコード修正(SWE-bench)の完遂率が指標に。

HLEスコア（ツール使用あり）



今後1～3ヶ月の主要モデル動向と予測

マルチモーダルによる「実作業」の自動化



マルチモーダルによる「実作業」の自動化

UI閲覧や画面からコードを生成し、アプリを構築するワークフローが生流になります。

Kimi K2.5：最大100体のエージェント並列稼働



Kimi K2.5：最大100体のエージェント並列稼働

人間の指示なしで多数のサブエージェントを自律構成し、作業濃度を最大4.5倍に高めます。

安全規制がGrok等の「攻めの機能」を抑制



安全規制がGrok等の「攻めの機能」を抑制

ディープフェイク規制により、xAIなどの一部モデルは機能制限や安全対策に追われる見込みです。