

AI 評価の新基準：Artificial Analysis Intelligence Index v4.2 の全貌

2026年9月公開
AI 評価指標 v4.2

実務型エージェント能力と長文読解の比重を大幅に強化し、モデルの「真の実力」を測る非公開テスト比率が倍増。最新モデルランキングに大きな変動。

評価メトリクスの劇的な進化



実務・エージェント能力
へのシフト (60%重み)

エージェント処理と長文実務文書の評価重みが計60%を占める構成へ。



非公開
テスト比率
(40%)

公開データ
(60%)

40% へ倍増

テスト対策（ゲーミング）を防ぐため、非公開データの割合を従来の2倍に強化。



難関新ベンチマークの導入
(AA-Briefcase)

数週間のプロジェクトを模した「AA-Briefcase」等を最大の評価軸に採用。

最新モデルランキングと市場勢力図



モデル選定の推奨基準



中国勢とオープンウェイトの躍進
Kimi K3 (中国) がオープン勢首位。
コスト効率では米中が拮抗。



モデル選定の新たな判断基準
総合スコアだけでなく、用途別のサブスコアとコスト効率での検証を推奨。