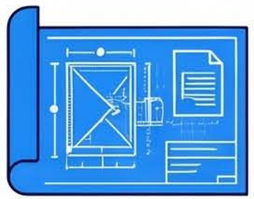
Gemini 4 Argon：長時間エージェント時代のフロンティアAIモデル

チャット知能から長時間タスク完遂型エージェントへの転換と、その技術的・経済的優位性

長時間タスクを完遂する技術基盤

「チャット知能」から「長時間エージェント」への移行

実コード改修や法務・財務調査など、複雑なタスクを最後まで完遂することに重点。









旧世代：
チャット知能

新時代：
長時間エージェント



100万トークンの「統合コンテキスト」導入

入力・思考・出力を1,048,576トークン内で共有し、継続プロトコルで大規模生成に対応。

初答約45秒・生成100+ t/sの「Deep Work」仕様



初答
~45秒
(思考重視)



生成 100+ t/s
(高速出力)

思考を挟むため初動は遅いものの、高い出力スループットで長時間の作業を効率化。

競合性能と攻撃的な価格戦略



実務ベンチマーク19項目中14項目で首位（同点含む）

経済価値の高い実務（Vals Index）や法務・財務評価で優れた成績を記録。

経済価値の高い実務（Vals Index）や法務・財務評価で優れた成績を記録。

100万トークンあたり入力$2 / 出力$10の低価格



最上位エージェント型モデルを経済的に大量稼働させられる攻撃的な導入価格。

小型モデルとの「階層型ルーティング運用」が最適



タスク入力



小型モデル
(例: Flash)

簡易処理はFlash等の小型モデルに任せる





高複雑度タスクをArgonへ割り当てる構成を推奨。

Gemini Notebook