Qwen4-Max 正式リリース：性能とコストの徹底解剖

モデル概要とコスト効率



2.4兆のMoE構造

活性710億で100万入力対応



性能評価と日本企業の留意点



専門強みと長尺弱み

法務に強く長尺操作に課題



東京リージョンの留意点



デプロイ表記Globalの確認要



段階的評価の推奨

成果総額で比較し機密入力を回避

コスト60%削減



入力$1.5・出力$4.5に値下げ



重み非公開のAPI提供

モデル本体の重みは非公開で、
API経由のみの利用となります。

料金比較（100万トークンあたり）

モデル	入力料金	出力料金	キャッシュヒット
Qwen4-Max（国際）	$1.50	$4.50	$0.075
Qwen3.8-Max（前世代）	$2.00	$6.00	$0.250
Claude Opus 5.5	$4.00	$20.00	$0.200

Gemini Notebook