

Qwen3.8-Max 徹底解剖：圧倒的コストパフォーマンスとガバナンスの光と影

次世代AIモデル「Qwen3.8-Max」の主要スペック、性能、および導入におけるリスクを俯瞰的に理解する。

光（性能・低価格）

圧倒的なスペックとコスト破壊



2.4兆パラメータの 超巨大MoE アーキテクチャ

アクティブパラメータ950億、
100万トークンの広大なコン
テキスト窓を持つマルチモー
ダルモデル。



米国フロンティア勢の 3～4分の1という衝撃価格

入力\$2／出力\$6（100万トークン辺り）
で、Claude Opus 5比で約7割安。

主要な競合モデルとのAPI価格比較（推定コスト比）

モデル	推定コスト比	特徴
Qwen3.8-Max	基準 (1.0x)	圧倒的安価・コーディング特化
Claude Opus 5	約3.0x～4.0x	高い推論能力・信頼性
GPT-5.6 Sol	約4.0x	フロンティア最高峰の性能

東京リージョン (ap-northeast-1) で即日利用可能

OpenAI互換APIを提供し、日本国内でのデータレ
ジデンシー要件にも対応。

影（法規制・運用負荷）

実力評価と避けて通れないリスク



コーディング・画像認識で 世界トップクラスの性能

Arenaランキングで中国モデル最上位。自律的な
研究再現デモなど高いエージェント能力を誇る。



中国国家情報法および 米国1260Hリストのリスク

政府による情報協力義務や、米国防総省の取引
制限リスト掲載による調達リスクが存在。



「名ばかりオープン」と ささやかれる運用難易度

ウェイト公開予定だが、2.4兆パラメータは
1.2TBのVRAMを要し自社ホストは事実上困難。