

AIモデル「MiniMax M3.1 Flash」が変える知財実務：コスト削減と導入戦略

MiniMax M3.1 Flashは、トークン単価を最大75%削減し、大量の特許業務コストを低減。機密保持と人間による検証を考慮した段階的導入が不可欠。

モデル特性と費用構造への影響



トークン単価が従来（M3）より低下

50%~75%

M3とM3.1 FlashのAPI利用料金比較（USD / 100万トークン）



通常入力・出力で50%減、キャッシュ読み取りで75%の削減を達成。



1,000,000トークンの
長文・マルチモーダル
対応

テキスト・画像・動画入力と
5段階の思考量制御に対応。



単純生成費より
「人の確認時間」が
総費用を左右
API価格低下の一方、
人間による確認時間等の
総費用管理が重要。

業務別の導入優先度と実証プロセス



公開情報の監視を「優先」
SDIや一次スクリーニングを優先
試行し、FTO等は慎重に判断。



未公開発明は
「条件付き」

30～60日間のスモールスタート実証（PoC）



公開特許
300～500件



従来の作業時間・
品質と比較