

Kimi K2.7 Code : 知的財産 (IP) 業務を再定義する自律型エージェントの衝撃

アーキテクチャの進化と特化設計

1兆パラメータMoE (1T) による極限の効率化



総パラメータ1兆 (1T) に対し、推論時は3% (32B) のみを活性化させる「スパースな活性化」により、低リソースでの高速処理を実現。

「考えすぎ」を 30%削減



先行モデルに比べ推論トークン (Thinking-token) を30%削減し、無駄な自己検証ループを排除。より直線的なタスク完遂を可能に。

決定論的な挙動を強制



決定論的な挙動を強制

Temperature (1.0) やTop_p (0.95) をシステムレベルで固定。クリエイティブさより、コード生成やAPI呼び出しの安定性を重視した設計。

オープンソース & 256Kコンテキスト



Kimi K2.8 → K2.7 Code

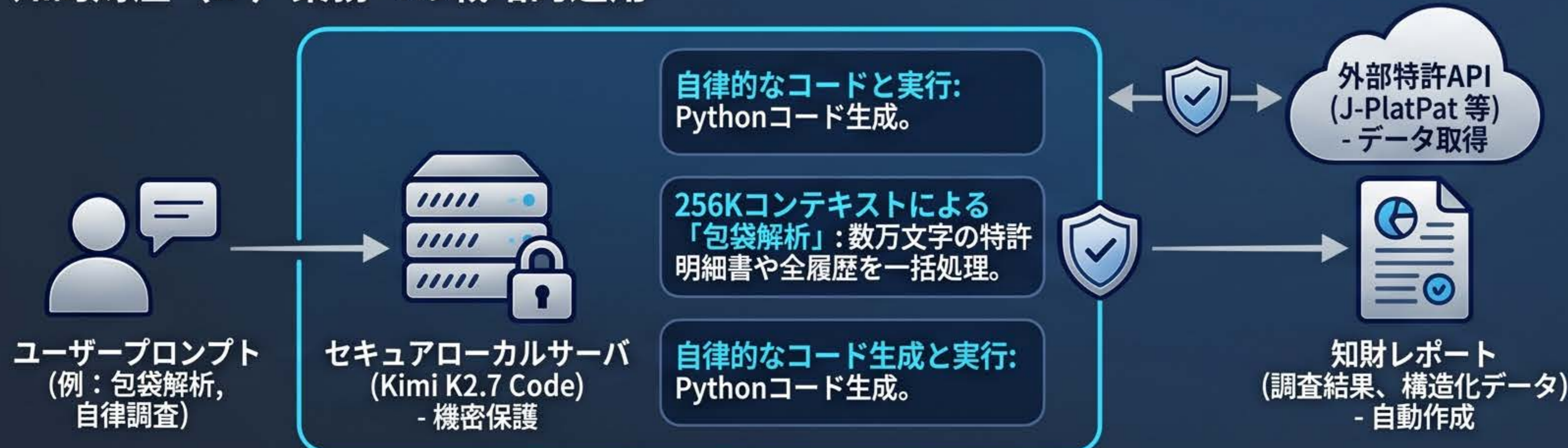
改変版MITライセンスで公開。256Kトークン超長文読解に対応。

パフォーマンス：公式スコアと実務評価の乖離



エージェント能力 (MCP) でClaude Opus (76.4点) を凌駕。低レベルコード生成における「正直な退行」あり。

知的財産 (IP) 業務への戦略的適用



未公開の発明提案書やFTO調査資料を外部APIに送ることなく、自社サーバー内のローカルLLMで安全に処理可能。J-PlatPat APIと連携した自律調査も実現。

破壊的なトークン経済性とライセンス



出力コストはClaude Opus 4.8の約1/6。プロンプトキャッシュ利用時は入力コストが\$0.19まで低下。

MIT 企業の防壁となる改変版MITライセンス: 基本的に無償利用可能。巨大プラットフォームには帰属表示を義務付け、独占を防止。