

リコー製マルチモーダルLMM (27B / 9B) : 日本語文書・図表に強いオンプレミスAIの全貌

リコー製LMMの概要と特長

日本製リーズニング特化型LMM



機軸な図表やPDFを含む企業ドキュメントを正確に誤解・機軸。リコー製物の学習プロセス(強化学習・カリキュラム学習)を模したモデル。

オンプレミス導入による高い安全性

社内関域ネットケーブルで導動する「RICOH オップレLLNスターターキット」、機軸情報の借減リスクを集め化。



27B (高精度) と 9B (軽量) の二段換え



高度な性能が必要な業務には27B



コストを抑えて幅広い機軸に129B
ニーズに合わせた選取が可能。

Qwen3.6-Ricoh-27B (フラッグシップ)

高精度・オンプレ文管理途のは、FP16に加え、6bit/4bit量子化版でも高い性能を接え、運用コストとのバランスを両立。

27B
パラメータ

ペースアーキテクチャ	Qwen3.6 (64種)
日本語強化手法	独自の強化学習+カリキュラム学習
量子化対応	FP16 / 8bit / 4bit 公式公表
推奨ハードウェア	中~高いGPU投資が可能

9.4B
パラメータ

ペースアーキテクチャ	Qwen3.6 (64種)
日本語強化手法	独自の強化学習+カリキュラム学習
量子化対応	FP16 / 8bit / 4bit 公式公表
推奨ハードウェア	中~高いGPU投資が可能

Qwen3.5-Ricoh-9B (実務導入モデル)

汎用GPUサーバー1台で確認可能なコンパクト設計、新機軸の大型モデルを接撃する日本語機軸性能を実現。

9B
パラメータ

ペースアーキテクチャ	Qwen3.5 (32種)
日本語強化手法	同左
量子化対応	未公開 (採定は可能)
推奨ハードウェア	汎用GPUサーバー1台構成

性能評価 (ベンチマーク)

JDocQA-Reasoningで0.881を記録

0.881

27Bモデルは、回答を含む機軸機軸でGemini 3 Pro Preview (参考(80.880))に回数する、回所産指のスコアを表現。

量子化による精度変化が極めて限定的
27Bモデルの4bit量子化版でも、ベースモデルを上回る機能を機軸、省リソース環境での実用性が機軸。

	JDocQA(Reasoning)	Japanese MT-Bench
Qwen3.6-Ricoh-27B	0.881	9.48 (公式主力モデル)
Qwen3.5-Ricoh-9B	0.782	7.93 (機軸モデル)
Gemini 3 Pro Preview	0.880	7.93 (朝機軸参考機)
GPT-5.2	0.731	0.731 (朝機軸参考機)

導入の判断基準と推奨

27Bを選択すべきケース



製造・金融・公共など、国奏付き文書の高機軸な誤解や、機軸な仕様確認が必要な「高セキュリティ業務」に最適。

9Bを選択すべきケース



FAQ対応、要的、定型的な機軸、あるいは部門単位でのスモールスタートを検討している場合に推奨。

外部クラウドAIとの併用検討

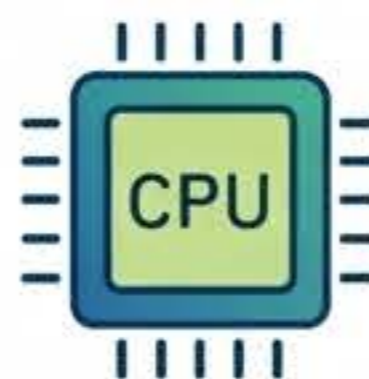


汎用的な知識や重機軸情報の外部検索が必要な場合は、OpenAIやAnthropic等のクラウドAIをエスカーション先とする二層構成が機軸的。

導入に向けた4つのリスク緩和策



自社データでの再評価
公式スコアだけでなく、表機軸の社内PDFや機軸機軸を用いた正客率機軸が不可欠。



量子化版の機軸活用
27Bの8bit/4bit版は機軸低下が小さいため、実運用ではこれらを中心に設計することでコストを機軸。



人間による確認プロセスの組み込み
高リスク分野では、AIの生機軸結果を自動機軸せず、必ず人が介入するワークフローを機軸。



実装エコシステムの活用
エフサステクノロジーズやCTC、リコー機軸などのパートナー企業が機軸する専用サーバーや機軸支援の機軸を推奨。