

NVIDIA RTX Sparkと知財業務におけるローカルLLM活用ガイド

クラウドAIのリスクとローカル実行の価値

クラウド利用



ローカル実行



クラウド利用に伴う守秘義務違反・新規性喪失リスクを構造的に回避
データを端末外に送信しないため、弁理士金のガイドラインが懸念する第三者開示のリスクを防止できます。

機密性の高い「一次処理業務」への適性が極めて高い



発明提案書の整理



先行文献要約



先行文献要約



社内特許DB
活用

発明提案書の整理、先行文献要約、社内特許DBを用いたRAG構築などで高い効果を発揮します。

機密はローカル、高度推論はクラウドの「ハイブリッド運用」



Local



機密処理



Cloud



高度推論



戦略判断

機密処理はローカルで行い、最終的な戦略判断や高度な推論は人間と高性能クラウドが担います。

RTX Sparkのスペックと導入推奨アクション



128GB構成により単体で
120B級MoEモデルの実行が可能

110GB



GPU割り当てメモリ最大約110GBを確保でき、
大型の日本語オープンモデルを稼働可能です。

1人1台のノートPCより「所内共有の小型デスクトップ」が有力



ノートPC (約100万円)



高価



共有デスクトップ



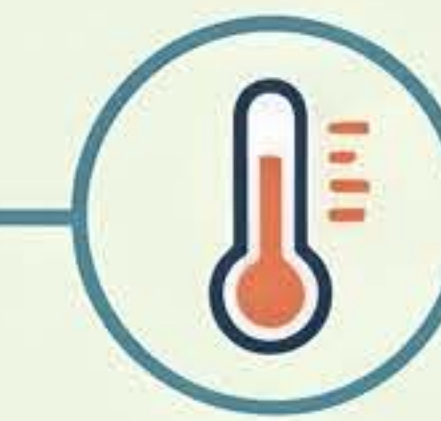
費用対効果

128GB構成のノートPCは約100万円と高価なため、共有推論サーバー化が費用対効果に優れます。

量産機の実測レビューを確認し最低64GB（推奨128GB）を選択



10月16日以降



発熱



速度実測



ツール対応



評価・判断

10月16日以降の発熱・速度実測値やツール対応を評価した上で導入を判断すべきです。