知財業務におけるLLM上位・中位モデル使い分け戦略

モデル選択を左右する3大特性

階層と性能の逆転

中位モデルが上位の性能を凌駕する場合あり。

Claude系

Fable 5.2 / Opus 5.5

誤答の割合(試算):
約13%〜17%

強み・得意分野:
総合指数首位・
長文ナレッジ業務

GPT系

Astra / Sol

誤答の割合(試算):
約19%〜21%

強み・得意分野:
科学系タスク・
作業手順構築

Gemini系

Carbon / Neon

誤答の割合(試算):
約5%〜6%

強み・得意分野:
米国法務リサーチ・
低ハルシネーション

単価差と費用対効果

費用差2〜4倍に対し
総合性能差は優差。

系統別の誤答パターン:
Claude・GPTは推測し、
Geminiは棄権する。

実務展開の推奨運用パターン

中位起点のカスケード



中位起点のカスケード

日用業務は中位で処理し難所のみ上位へ昇格。

事実照会の原文検証



事実照会の原文検証

判例や条文照会は原文添付の検証が必須。

異系統による相互検証



異系統による相互検証

重要な成果物は別系統LLMでクロスチェック。

Gemini Notebook