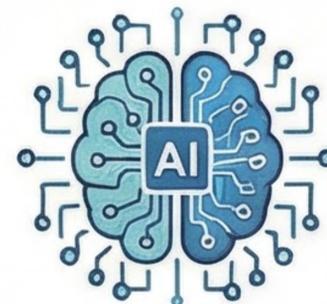


知財実務におけるLLM導入の新基準：ARC-AGI-2による性能差と活用戦略

ARC-AGI-2スコアの劇的な進化

従来LLM
0%
LLMの推論能力



高性能LLM (Gemini 3)
84.6%

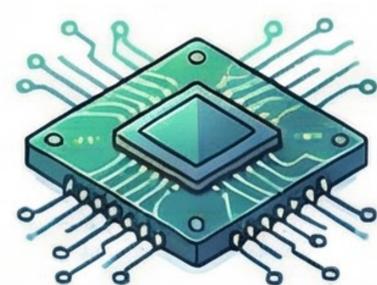
流動性知能（未知の課題への適応力）が知財の中核プロセスを支える

推論能力の差がもたらす知財タスクへの影響

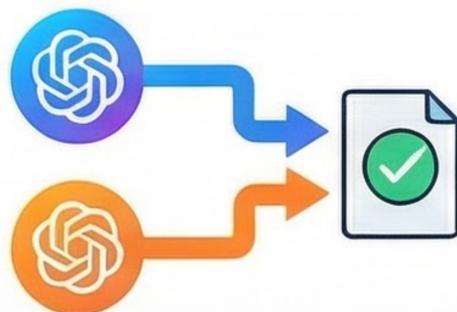
高性能LLM (想定)	指標	低性能LLM (想定)
ARC-AGI-2 スコア: 50% ~ 85% 程度		ARC-AGI-2 スコア: 0% ~ 5% 程度
 侵害分析	知財タスクの例	 抽出
 要素分解・仮説生成		 抄録作成
 請求項解釈		 形式整形
人的レビュー負荷: 高 (整合性確認が必須)	人的レビュー負荷	人的レビュー負荷: 非常に高 (誤り混入が多いため)

タスクの難易度に応じたモデルの使い分けが推奨される

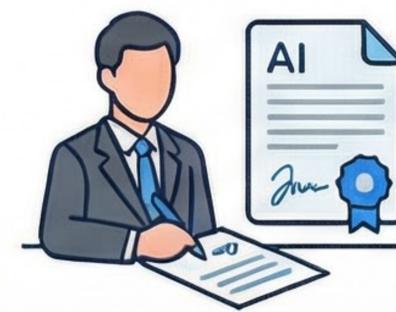
リスク管理と導入・運用の指針



高性能モデル特有の「巧妙な幻覚」
整合性のある虚偽を生成するため、形式チェックだけでは見逃す危険がある。



マルチホーミングと二重化レビュー
タスクを動的ルーティングし、別システムモデルでの反証探索を組み込む。



弁理士による最終責任の原則
AI出力の正確性は保証されず、人間による確認責任 (HITL) が前提条件。