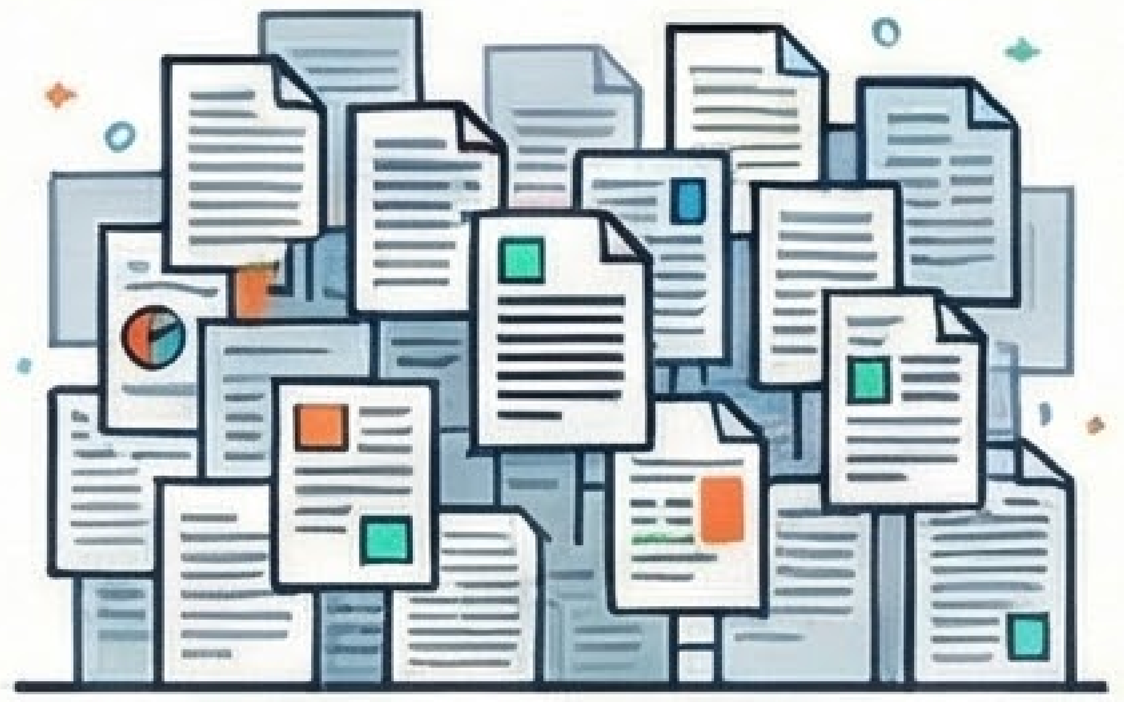
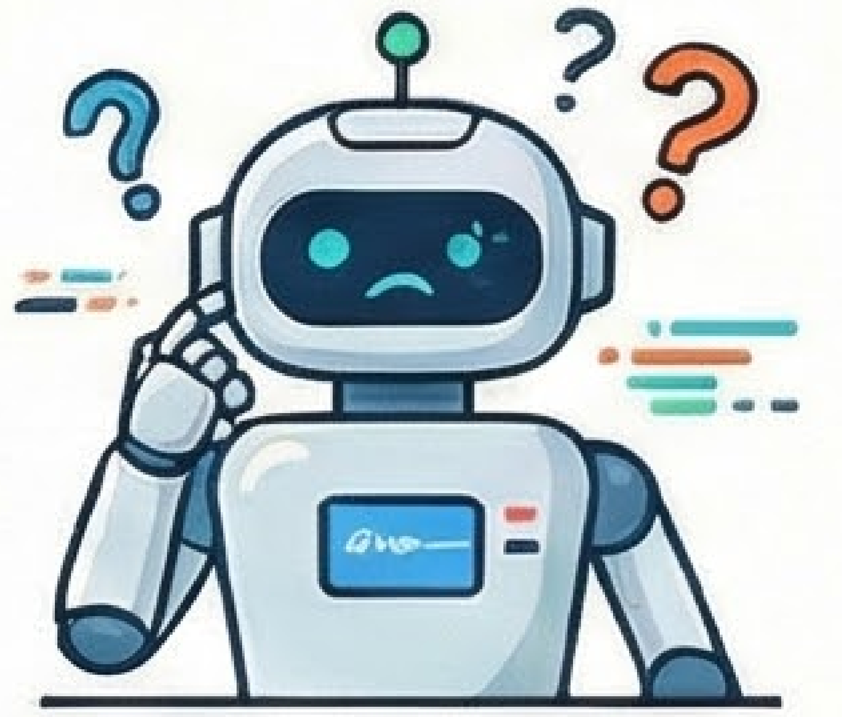
知財AIの実力検証：PatentBenchが示す専用AIの優位性

知財調査の「壁」と汎用AIの限界



特許調査は退屈な業務

100件以上の特許公報を精査するには数日～数週間を要し、熟練者でも重要文献を見落とすリスクが常に存在する。



汎用AIは特許実務に不向き

一般テキストで学習した汎用LLMは、特許特有の専門用語やクレーム文本を理解させず、誤回答は深刻な法的リスクに直結しかねない。

公平で実践的な評価設計

PatentBenchとは？

知財業務に特化した全評価の特別的ベンチキーク、AIの特許調査能力を客観的かつ容量的に評価するためにPatsnapが開発。

公平で実践的な評価設計

約840件の多様な技術分野・言語（英語60%、中国語22%）の最新特許案件をテストデータとし、AIが学習素をすない新規の課題で性能を測定する。

新たな評価基準「PatentBench」の登場

評価指標1：Xヒット率

約840件の多数な技術分野・言語（英語60%、中国語22%）の最新特許案件をテストデータとし、AIが学習素をすない新規の話題で性能を測定する。

評価指標1：Xヒット率

調査対象のうち、少なくとも1件の最重要先行文献（X文献）を発見できた案件の割合。「有効な文献を見つけ出す能力」を示す。

評価指標2：Xリコール率

正確となる全X文献のうち、AIが発見できた文献の割合。「どれだけ網羅的に（漏れなく）探せたか」を示す。

衝撃のベンチマーク結果：専用AI vs 汎用AI

専用AI「Patsnap Eureka」が汎用AIを圧倒

Patsnap Eureka（専用AI）

Xヒット率（有効文献の発見率）	Xリコール率（網羅率）
81%	36%

新規性調査において、Patsnap Eurekaは有効文献の発見率（ヒット率）と網羅性（リコール率）の両方で汎用LLMを大きく上回った。

汎用AI
(ChatGPT-o3, DeepSeek-R1)

	Xヒット率	Xリコール率
ChatGPT-o3	32%	11%
DeepSeek-R1	9%	3%

高性能の秘密：Patsnap Eurekaの仕組み

技術的特徴の抽出

入力された発明内容からキーポイントを抽出し、検索の土台を構築する。

クレーム要素の分解

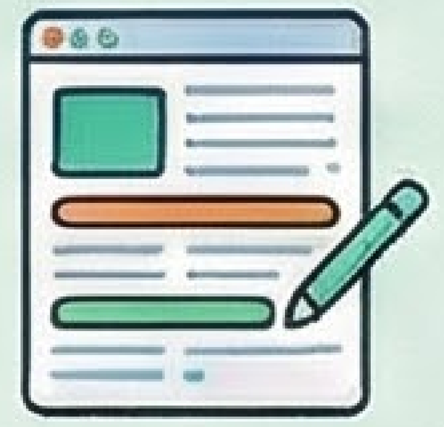
特許クレームを構成要素ごと分解し、評価すべきポイントを明確化する。

複合検索式の自動生成

キーワード、特許分類、質問の構成要素などを組み合わせた重層な検索クエリを生成する。

検索とスクリーニング

世界中の特許・論文DBから関連文献を抽出し、関係性に応じ候補を絞り込む。



クレーム対応付け（マッピング）

先行文献のどの部分が発明の新規性を否定するかを性能でハイライトし、対応関係を示す。

比較表・レポートの作成

調査結果を先行技術との対比表や報告書の形式で自動作成し、専門家の判断を助ける。

実務導入への道：可能性と課題

導入のメリット

調査効率の飛躍的向上、グローバルな多言語検索への対応、専門家が高付加価値業務へ集中できる環境の実現。

導入時の課題

AIの結果を鵜呑みにできない信頼性の問題（リコール率36%）、判断根拠の説明可能性、機密情報を守るセキュリティ対策が不可欠。

人間との協働が鍵

AIは専門家を代替するのではなく、強力なアシスタント。AIが下調べを行い、人間が最終判断と戦略立案を担うという協働が重要になる。