

PatentBenchとPatsnap Eureka：知財AIの実力を測る 新指標と実務の未来

汎用LLM対比における「証拠に基づく」
特許調査AIの有効性検証



要旨：なぜ今、知財専用AIが必要なのか



定量的な証拠

Patsnap社が公開した評価基準「PatentBench」において、専用AI「Eureka」は新規性調査のX文献検出率81%を達成。対する汎用モデル（ChatGPT-o3）は32%に留まる。



構造的な違い

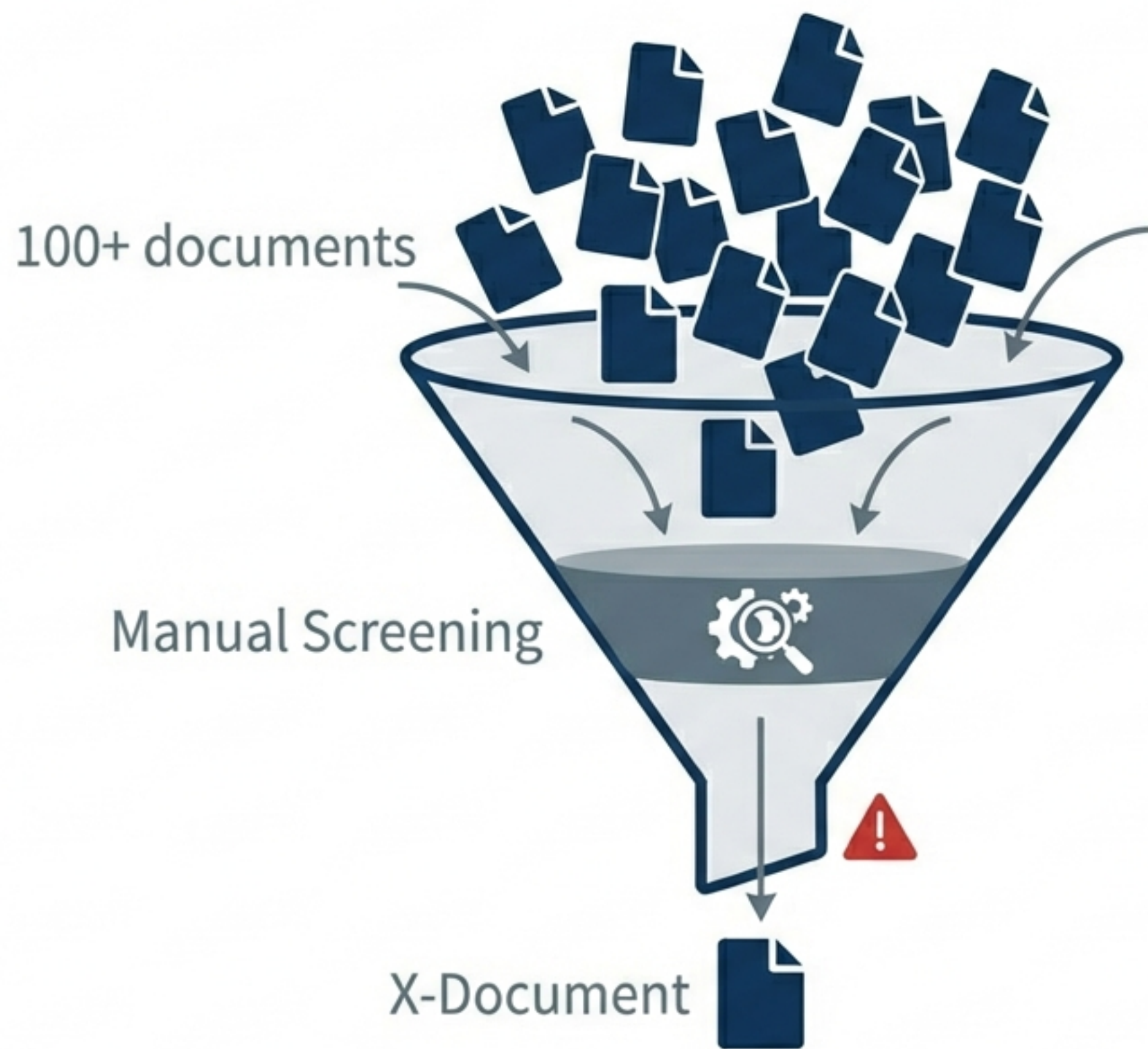
汎用LLMは「ハルシネーション（もっともらしい嘘）」や専門知識の欠如により、特許調査の致命的リスクとなる見落としが発生しやすい。



実務への影響

Eurekaは調査プロセスの自動化（検索・スクリーニング・マッピング）により、実務者が戦略立案に集中できる「Human-in-the-Loop（人間とAIの協働）」環境を実現する。

知財調査における「壁」と現行プロセスの課題



膨大な作業量：

1件の調査で100件以上の公報精査。数日から数週間を要する。

専門用語の壁：

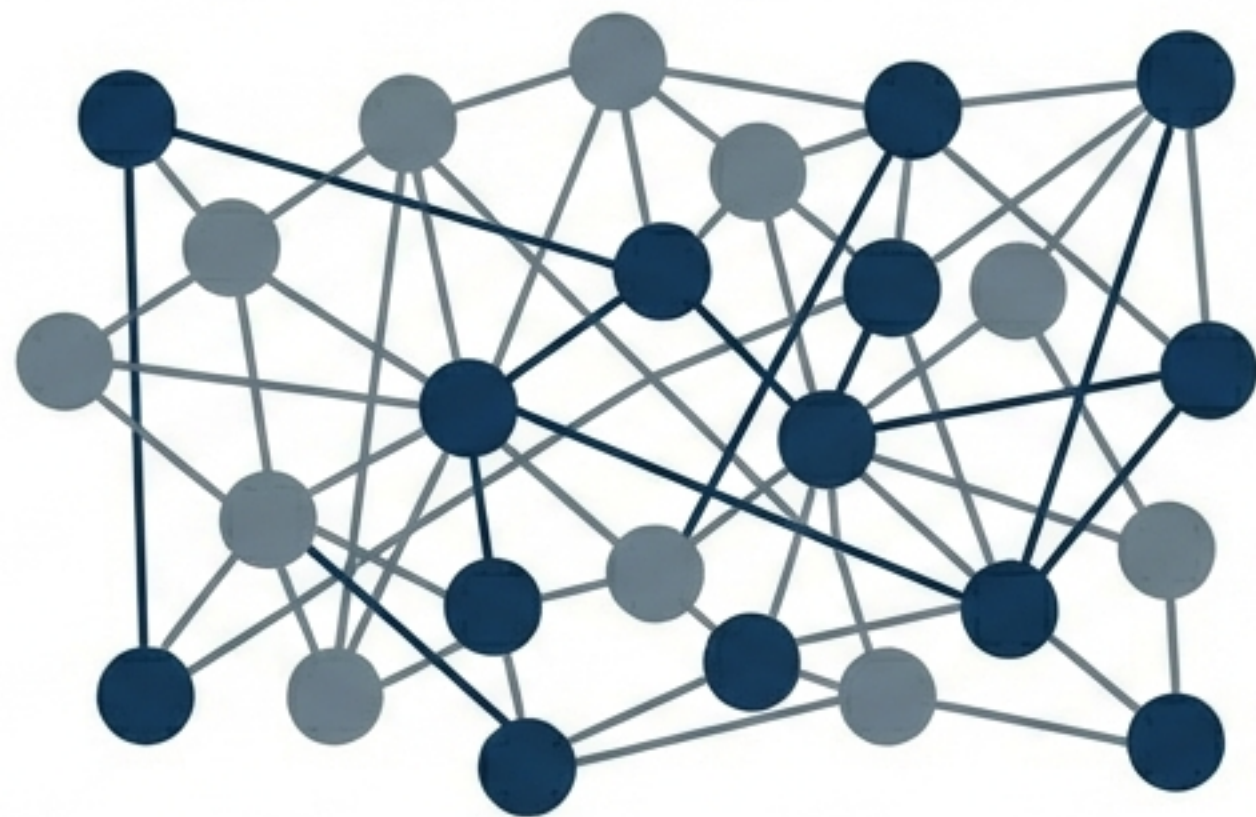
特許独特のセマンティクスや用語のゆれが検索を困難に。

人的リスク：

疲労や時間制約によるクリティカルな文献の見落とし。

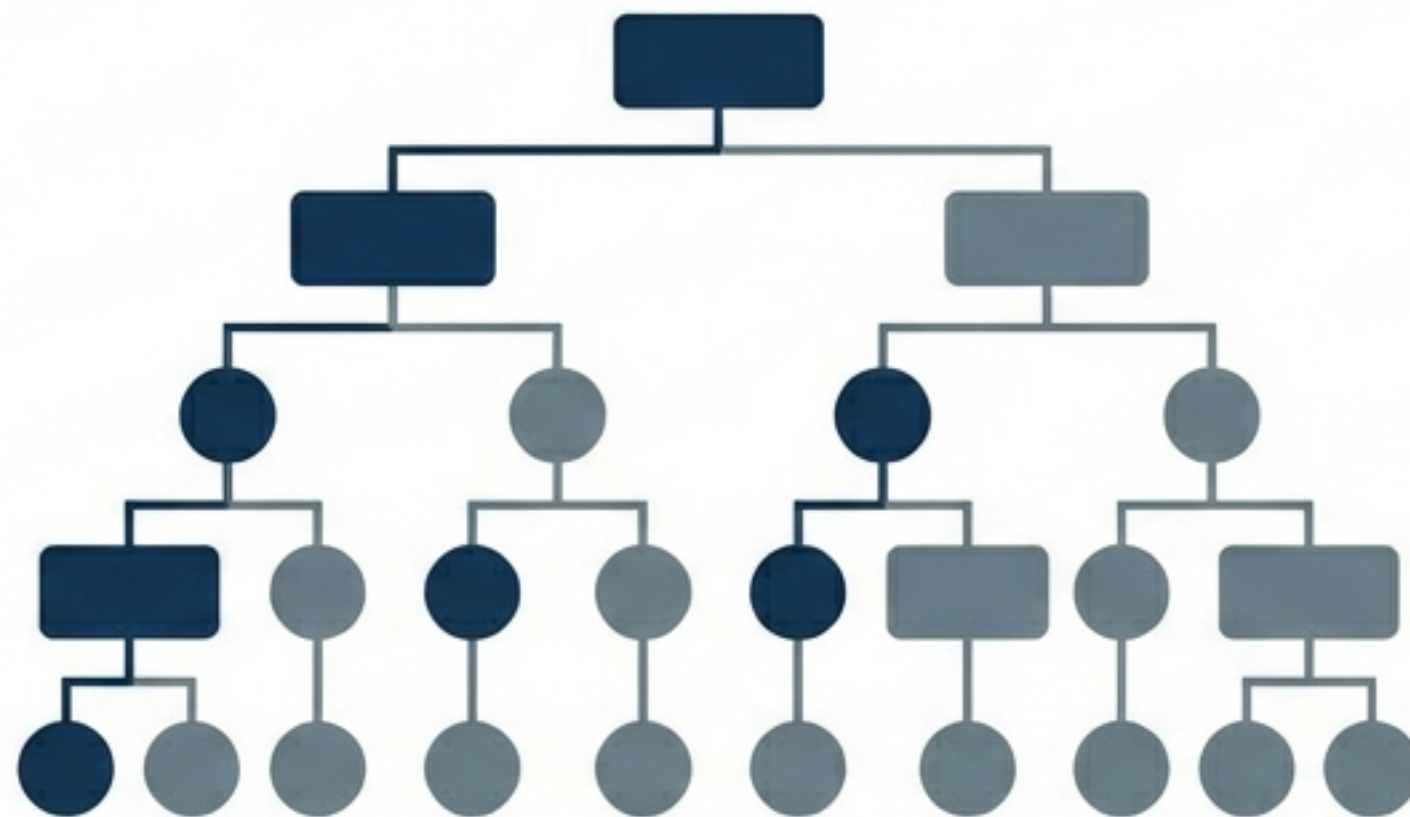
汎用LLM（ChatGPT・DeepSeek）の限界

General Web Data



学習データの質：ウェブ上の一般テキストで学習。特許特有の論理構造を理解していない。

Patent Specific Logic



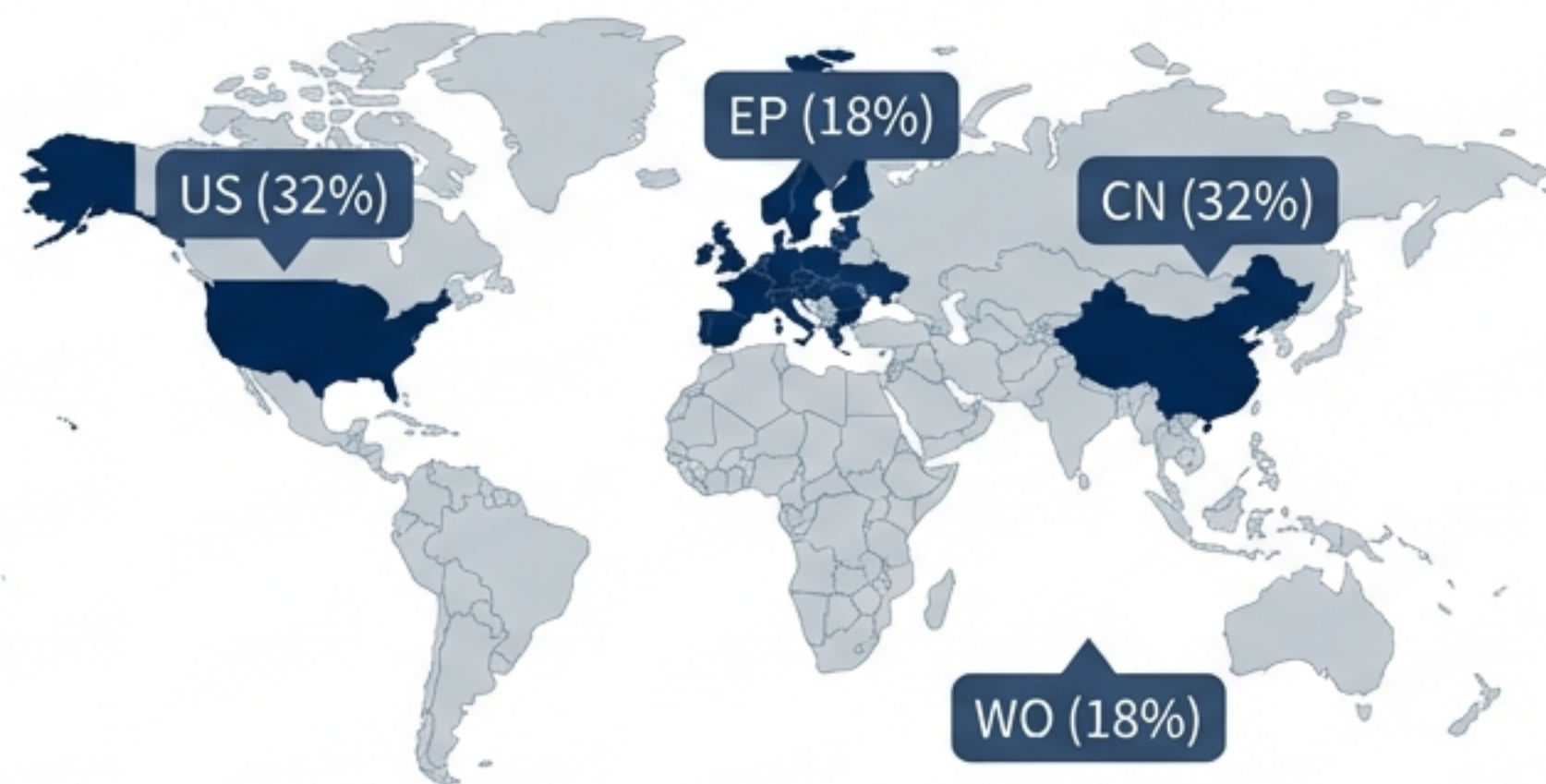
ハルシネーションの危険性：もっともらしい嘘は、法的判断において致命的なリスクとなる。

結論：「会話ができるAI」と「実務を遂行するAI」は異なる。
汎用モデルは論理的整合性が求められるタスクには不向きである。

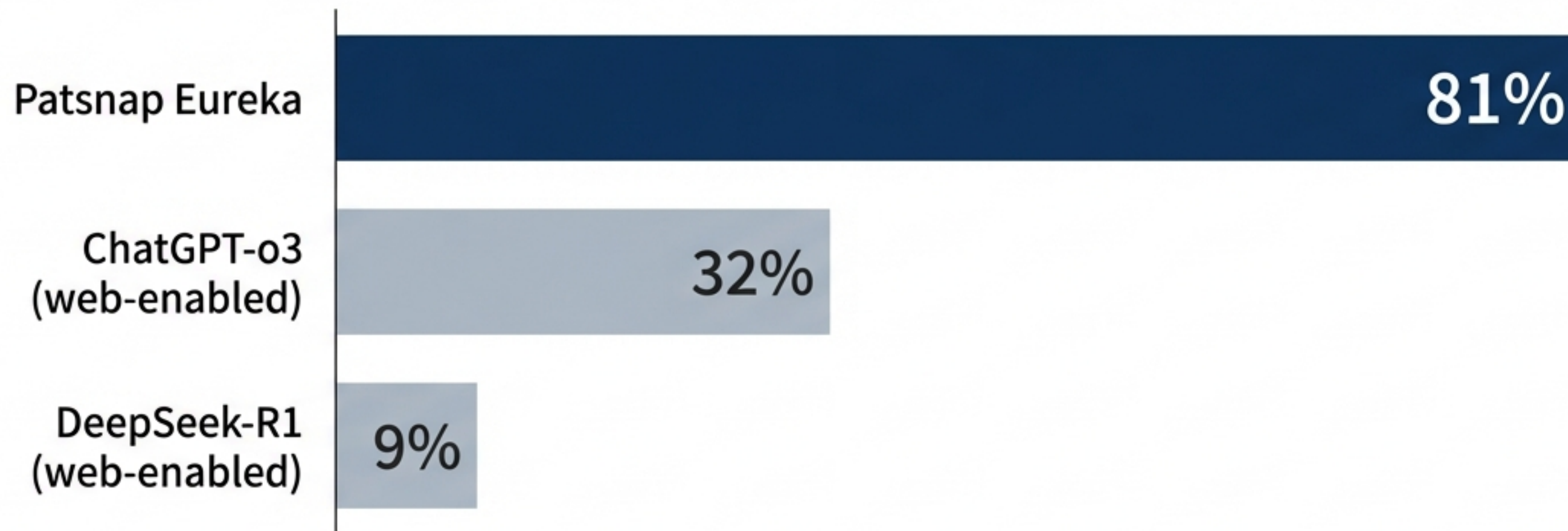
PatentBench：世界初の知財特化型AI評価指標

- **目的**：新規性調査（先行技術調査）におけるAI能力の客観的測定。
- **データセット**：約340件の多言語・多地域サンプル。
- **評価手法**：実際の審査で引用された「X文献（新規性を否定する文献）」を標準回答とし、AIの発見率を判定。

Evaluation Scope: Multi-lingual & Multi-jurisdiction



新規性調査における圧倒的な格差：Xヒット率

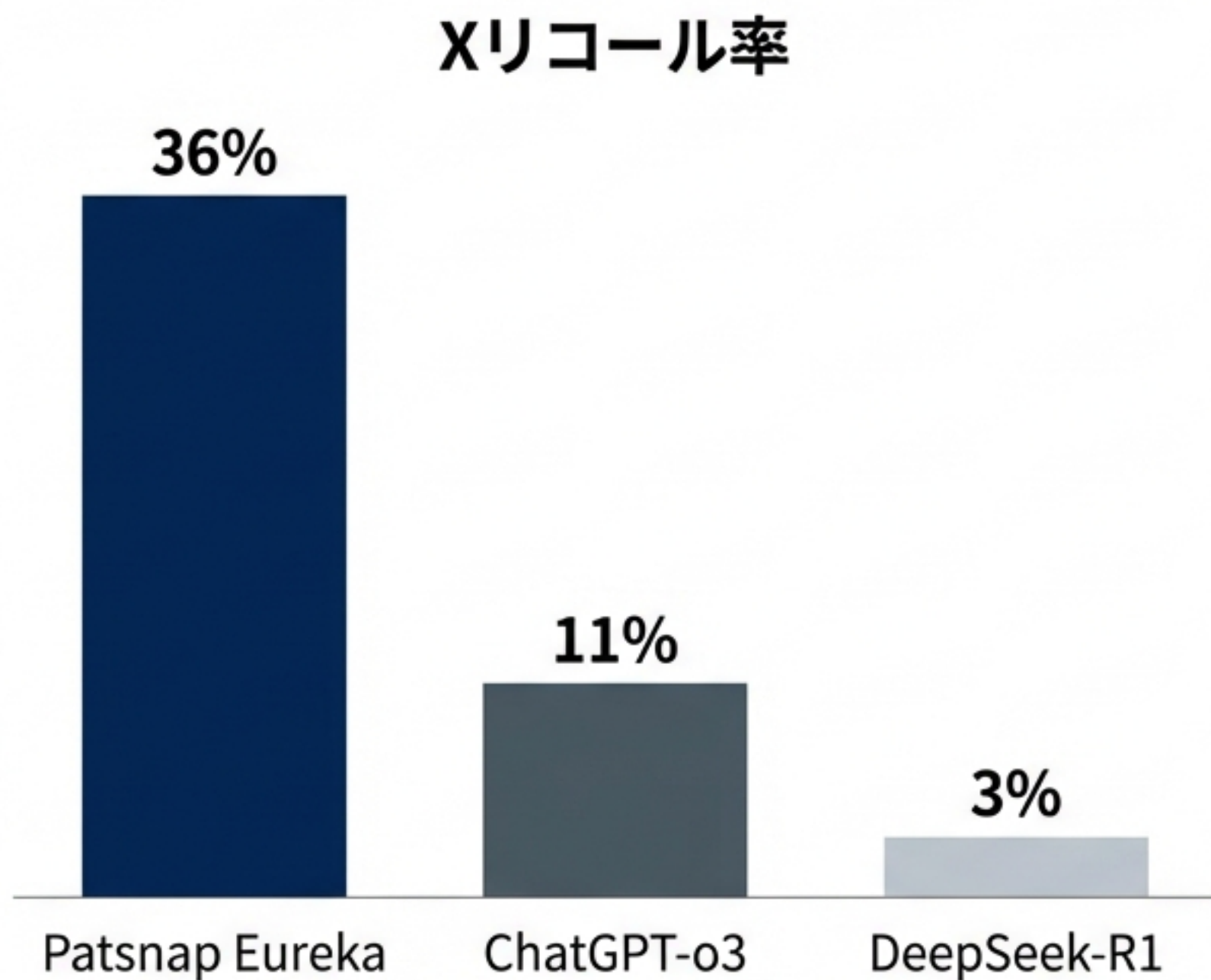


Noto Sans JP Regular

指標の定義：Top100の検索結果中に、少なくとも1件の「X文献」が含まれる割合。

インサイト：Eurekaは5回中4回以上の確率で先行技術を発見。DeepSeekは1割未満。

網羅性の検証：Xリコール率（X Recall Rate）



指標の定義：標準回答に含まれるX文献総数に対し、AIが拾い上げることができた文献数の割合。

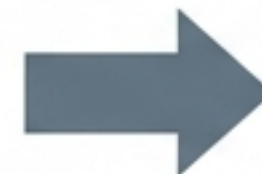
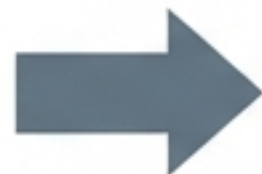
インサイト：汎用LLMは約9割の関連文献を見落とす（リコール率11%以下）。Eurekaは3倍以上の網羅性を示し、組み合わせ検討に優位。

具体事例による検証：見落としの現実的リスク

AI Model	Found Families	Missed Families	Recall Rate
Patsnap Eureka	4 Families (Full)	0	100%
ChatGPT-o3	1 Family	3 Families (Missed)	25%
DeepSeek-R1	0 Families	4 Families (Missed)	0%

汎用LLMは、深い専門知識と検索ロジックが要求される特許調査で致命的な「検索漏れ」を起こす。

なぜEurekaは高精度なのか： 領域特化型エージェントの仕組み



独自データ基盤
(2億件以上の特許・論文)

専門ロジック
(審査官の判断基準)

RAG + RAT技術
(ハルシネーション抑制)

高精度アウトプット

専門家によるクリーニング済み

思考プロセス強化による推論

調査プロセスの完全自動化と可視化

1

構成要件の分解

クレームを自動的に構成要素に分解し、技術的特徴を抽出。

2

複合検索式の生成

キーワード、セマンティック、特許分類（IPC/CPC）を組み合わせた高度なクエリ作成。

3

マッピング

発見した先行文献の記載が、発明のどの構成要件に対応するかを自動ハイライト。

4

スクリーニング

重複排除とランク付けを行い、ノイズを低減。

説明可能性（Explainability）と実務直結のアウトプット

Claim Chart

Claim Element	Prior Art Evidence
[Component A] ...	[Excerpt from Patent X] ...
[Component B] ...	[Excerpt from Patent Y] ...
[Component C] ...	[Excerpt from Patent Z] ...

- クレーム対比表の自動生成：新規性判断の根拠が一目でわかる。
- 編集可能なレポート：Word/Excel形式。わずかな修正で公式文書化。
- ブラックボックスの解消：根拠の可視化により検証が容易。

人間とAIの協働モデル：Human-in-the-Loop



AIは人間を代替するのではなく、能力を拡張する「増幅器」として機能する。

セキュリティ体制と公的機関の導入トレンド



Enterprise Security

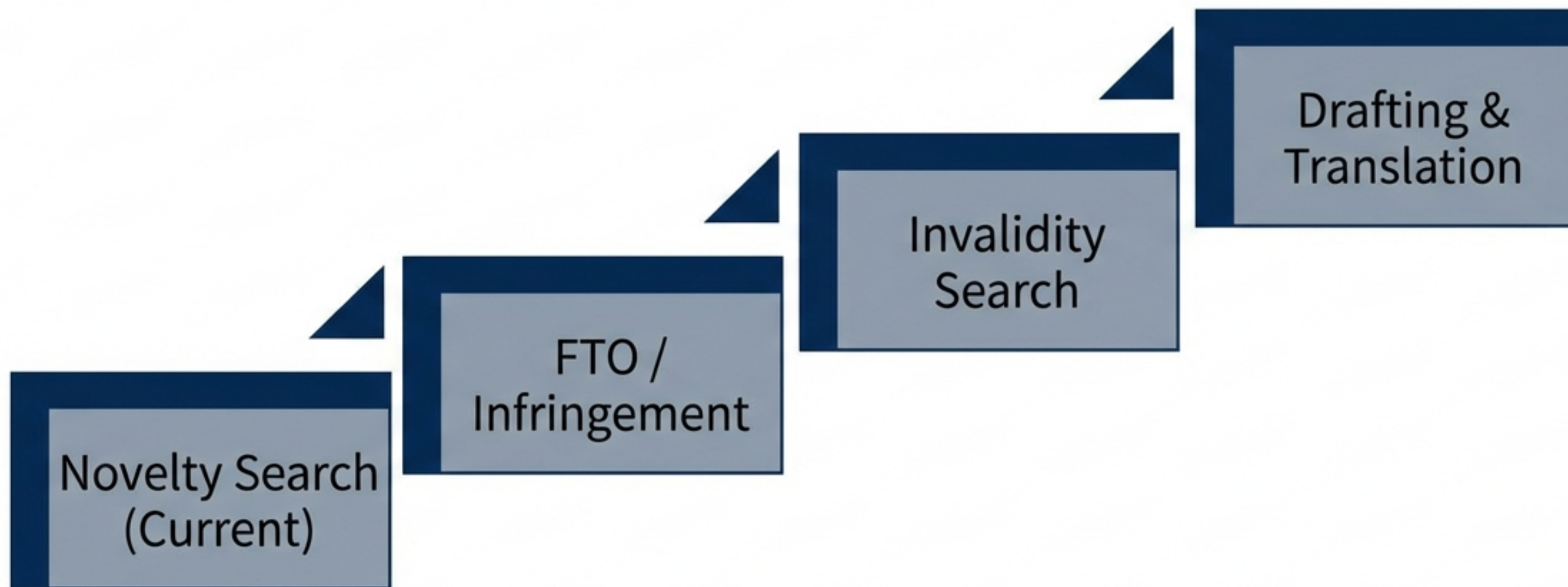
- ISO 27001認証取得。
- オンプレミス（自社サーバー）導入対応。
- データ隔離環境（学習への利用なし）。



Public Sector Adoption

- USPTO（米国特許商標庁）：AIによる先行技術検索パイロットを開始。
- JPO（日本特許庁）：検索補助におけるAI活用を推進。

知財AIの今後の展望とロードマップ



ツールの淘汰：「AIを使うか」から「どの専用AIを使うか」へ。
PatentBenchのような客観指標が選定の鍵となる。

結論：AIを味方につけ、知財実務を次世代へ



1. 汎用AIのリスク

一般的なLLMを知財調査に使うことは、精度とセキュリティのリスクが高い。

2. 専用AIの必然性

PatentBenchデータは、Eurekaが実務レベルの精度（Hit率81%）に達していることを証明した。

3. プロフェッショナルの進化

信頼できるパートナーとしてのAIを選定し、よりクリエイティブで戦略的な業務へ。