



# PatentBenchとPatsnap Eurekaによる知財AIの評価と実務展望

## 概要

特許領域に特化したAI評価基準「PatentBench」がPatsnap社により発表され、AIの特許調査能力が初めて定量的に比較されました<sup>1</sup>。PatentBenchは新規性調査（先行技術調査）という特許実務の中核タスクで、PatsnapのEureka新規性調査エージェントと汎用モデル（ChatGPT-o3、DeepSeek-R1）の性能を同一条件で評価しています<sup>2</sup>。その結果、Patsnap EurekaはX文献（新規性を奪う先行文献）の検出率81%、リコール率36%を達成し、汎用モデル（検出率32%/9%、リコール率11%/3%）を大きく上回りました<sup>3</sup>。このリポートでは、PatentBenchの設計手法と評価指標、ベンチマーク結果から見えるAI支援の精度と限界、および知財専用AIエージェント「Patsnap Eureka」の機能と実務導入の可能性・課題について、知財実務の専門家向けに詳細に解説します。特に、AIによる新規性調査やクレーム解釈支援の現状と展望、そして特許庁や企業知財部門がこうしたAIを活用する際のポイントを考察します。

## PatentBenchとは何か：知財専用AIベンチマークの背景

生成AI・大規模言語モデル（LLM）の進化により、文章生成や要約・翻訳といった一般タスクでAIの高性能が示されています。しかし、知財業務——例えば先行技術文献の網羅的検索（新規性調査）やクレーム解釈といった特許実務の中核業務をAIが担えるのかについては、未だ疑問が残ります<sup>4</sup>。特許文献は高度に専門的かつ曖昧な表現が多く、企業や分野ごとに異なる用語ゆれもあるため、熟練者でも重要文献の見落としリスクが常につきまといます<sup>5</sup>。100件以上の特許公報を精査するには数日～数週間を要するのが実情であり、これがイノベーションを支える知財調査の「壁」となっています<sup>5</sup>。

近年、多くの企業がこの過酷な業務負荷を軽減すべく生成AIの導入を検討しています。しかし、大半の汎用LLMはウェブ上的一般テキストから学習しており、「特許」という特殊な世界は全く別のルールで動いています<sup>6</sup>。汎用モデルは特許独特の専門的セマンティクスやクレーム体系、知財固有のワークフローを十分に理解できず、特許実務における判断を伴う作業には構造的な限界があります<sup>7</sup>。実際、汎用LLMが生成したもっともらしい誤回答（ハルシネーション）を鵜呑みにすれば、単なるミスでは済まされず、製品戦略の誤りや深刻な法的リスクに直結しかねません<sup>8</sup>。このような背景から、「会話ができるAI」ではなく業務文脈を理解して実務を遂行するAIへの期待が高まっています<sup>9</sup>。Patsnap社はこの課題に真正面から取り組み、知財実務専用のベンチマーク「PatentBench」を構築することで、AIの可能性と限界を客観的に示そうとしています<sup>10</sup><sup>11</sup>。

PatentBenchは、知財業務にフォーカスした世界初の包括的ベンチマークであり、特許AIが備えるべき核心能力を測定するよう設計されています<sup>12</sup>。特に第1弾として、特許実務でも最も難易度の高いタスクの一つ「新規性調査」にAIがどこまで有用かを検証する目的で作られました<sup>13</sup>。PatentBenchは単なる実験ではなく、実際の特許業務でAIが直面する課題（先行文献検索、特許クレームの読解、新規性の判断など）に基づいて設計されている点が特徴です<sup>14</sup><sup>15</sup>。これにより、特許領域におけるAI活用が「興味本位の模索」から「測定可能な進歩」へと進化する礎を築くことを目指しています<sup>16</sup>。

## PatentBenchの設計と評価手法

ベンチマーク手法として、PatentBenchでは専門家が厳選した多数のテストサンプルを用意し、それぞれに「テスト質問（検索課題）」と「標準回答（想定すべき先行文献集合）」が設定されています<sup>17</sup>。新規性調査では本来「完全な正解集合」を定義することは難しいため、実際の審査で引用されたX文献（新規性を否定する先行技術文献）などを基に「理想解に近い回答セット」を構築するアプローチが取られました<sup>18 19</sup>。各サンプルは現実の特許出願を模した問題文（発明の説明やクレーム）を入力とし、AIが提案した先行技術文献リストを、専門家が用意した「標準回答」の文献集合と比較して評価します<sup>17 20</sup>。

データセットの構成にも工夫があります。統計的に有意な評価を行うため、数百件規模（現在約340件）のサンプルが用意されており、技術分野や地域・言語の偏りを避けるよう設計されています<sup>21</sup>。例えばIPC分類（国際特許分類）では主流技術からニッチ分野までバランスよく分布し、言語も英語68%・中国語32%と多言語対応<sup>22</sup>。出願官庁も米国特許庁(US)・中国(CN)が各32%、欧州特許庁(EP)・WIPO出願(WO)が各18%と主要管轄を網羅しています<sup>23</sup>。さらに、評価サンプルはAIモデルの学習済み知識と被らないようごく近年の出願案件から選ばれています<sup>24 25</sup>。こうした多様で新しいデータに基づき、各AIツールが未知の発明に対してどこまでの確に先行技術を探せるかを試す構成です。

評価指標としては、特許実務に即した2つの軸が設定されました<sup>26</sup>：

- Xヒット率（X Hit Rate）：各テストでAIが少なくとも1件は「標準回答」に含まれるX文献を検索上位にヒットさせた割合<sup>27</sup>。要するに「該当クエリで何割のケースで有効文献を見つけられたか」を示す指標です<sup>28</sup>。特許審査や無効資料調査では「せめて1件でもクリティカルな文献を見逃さない」とが重要なため、このヒット率は意思決定の迅速化に直結する指標と位置付けられます<sup>27</sup>。
- Xリコール率（X Recall Rate）：標準回答に含まれる全X文献数に対し、AIの検索結果リスト（上位100件以内）に含まれたX文献数の割合<sup>27 29</sup>。これは「模範解答の何割をAIが拾えたか」を示し、どれだけ漏れなく包括的に探せたかを評価します<sup>27 29</sup>。新規性判断では複数の先行技術を組み合わせて検討するケースもあるため、関連文献を網羅的に収集できるかを測るこのリコール率は特許クレームの精緻な分析や強い特許クレーム作成に不可欠とされています<sup>30</sup>。

評価は各AIモデルについて、提供した検索結果の上位100件を対象に、このXヒット率とXリコール率を算出する形で行われました<sup>3</sup>。なお、Xヒット率はトップ1件・3件・5件でのヒット状況も細かく分析されており、上位数件の結果にどれだけ有効文献が含まれるか（精度の指標）も検討されています<sup>31</sup>。総じて、PatentBenchはサンプル設計から評価軸に至るまで、実務に直結する有用性を多面的に測れるよう工夫されています<sup>32 33</sup>。

## PatentBenchでの評価結果：専用AIと汎用モデルの比較

PatentBenchによる新規性調査ベンチマークの結果、Patsnap Eurekaの新規性調査AIエージェントが他の汎用AIモデルを大きく凌駕する性能を示しました<sup>3</sup>。以下の表に、各モデルのXヒット率およびXリコール率をまとめます<sup>3</sup>。

モデル	Xヒット率 (Top100内にヒット)	Xリコール率 (Top100内網羅率)
Patsnap Eureka 新規性調査エージェント	81% <sup>3</sup>	36% <sup>3</sup>

モデル	Xヒット率 (Top100内に ヒット)	Xリコール率 (Top100内 網羅率)
ChatGPT-o3 (汎用LLM + 検索)	32% <small>3</small>	11% <small>3</small>
DeepSeek-R1 (汎用LLM + 検索)	9% <small>3</small>	3% <small>3</small>

3 ※Xヒット率：Top100の検索結果中に「標準回答のX文献」が1件以上含まれる割合、Xリコール率：標準回答に含まれるX文献総数に対しTop100中に含まれる割合。

図1：PatentBenchにおける各AIモデルのXヒット率比較（青：Patsnap Eureka、紺：ChatGPT-o3、灰：DeepSeek-R1）3

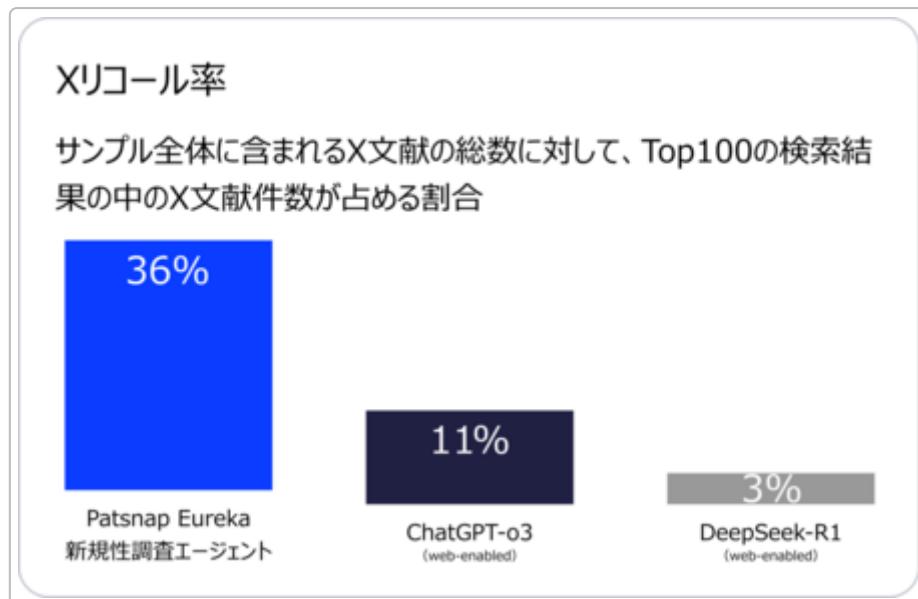


図2：PatentBenchにおける各AIモデルのXリコール率比較（青：Patsnap Eureka、紺：ChatGPT-o3、灰：DeepSeek-R1）3

Patsnap Eurekaは全テストケースの81%で少なくとも1件の有効文献を上位結果に含め（Xヒット率）、標準回答のX文献全体の36%を拾い上げました（Xリコール率）3。これに対し、汎用LLMを用いたChatGPT-o3はわずか3割強（32%）のケースでしか有効文献を見つけられず、DeepSeek-R1に至っては1割未満（9%）という低いヒット率でした3。網羅性を示すリコール率でも、ChatGPT-o3は11%、DeepSeek-R1はわずか3%に留まり、汎用モデルでは新規性を否定する重要文献の大半を見落としてしまう現状が浮き彫りとなりました3。これは、特許専門のデータとロジックで最適化されたAIと、一般知識のLLMとの差が如実に現れた結果と言えます。

特筆すべきは、Patsnap Eurekaエージェントが検出率・網羅率の双方で他モデルを圧倒した点です3。Xヒット率81%ということは、5件中4件以上の割合で有効先行文献を発見できたことを意味し、新規性調査の初期段階でキードキュメントを捉える能力が飛躍的に高いことを示しています<sup>27 28</sup>。またXリコール率36%は、標準回答に含まれる複数の関連文献のうち約1/3を自動検索でカバーできた計算です【22†】。汎用モデルではこの値が1割前後に過ぎないため、Eurekaはより多面的に先行技術の網を張り巡らせ、漏れを大幅に減らしていると解釈できます<sup>3</sup>。実務上、新規性の判断には一件だけでなく複数文献の組合せ検討も重要な場合があるため、幅広く関連文献を拾える能力は非常に有用です<sup>30</sup>。

ベンチマーク中の具体例として、あるテストでは**発明の明細（問題文）**を入力として各AIが100件の検索結果を提示し、それを標準回答のX文献集合と比較しました<sup>34</sup>。Eurekaエージェントは標準回答に含まれる**4つの関連特許ファミリーすべて**をトップ100内に検出し、Xヒット率100%・Xリコール率100%という完璧な結果を残しました<sup>34</sup>。一方、ChatGPT-03もトップ100内に標準回答文献を含めたものの1ファミリーのみ（他3件は漏れ）で、リコール率25%にとどまりました<sup>35</sup>。DeepSeek-R1は残念ながら該当文献を一つも発見できず、リコール率0%という失敗例も見られました<sup>35</sup>。このように**個別ケースでも汎用LLMは見落としが多く**、深い専門知識や検索ロジックが要求される特許調査には不向きであることが示唆されます<sup>36</sup>。

以上の結果から、汎用LLMは推論や文章生成では卓越していても、**特許の新規性調査といった高度に専門化されたタスクでは苦戦することが明らかです**<sup>36</sup>。一方、Patsnap Eurekaのようにドメイン特化したAIツールは、精度・再現性の両面で優れたパフォーマンスを発揮し、**特許ワークフローに不可欠な精度と網羅性を満たし得ることが証明されました**<sup>36 37</sup>。これは、特許専用に調整されたAIが、汎用モデルには困難なクレーム文言の解釈や技術的特徴の抽出、巧妙な検索戦略を習得しているためであり、専門領域では汎用AIに代わり専用AIが**本領を発揮することを示すエビデンスとなっています**<sup>36 38</sup>。

## 知財調査AIエージェント「Patsnap Eureka」の機能と実務応用

PatentBenchで高精度を示した**Patsnap Eureka**とは、特許・知財業務に特化して開発されたAIエージェント群です<sup>9</sup>。Patsnap Eurekaは単一の対話型AIというより、知財・R&D分野の**複数の専門エージェント**から構成されるプラットフォームであり、具体的には「新規性調査エージェント」「FTO（特許侵害可避性）調査エージェント」「明細書ドラフティング支援エージェント」「特許翻訳エージェント」等のモジュールが実装されています<sup>39</sup>。これは、汎用LLMが直面する幻覚（ハルシネーション）や実行力の限界を克服するために、**タスク固有の知識と手順を組み込んだ「領域特化型AIエージェント」**が有効であるとの考えに基づくもので<sup>9 6</sup>。Patsnap Eurekaはその先駆けとして登場し、特許実務で頻発する高負荷タスクを自動化・高速化することで専門家を支援することを目指しています<sup>9</sup>。

中でも**新規性調査エージェント**は、特許調査業務における一連のステップをほぼエンドツーエンドで自動化する点が画期的です<sup>40</sup>。具体的な機能を段階的に見ると、以下のようにになります<sup>40</sup>：

- ・**先行技術抽出**：入力された発明の明細やクレームから、発明のキーポイントや技術的特徴を抽出します。技術分野や関連キーワードを解析し、検索クエリ作成の下地を構築します。
- ・**クレーム要素の分解**：特許クレームを構成要素に分解し、各要素ごとに新規性評価すべきポイント（**発明特有の技術的特徴**）を洗い出します。これにより検索のターゲットを明確化します<sup>41</sup>。
- ・**複合検索式の生成**：抽出したキーワード・要素を組み合わせ、特許データベース向けの高度な検索クエリを自動生成します<sup>41</sup>。キーワード検索だけでなく、**語義や技術概念の類似**まで捉えるセマンティック検索や、特許分類（CPC/IPC）情報も活用した複合検索式を構築します。
- ・**検索・スクリーニング**：世界中の特許・論文データベースから関連文献を検索し、数万件の候補から**上位有力文献**を絞り込みます。PatsnapのAIは**大規模言語モデル+RAG（Retrieval-Augmented Generation）**技術で最新データまで横断的に検索しつつ、関連度スコアに基づきランキングします<sup>37</sup>。同時に**重複特許ファミリーの排除**や公開日等による絞り込みも行い、ノイズを低減します。
- ・**クレーム対応付け（マッピング）**：検索で得られた有力な先行文献について、その文献中のどの記載が**対象発明のクレーム要素に対応するか**を自動で対応付けます<sup>41</sup>。例えば先行特許の具体的実施例中の構成が、新規性を否定すべきクレーム要件に該当することを示すハイライトや説明を生成します。これにより文献のどの部分が発明の新規性を破るポイントか一目で分かります。
- ・**比較表・レポート作成**：以上の結果を総合し、**先行技術対比表**や**調査報告書**を自動生成します<sup>41</sup>。比較表ではクレーム要素ごとに**対応する先行文献**の記載箇所を列挙し、新規性〇×の判断材料を整理します。またレポートには発見文献の要約や法的評価（特許分類や引用関係など）も含まれ、調査結果をエビデンス付きで体系的にまとめます。

Patsnap Eurekaの強みは、こうした調査プロセス全体を自動化しつつも、結果を専門家が検証・編集しやすい形式で出力する点にあります<sup>41</sup>。生成されたサーチレポートやクレーム対比表は、引用箇所や根拠が明示された上でWord/Excel等に編集可能なフォーマットで提供され、ユーザー（弁理士・審査官）が追記修正してそのまま公式文書に流用できる実用性を備えています<sup>41</sup>。例えば、特許事務所で先行技術調査報告書をクライアント向けに作成する場合でも、AI出力をベースにわずかな加筆で完成させることができます。これにより、従来1~2週間かかっていた調査業務が数分~数時間で完了し、報告されています<sup>41</sup>。実際、Patsnapによれば同社エージェントの導入で「それまで3日かけていた特許調査を3時間で終え、残りの時間を戦略立案に充てる」といったワークフロー改革が起きているとのことです<sup>42</sup> <sup>43</sup>。このように、Eurekaエージェントは定型的で時間のかかる作業をAIに任せ、人間の専門家は結果の判断と戦略策定に注力できる環境を実現します<sup>42</sup> <sup>43</sup>。

Patsnap Eurekaの高性能の背景には、圧倒的な専門データ量と実務知識の組込みがあります<sup>44</sup>。Patsnapは10年以上にわたり特許・技術情報に特化したデータ基盤を構築しており、現在174か国・2億件以上の特許データ、さらに2億件以上の学術論文など関連ドキュメントを収録しています<sup>44</sup>。これら膨大なデータは、重複特許のファミリー統合や書式統一、専門家による属性ラベリングなど多段階のクレンジング・正規化処理を経て、高品質なAI訓練用コーパスとなっています<sup>44</sup>。加えて、世界15,000社以上のイノベーション企業・特許事務所にサービス提供してきた経験から、特許調査や審査・出願業務での実際の判断基準や手順を体系化し、それらをAIエージェントの行動ロジックに反映させています<sup>45</sup>。簡単にいえば、「何をもってクレームの新規性が破られると判断するか」「クレームのポイントをどう読み解くか」といった人間の専門知識をAIに埋め込んでいるのです<sup>45</sup>。その結果、Eurekaエージェントは特許特有の言語表現やクレーム体系を理解し、単なるキーワード一致ではない意味理解に基づく検索や、法律実務に則したアウトプット形式（比較表やクレームチャート）を生成できるようになっています<sup>46</sup> <sup>47</sup>。このドメイン特化型の訓練と、RAG（検索強化型生成）やRAT（Retrieval-Augmented Thinking）といった先端AI技術の組み合わせにより、Eurekaは高精度で幻覚の少ないプロフェッショナル向けAIソリューションを実現しているのです<sup>48</sup> <sup>49</sup>。

以上を踏まえ、Patsnap Eurekaは単なるChatGPT的なチャットボットではなく、知財実務の高度なパートナーと位置付けられます。Eurekaは専門家の指示を待つだけでなく、自ら適切な検索式を立案し、大量のデータから重要情報を取捨選択し、理解しやすい形で提示する「実働型AIエージェント」です<sup>9</sup>。これにより、従来は専門家のみが対応可能だった複雑作業をAIが肩代わりし、最終判断や戦略立案に人間がリソースを振り向けるという、新しい人間-AI協働の形が生まれています<sup>50</sup> <sup>51</sup>。特許クレームの解釈支援という点でも、クレーム要件を自動解析し先行技術との対応付けを行う機能は、従来は人手で行っていた要素比較・無効化チャート作成を飛躍的に効率化します。これは弁理士・審査官の視点から見れば、クレーム読み解にかかる時間を減らし、より多くの時間を発明の本質的評価に割けることを意味します。実務経験者にとって、Eurekaは「下調べと情報整理を任せられる優秀な補佐」であり、知財実務全体の生産性向上につながるツールなのです<sup>50</sup> <sup>52</sup>。

## 特許庁や企業知財部門への導入：課題と可能性

特許庁（審査官）や企業の知財部門がこのような特許調査AIを導入する場合、考慮すべき可能性と課題がいくつか存在します。

### 導入の可能性・メリット

- 調査効率・網羅性の飛躍的向上: PatentBenchの結果が示す通り、専用AIエージェントは従来の人手や従来型検索ツールでは難しい網羅的検索を短時間で実行できます<sup>3</sup>。特許庁では出願件数の増加に対し審査官一人当たりの負荷が課題となっていますが、AIが下調べを代行することで審査の迅速化と見落としリスク低減が期待できます。実際、米国特許商標庁(USPTO)もAIによる先行技術検索結果を審査官に提供する試行を開始しており、AI活用でより関連性の高い文献を上位表示して見逃しを防ぐことを目指しています<sup>53</sup>。USPTOの発表によれば、技術の高度化・文献爆発に対応するためAIで検

索結果を拡張・ランキングすることで「従来は上位に現れなかった関連文献を容易に発見できるようになる」計画です<sup>54</sup>。

- **多言語・グローバル検索対応:** AIは巨大な多言語データを扱えるため、例えば日本語の発明に対して中国語や英語の先行技術を同時に検索・翻訳してくれるなど、国際出願への対応力が上がります。Patsnap Eurekaは英中日の特許文献データを統合的に利用でき<sup>44</sup>、ユーザーの言語に翻訳して結果を提示する機能も備えているため、企業の知財部門が海外特許調査を自前で完結させることも可能になります。特許庁においても、各国の関連文献をAIが収集・翻訳して提示すれば、審査官が他言語文献を調査するハードルを下げられ、国際調査報告の質向上につながるでしょう。
- **専門家リソースの高度化:** 単純作業の自動化により、人間の専門家は判断・戦略立案といった高付加価値業務に集中できます<sup>50</sup>。企業知財では、AIが特許調査や要約作成を担うことで、知財担当者は得られた知見をもとに特許ポートフォリオ戦略や他社特許のリスク分析に時間を割けます。特許庁でも、先行技術のピックアップにかかる時間が減れば、審査官は発明の進歩性評価やクレーム解釈など本質的審査により注力できます。これは審査の質を維持向上しつつ量もさばくために有効です。
- **客觀性・知見の共有:** AIが関与することで、調査の再現性・客觀性が向上する側面もあります。人間の検索はどうしても属人的ですが、AIツールは同じ入力に対し同じように処理を行うため、結果のばらつきが減ります。審査官の間や企業の担当者間で経験差による見落としを補完し、知見を組織内で共有する手段となりえます。特許庁でAI検索結果を一律提供すれば、審査の一定水準の均一化にも寄与するでしょう。

## 導入の課題・注意点

- **信頼性・誤検出のリスク:** いかに専門特化とはいっても、AIによる検索結果にも誤りや過不足は残存します。PatentBenchでもリコール率36%であり、残り約64%の有効文献は拾えていないことに留意が必要です【22†】。したがって、AIが提出した結果をそのまま鵜呑みにせず、最終判断には人間の確認が不可欠です<sup>51</sup>。AIが提出しなかった文献にも重要なものが存在する可能性を踏まえ、審査官や調査担当者は補完的なサーチや結果の精査を続ける必要があります。また、AIが誤って無関係な文献を提示する（誤検出する）ケースも考えられるため、その理由を説明できる仕組み（後述のExplainability）を整えた上で、人間がチェック・除外するプロセスが重要です<sup>51</sup> <sup>55</sup>。
- **説明可能性（Explainability）:** 特許審査や係争では、なぜその文献が選ばれたのかを説明できなければなりません。ブラックボックスなAIでは、結果だけ提示されても審査官や弁理士は判断に使えません<sup>55</sup>。この点、Eurekaは文献の該当箇所ハイライトやクレーム対応表など理由を示す機能を有しています<sup>41</sup>、さらにAIの内部判断ロジックの可視化も求められるでしょう。例えば「このキーワード類似度により候補に上がった」等の説明を出せると、ユーザーが納得・信頼しやすくなります。特許庁が導入するには、審査結果が後に争われることも踏まえ、AIによる検索過程を証跡として残すことや、審査記録に組み込む手順の検討も必要です。
- **人間との協調（Human-in-the-Loop）:** AIに全て任せきりにせず人間との協調で使う枠組みが重要です<sup>51</sup>。例えば特許庁では、AIが提出した10件程度の参考文献リスト（審査官への助言）をまず示し、審査官はそれらを検証・取捨選択した上で追加検索も行う——といったハイブリッドな運用が考えられます。実際、USPTOが2025年に開始したAI検索パイロットでは、AIが出願の内容と分類から自動検索した上位文献10件前後を審査開始前に提示しつつ、出願人にもそれを参考情報（応答義務なし）として知らせる仕組みを試験導入しています<sup>56</sup> <sup>57</sup>。このようにAIをアシスタント役と位置付け、人間が最終的な役割を担う形であれば、AIのバイアスや誤りによる致命的なミスを防ぎつつ効率化を享受できます。

- ・**インフラ・セキュリティ**: 特許庁で扱う出願書類や企業の未公開発明情報は極めて機密性が高く、外部クラウドAIサービスにそのまま投入することは許されません<sup>58 59</sup>。したがって、**オンプレミス環境**でAIを運用できるか、データの守秘性をどう確保するかが課題となります。Patsnap Eurekaはクラウド版だけでなくファイアウォール内へのオンプレ導入にも対応し、ISO 27001認証など企業向けの厳格なセキュリティ対策を謳っています<sup>58 59</sup>。特許庁の場合、自庁内サーバーでAIモデルを動かし、外部と切り離した形で検索することが求められるでしょう。また、モデルが学習のためにユーザーの機密データを蓄積しないようデータ隔離を徹底する仕組みも必要です<sup>58</sup>。技術インフラの整備に加え、社内規程やガイドラインでAI利用時の注意事項（入力する文書の選別等）を定め、情報漏洩リスクを管理することも重要です。
- ・**法規制・ポリシーの整備**: AI活用が進むにつれ、法的・倫理的なルール整備も避けられません。特許審査では審査官が審査したことが重要であり、AI判定に丸ごと依拠することは許されないでしょう。しかしこまでをAIに任せたか明確にする必要が出てくる可能性があります。また、AIが新規性・進歩性の判断まで踏み込むようになると、判断過程の透明性や、AI出力に誤りがあった場合の責任の所在なども議論を要します。現状ではAIは検索補助ツールとの位置づけですが、特許庁内での公式利用にあたってはその役割範囲を規定し、品質評価や定期的なモデル検証（バリデーション）のプロセスを設けることが望ましいでしょう。企業においても、AIが作成したドラフトやレポートをそのまま使う際のチェック体制や、不適切な利用を避けるための指針が必要です。

以上のように課題はありますが、これらは適切なガバナンスと人間の関与によって対処可能であり、メリットと比べて決して越えられない壁ではありません。むしろ、現在各国の特許庁もAI活用に前向きであり、USPTOでは自動検索結果を審査官と出願人に共有する試みを通じてAIの有効性を評価しようとしています<sup>60 57</sup>。欧州特許庁(EPO)も分類検索や機械翻訳でAI技術を先行導入してきており、今後はより生成AI的な文書要約や先行技術マッチングへの関心も高まる見られます。日本特許庁(JPO)でも過去に特許分類付与への機械学習活用を試行するなどの動きがあり、今後Patsnapのような外部ソリューションとも連携しつつ審査効率と質の両立を図っていく可能性があります。

## まとめと今後の展望

PatentBenchの登場とそれが示した成果は、**知財業務におけるAI活用が実用段階に入ってきたことを象徴**しています。Patsnap Eurekaの新規性調査エージェントは、汎用AIでは太刀打ちできなかった特許特有の課題に対し、専門特化したAIならではの高精度な解決策を提示しました<sup>3</sup>。もっとも、現時点でもAIは人間専門家の代替ではなく**強力なアシスタント**という位置づけです。特許調査・分析のプロセス全体を効率化しつつも、最終的な判断やクリエイティブな戦略立案は依然として人間の役割が重要です<sup>51 61</sup>。したがって、知財実務者はAIの利点と限界を正しく理解し、自らの知見と組み合わせて使いこなす必要があります。

今後の展望として、PatentBenchは新規性調査だけでなく**他の特許タスクにも評価範囲を拡張していく予定**です<sup>62</sup>。特許クレームの解釈能力、明細書のドラフティング補助、拒絶理由応答の作成、訴訟における侵害分析など、多岐にわたるシナリオでAI性能がベンチマークされることでしょう<sup>63 64</sup>。これにより、どの領域でAIが実用水準に達し、どこに改善の余地があるかが明確になり、業界全体で**知財AIの技術水準が底上げ**されると期待できます。Patsnapのような先行企業だけでなく、各国特許庁や他のソフトウェアベンダー、オープンソースコミュニティも含めた広範な協力の下で、特許分野に特化したAI技術がさらに発展していくでしょう。

企業の知財部門や特許法律事務所にとって、**AI活用はもはや「検討すべきか？」ではなく「どれを使うべきか？」の段階に入っています**<sup>65 66</sup>。本レポートで述べたように、汎用モデルをそのまま当てはめるのでは不十分であり、特許専門データで鍛えられたツールを選定することが重要です<sup>46</sup>。その際、PatentBenchのような客観指標がツール選択の指針となり、例えば「新規性調査にはPatsnap Eurekaが現状では最も信頼できる」といった判断が可能になります。もっとも、将来的には競合する他の特許AI（例えばGraphベースのIPRallyやオープンソースのPQAIなど）も進化し、性能が拮抗する場面も出てくるでしょう<sup>67 68</sup>。ユー

側は引き続きベンチマーク結果や実地検証を注視し、自組織のニーズに合ったAI活用法をアップデートしていく必要があります。

最後に強調すべきは、**知財プロフェッショナルの役割はAIによってさらに重要になる**という点です。AIは知的作業を代替するのではなく、**増幅器（augmenter）**として働きます<sup>51 61</sup>。大量の情報を収集・整理する力が飛躍的に高まった分、どの情報をどう使うか、得られた示唆をいかに戦略に落とし込むかが、これまで以上に成果を分けるでしょう。特許という制度自体、そしてそれを扱う実務のクリエイティビティや判断力は、人間にしか発揮できない部分が残ります。しかしAIを正しく活用できる専門家は、そうでない者に比べ圧倒的なアドバンテージを持つ時代となります。**知財・特許業務とAIの協働**は今まさに幕開けの段階ですが、PatentBenchに見るよう確かな進歩が始まっています。知財実務者はこの波を捉え、AIエージェントを味方につけることで、これまでにない高みの付加価値サービスを提供できるでしょう。それが「知財の仕事をAIに任せられるのか？」という問い合わせに対する、現場からのひとつの答えとなるはずです<sup>4</sup>。

**参考文献・出典:** PatentBenchプレスリリース<sup>1 3</sup>、共同通信PRワイヤー<sup>69 70</sup>、Patsnap公式ブログ<sup>46 51</sup>、FedScoop<sup>53</sup>、USPTO連邦公報抜粋<sup>71 72</sup>など。

---

<sup>1</sup> <sup>2</sup> <sup>3</sup> 知財の仕事をAIに任せられるのか？ - Patsnapは知財専用AI評価のグローバルスタンダードを発表 | PRワイヤー | 沖縄タイムス+プラス

<https://www.okinawatimes.co.jp/articles/-/1752318>

<sup>4</sup> <sup>5</sup> <sup>6</sup> <sup>7</sup> <sup>8</sup> <sup>9</sup> <sup>10</sup> <sup>11</sup> <sup>39</sup> <sup>40</sup> <sup>41</sup> <sup>44</sup> <sup>45</sup> <sup>69</sup> <sup>70</sup> 知財の仕事をAIに任せられるのか？ - Patsnapは知財専用AI評価のグローバルスタンダードを発表 | Patsnapのプレスリリース | 共同通信PRワイヤー

<https://kyodonewsprwire.jp/release/202512181268>

<sup>12</sup> <sup>13</sup> <sup>14</sup> <sup>15</sup> <sup>16</sup> <sup>63</sup> <sup>64</sup> Introducing PatentBench: Setting the Standard for Patent AI | Patsnap

<https://www.patsnap.com/resources/blog/introducing-patentbench-a-new-standard-for-ai-built-for-patent-tasks/>

<sup>17</sup> <sup>18</sup> <sup>19</sup> <sup>20</sup> <sup>21</sup> <sup>22</sup> <sup>23</sup> <sup>24</sup> <sup>25</sup> <sup>26</sup> <sup>27</sup> <sup>28</sup> <sup>29</sup> <sup>30</sup> <sup>31</sup> <sup>32</sup> <sup>33</sup> <sup>34</sup> <sup>35</sup> <sup>36</sup> <sup>37</sup> <sup>38</sup> <sup>42</sup> <sup>43</sup> <sup>50</sup> <sup>62</sup> PatentBench - AI-driven Novelty Search Benchmark | Patsnap

<https://www.patsnap.com/benchmark>

<sup>46</sup> <sup>47</sup> <sup>51</sup> <sup>52</sup> <sup>55</sup> <sup>61</sup> <sup>65</sup> <sup>66</sup> <sup>67</sup> <sup>68</sup> Which AI Patent Tools Actually Work Well in 2025?

<https://www.patsnap.com/resources/blog/articles/ai-patent-tools-that-work-2025/>

<sup>48</sup> <sup>49</sup> <sup>58</sup> <sup>59</sup> AI Agents for IP & R&D Innovation Platform | Patsnap Eureka

<https://www.patsnap.com/products/eureka/>

<sup>53</sup> <sup>54</sup> US Patent Office eyes using AI to improve ‘prior art’ searches | FedScoop

<https://fedscoop.com/patent-office-eyes-ai-prior-art-searches/>

<sup>56</sup> <sup>57</sup> <sup>60</sup> <sup>71</sup> <sup>72</sup> USPTO Will Launch Pilot to Evaluate Results of New AI Search Tool for Patent

Applications - IPWatchdog.com | Patents & Intellectual Property Law USPTO Will Launch Pilot to Evaluate Results of New AI Search Tool for Patent Applications

<https://ipwatchdog.com/2025/10/07/uspto-launch-pilot-evaluate-results-new-ai-search-tool-patent/>