特許実務AIの現在地：ベンチマーク「PatRe」が示す能力と限界

2026年の研究「PatRe」は、特許審査官による拒絶理由通知（OA）と出願人による意見書（Rebuttal）の生成を、AIがどの程度正確に行えるかを検証した初のベンチマークです。検証の結果、AIは「形式的な文書作成」には長けているものの、「実質的な法的判断」には重大な欠陥があることが浮き彫りになりました。

AIができること・できないこと（得意 vs 苦手）

「反論」は得意だが「問題発見」は苦手

意見書生成では高い網羅性を示す一方、拒絶理由の能動的な発見・構築ではスコアが大幅に低下します。

外部証拠への強い依存（引用精度 5.1% → 74.3%）

内部知識のみでは引用文献の「幻覚」が多発しますが、適切な資料を与えると精度が劇的に向上します。

「形式」は流暢だが「論理」は脆弱

文章の明快さやスタイルは高評価ですが、法的根拠（Soundness）や論理の完結性は極めて低いのが現状です。

実務導入における3つの重大リスク

「過剰批判バイアス」による拒絶の捏造

本来特許査定されるべき案件に対し、AI審査官は「厳しすぎる」不適切な拒絶を行う傾向があります。

条文適用の論理不整合（特に§112）

明細書全文を参照できない設計上の制約もあり、記載要件等の判断において60%もの過剰執行が見られます。

「人間による監督」が不可欠

AI判定（LLM-as-a-judge）は人間評価よりスコアが高く出る傾向があり、専門家による実質的な点検が必須です。

AIの自己評価と人間評価の乖離