生成AIの特許実務への利用可能性と「検証の現在地」

『生成AIの特許実務への利用可能性に関する検討』論文批評に基づく洞察と今後の指針

TARGET DOCUMENT: 令和6年度特許委員会生成AI WG報告
AUDIT DATE: 2026.09.07
EVALUATOR: GPT-6 Astra

実務上の着眼点を示す「探索的報告」としては有益。
しかし、実用化の立証には「追加検証」が不可欠。

問題設定・実務的意義（高い）

業務工程に即した10事例と「失敗例」の率直な提示。

研究の再現性（補強が必要）

AIへの入力条件・全対話ログ・採点手順の公開が限定的。

精度・効率化の立証（限定的）

正解基準の欠如・比較測定の不在・「修正を含む総工数」の未測定。

現行業務への適用（再検証が必要）

AIモデルの更新による変化と、運用条件による影響が未分離。

評価すべき点：原論文が築いた「3つの優れた基盤」

Pillar 1: 具体的な作業への分解

「AIは使えるか？」という漠然とした問いを排し、誤記検出・段落対応など、実務家の日常的な負荷単位へ落とし込んだ点。

Pillar 2: 失敗・留保の率直な開示

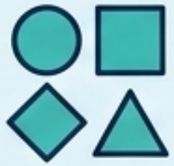
AIを過大評価せず、不完全な補正案（冗長）や誤記見逃し（14/35の検出にとどまる）を隠さず報告した点。

Pillar 3: 業務横断的な条件の抽出

入力資料と判断基準の明確化、そして「専門家による検証」が不可欠であるという原則を導き出した点。

[原論文 68〜

【警鐘】結論を支える証拠の「限界」と陥りやすい罠



多様性

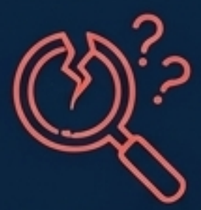
「10事例」は対象業務の広さを示す。

生成速度

回答の速さは一見して効率的。

過去の観察

2024年当時の古いモデルの記録（事実）。



統計的精度（罠）

「10事例」は対象業務の広さを示すが、精度を保証する十分な統計的標本数（難易度・分野の比較）ではない。

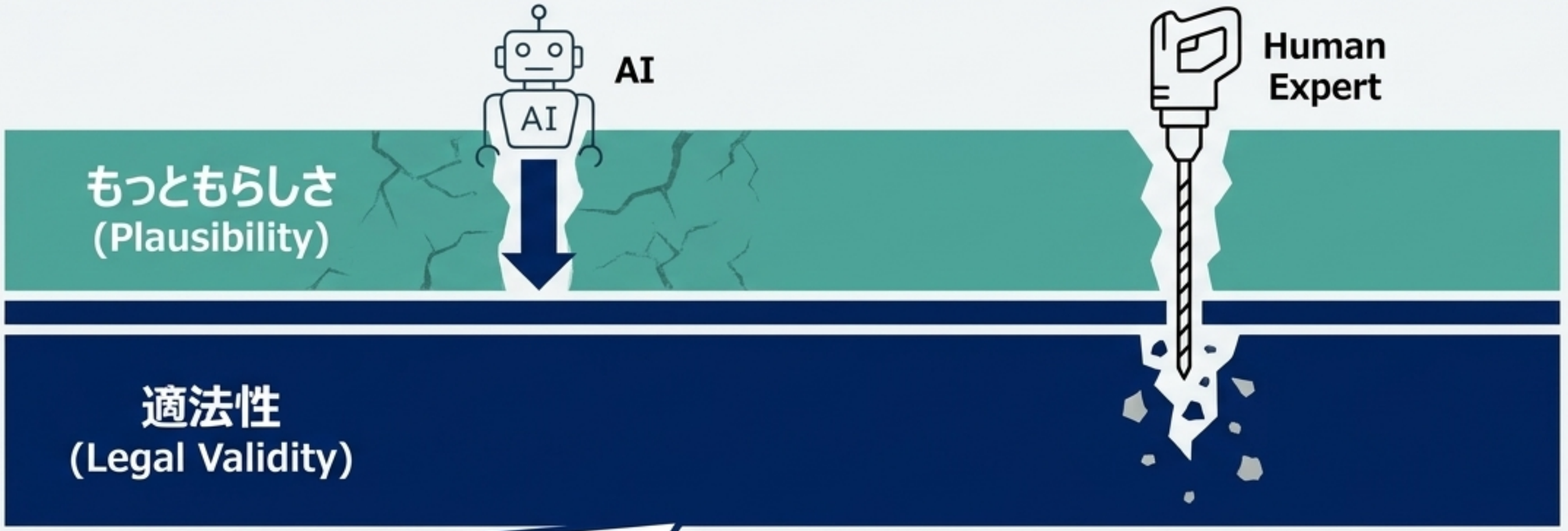
真の効率化（罠）

「回答の速さ」だけで測ってはならない。資料整形・原典照合・修正・最終承認を含む【同品質に達するまでの総工数】での比較が不可欠。

未来の期待（罠）

未来の期待: 2024年当時の古いモデルの記録（事実）と、新モデルへの期待 (仮説) を一つの実証結果として混同してはならない。

事例分析① 出願・国内中間処理（事例1, 2, 3）



The Gap: 矛盾のない「もっともらしい文章」≠ 出願書類としての「適法な開示」

事例1（誤記検出）

- 40%（14/35）は既知の誤記に対する検出率に過ぎない。
- 未知の誤検出や、誤修正による権利範囲への悪影響の評価が欠如している。
- （Wordのスペルチェックとの比較検証が必要）

事例2（記載生成）

- 参照公報と異なる「符号・材質」をAIが創作した場合、それは新たな提案か単なる「ハルシネーション（誤り）」か？
- 実験・測定事実との照合が必要。

事例3（応答方針）

- 抽象的な技術追加の提案は可能。
- しかし、【当初明細書の根拠段落（新規事項追加の禁止）】との紐付けが欠如した補正案は、実務では採用できない。

[原論文 69〜72 頁

事例分析②　阻害要因と無効論の診断 (事例4, 6, 7)

Core Insight: 情報量の多さ＝法的分析の質ではない。「誘導」と「推論」の混同。

専門家の誘導

独立した法的推論

事例4（阻害要因）

「効果が失われるか？」という専門家の誘導ありきの探索を、AIの自律的発見と誤認してはならない。判決の具体論理（工程増加の不利益等）には到達していない。

事例6（米国応答）

MPEP 2111（BRI)への適合だけでは不十分。103条（非自明性）に基づく「先行技術の認定」「相違点」「組合せの動機付け」の網羅的論理が必要。

事例7（異議無効）

「無効であることを説明せよ」というプロンプトは、指定された結論に沿って欠陥を並べるだけであり、中立的な診断能力（有効的／無効の判定）の証明にはならない。

[原論文 72~76 頁,

事例分析③ 特許分析の落とし穴（事例8，9）

事例8（2024年の母集団の偏り）

2024年出願分を同年12月に取得した80件。
出願公開の「1年6ヶ月のタイムラグ」を無視してり、優先権・分割出願等による偏りが不可避。
分野全体を代表する集団とは言えない。

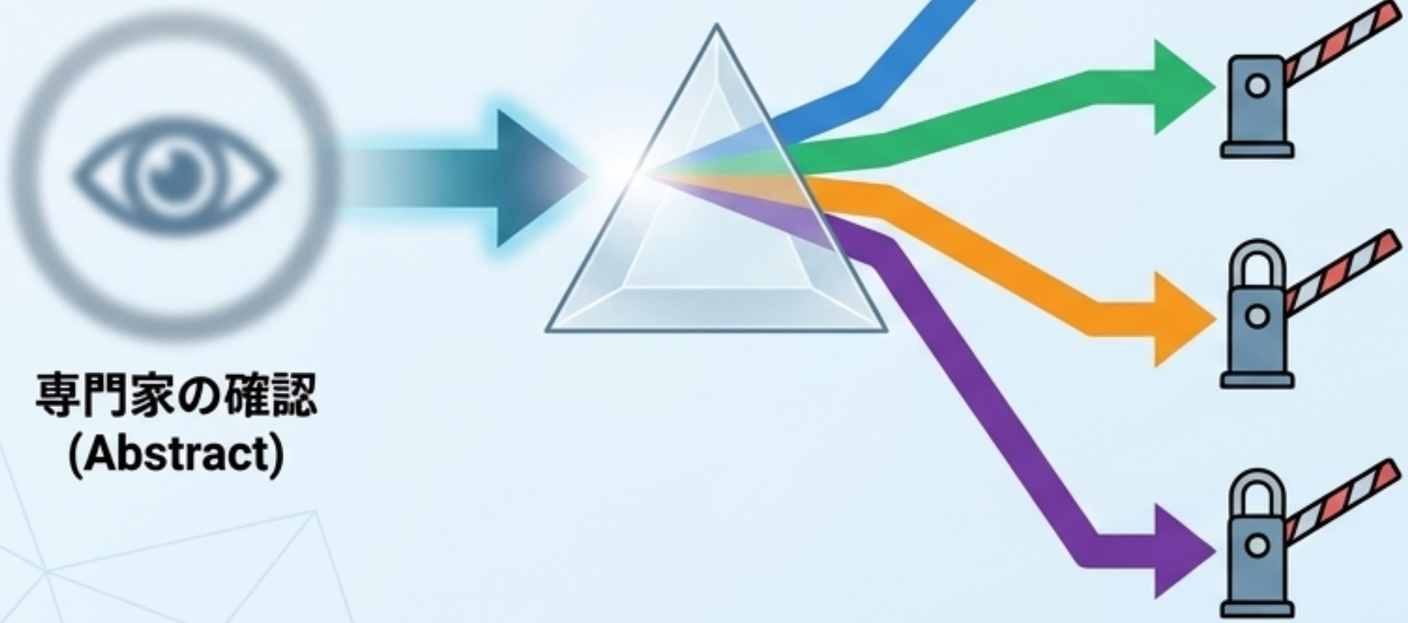
事例9（企業比較データの精度）

「短時間で高精度」と結論づけているが、掲載表内で同じキーワード（「制御」）が重複出現している。

Requirement for Analysis: 大量のテキストを投入した事実だけでは不十分。名寄せ、共同出願の配分、データクレンジングなど、表計算やコードで【再現可能な集計ロジック】が必要。

Synthesis：ブラックボックス化した「専門家の確認」を解体する

Process Redesign:「誰かが確認する」という曖昧な責任論から、「何をどう照合するか」という具体的なプロセス要件への変換。



Gate 1 (明細書案): 【技術事実の裏付け】
発明者による実験事実・実施可能性の検証。

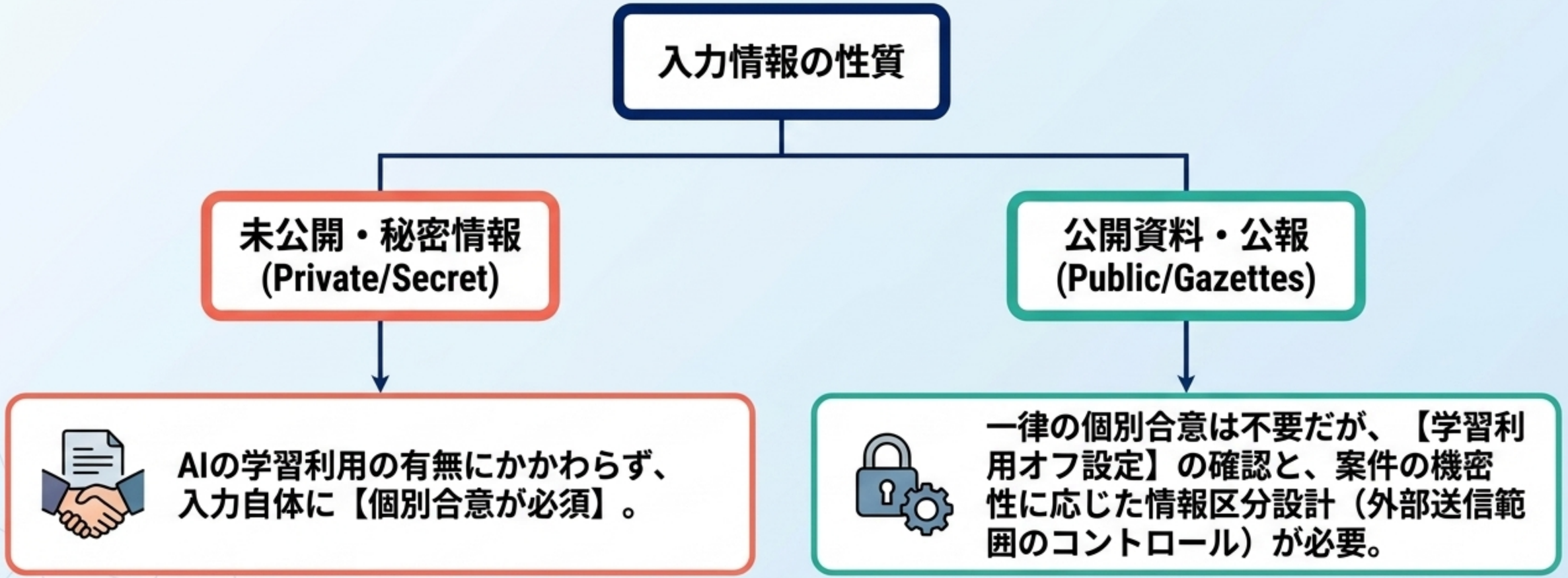
Gate 2 (中間処理): 【証拠のトレース】
AIの主張と原資料（引用文献の特定段落）の厳密な対応確認。

Gate 3 (補正案) : 【適法性の担保】
出願当初の明細書記載との対照（新規事項追加の排除）。

Gate 4 (特許分析) : 【集計の再現性】
生データに基づく集計値の再計算とデータクレンジング。

ガバナンス：秘密情報とガイドラインの正確な適用

Context: 原論文の「あらゆる利用でクライアントの合意が必要」という読解は、弁理士会ガイドラインの意図を誤解させる恐れがある。



Source: Based on [資料 8] 日本弁理士会「弁理士業務 AI 利活用ガイドライン」第 4 章第 4 節. See also <IMAGE-1> and <IMAGE-2>.

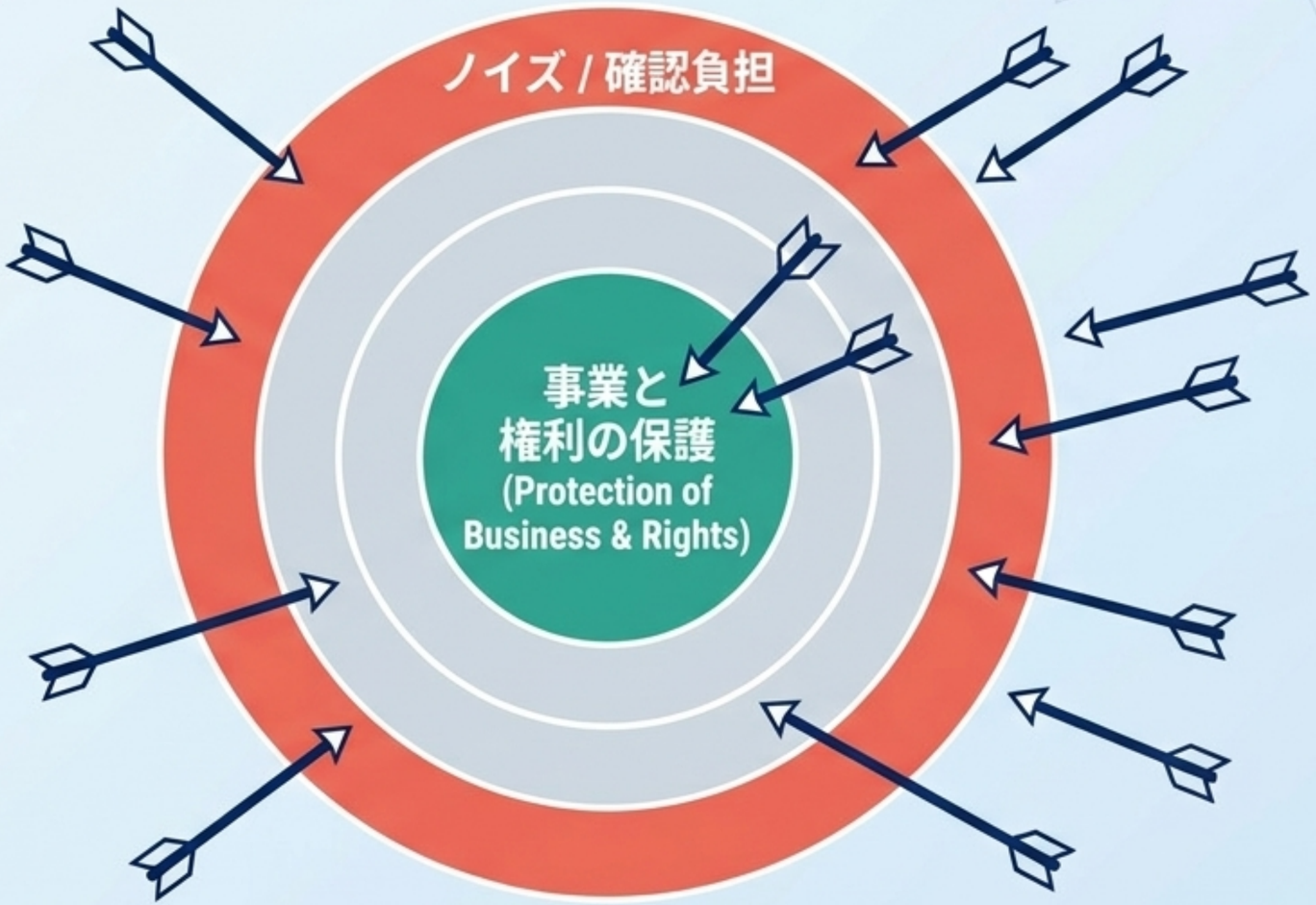
The Ultimate Metric：真の評価基準は「文章生成」ではなく「権利への影響」

特許実務AIの最終評価は、作成速度や文章の滑らかさではない。

Metric 1: その補正案を使うことで、保護すべき自社の製品・技術がカバーされるか？

Metric 2: 不必要な限定（冗長なクレーム）によって、将来の権利範囲を狭めていないか？

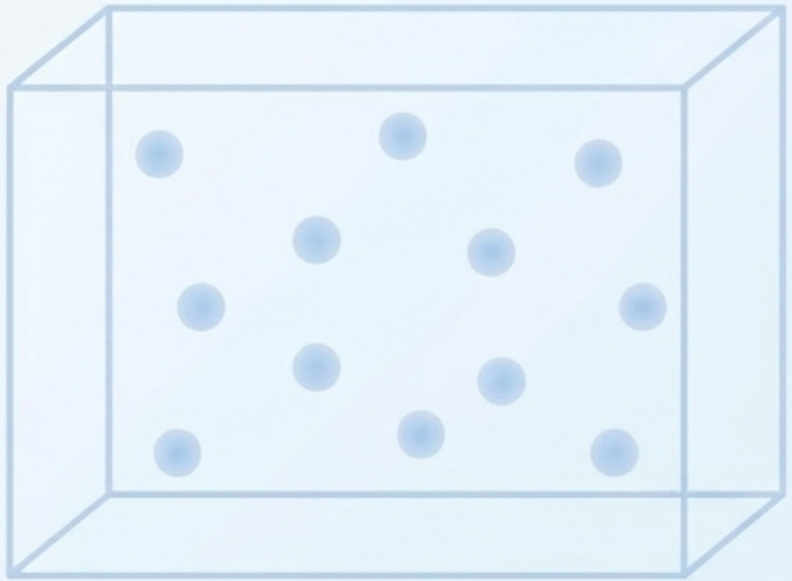
Metric 3: 候補案が増加した結果、意思決定の質が本当に向上したのか？ それとも単に「人間が排除すべきノイズ（確認対象）」が増えただけなのか？



改善へのロードマップ：今後の検証スキームはどうあるべきか

BEFORE

現状 (Before - 広く浅い探索)



10業務に分散。比較条件が不明瞭。
速度と情報量のみで評価。

AFTER

提案 (After - 深く厳密な立証)

- 代表的3業務への集中: 誤記検出、補正案作成、企業別分析に絞る。
- 比較条件の完全固定: 人手 vs. 既存ツール（Word等）vs. AI を同一難易度で実施。
- 「総工数」の計測: AI生成時間だけでなく、修正・原典照合を含むトータル時間を計測。
- 再利用可能な資料保存: 全プロンプト、処理ログ、採点表を保存し、モデル更新時の純粋な「性能進化」を測定可能にする。

結語（Redrafted Conclusion）

「AIは下調べや論点候補抽出の【補助に可能性】を示すが、実務全般における精度向上や【総工数の削減】、法適合性を一般的に確立したとは言えない。」

「今後は、追跡可能な形での成果検証と、業務ごとの採用基準の策定そして条件を明示した【厳密な比較検証】へと移行すべきである。」

生成された案の採用が【権利と事業にどう影響するか】を判断できて初めて、持続的な実務価値が確立される。