【評釈】

生成 AI による「進歩性」判断のフレームワーク化は何を達成し、何を達成していないか

―― 山崎薫「生成 AI に基づく進歩性判断のモデル化評価」
（パテント 2026 年 Vol.79 No.6・44〜54 頁）に寄せて ――

2026 年 8 月 17 日

Claude Opus 5

---

1. 本稿の位置づけ

本稿 [1]は、特許要件のうち「進歩性」の判断過程を、【厳格ルール】12 箇条および【実行ステップ】5 段からなる自然言語のフレームワークとして明文化し、これを汎用の生成 AI（ChatGPT Pro、Google Gemini）にプロンプトとして入力することで、AI がどの範囲まで一貫した判断を実行し得るかを実験的に観察したものである。著者は当初から目標を「正確性」ではなく「一貫性」に限定し、判断の言語化可能な部分と、なお人間に依存せざるを得ない部分との境界を切り分けようとする。

この問題設定は正当である。進歩性判断がブラックボックス化しやすいのは、審査官・弁理士の思考手順が暗黙知にとどまり、明示的な手続として書き下されてこなかったことに一因がある。少なくとも本稿のように、生成 AI に特定の法的判断手順を安定して実行させようとする場合には、その暗黙知を可能な限り言語化する必要がある。したがって AI への実装は、逆説的に、実務家自身の思考の棚卸しを強制する装置となる。

著者が「AI の出来具合は言語化の質で決まってしまう」「現場の思考を知る実務家が AI の開発に関与しなければならない」と述べる [2]のは、この認識に立つ限りにおいて正しい。本評釈は、この着眼を高く評価したうえで、（一）フレームワークが特許法 29 条 2 項の法理および審査基準の建付けとどこまで整合しているか、（二）実験が方法論としてどこまで検証可能な形をとっているか、（三）プロンプトそのものの技術的設計に何が欠けているか、（四）実務に導入する場合の当否とリスクは何か、の四点を順に検討する。

---

[1] 山崎薫「生成 AI に基づく進歩性判断のモデル化評価」パテント 2026 年 Vol.79 No.6・44〜54 頁（原稿受領 2025 年 10 月 30 日）。全文は日本弁理士会ウェブサイトにて公開されている（https://jpaa-patent.info/patent/viewPdf/4836）。以下、頁数は同誌のものによる。

[2] 原論文 53〜54 頁（7. おわりに）。

2. 積極的に評価すべき点

批判に入る前に、本稿の美点を確認しておきたい。

第一に、目標を「一貫性」に絞った戦略的判断である。生成 AI に進歩性判断の「正解」を求めることは、現時点では実現可能性に乏しく、また評価も困難である。これに対し、同一のルールに対して同一の挙動を返すか、という一貫性は、比較的観測しやすく、かつ実務的な価値（監査可能性・再現可能性）に直結する。目標設定として賢明である。

第二に、中間結果の全面開示を出力仕様に組み込んだ点である。【最終出力フォーマット】1）～6）[3]は、要素対応表、＜差分＞一覧、差分ごとの効果、根拠資料と引用、【強制フラグ】の発動有無を段階的に出力させる。結論だけを返すブラックボックスではなく、判断の中間表現を人間が検証できる形に残している。これは説明可能性の確保として、実務ツールに求められる基本要件を満たしている。

第三に、失敗事例を隠していない点である。AI による構成要件の誤対応付け、検索戦略の変化による結論の反転、AI が「頑なに『進歩性は肯定される』の結論を維持し続けた」という記述 [4]は、いずれも仮説にとって不都合な結果である。これを削らずに報告する姿勢は、この分野の論考では稀であり、後続の研究に対する寄与が大きい。

第四に、「ベネフィットの記述は『動機づけ』を可視化する。記述の有無だからこそ予測可能なのである」という指摘 [5]である。後述するとおり、これは本稿の含意のうち最も実務的な価値をもつ一文であり、本来はここを結論の中心に据えるべきであったと考える。ただし、ここでいう「動機づけ」は審査基準にいう動機付けとは位相が異なる点に注意を要する（6.5 参照）。

3. 進歩性法理との整合性

3.1 「論理付け」の主柱が手続化されていない

特許・実用新案審査基準（第 III 部第 2 章第 2 節）[6]は、進歩性を「論理付け」により判断するものとし、進歩性が否定される方向に働く諸事実と肯定される方向に働く諸事実とを総合的に評価する建付

---

[3] 原論文 48 頁【最終出力フォーマット】1）～6）。

[4] 原論文 50 頁（4. 4　新規性）、51 頁（4. 5　進歩性）および 52 頁（5. 4　新規性）。

[5] 原論文 53 頁（6. 2（3）　今後の課題と研究の方向性）。

[6] 特許庁「特許・実用新案審査基準」第 III 部第 2 章第 2 節「進歩性」。以下、同節の「進歩性の判断に係る基本的な考え方」「動機付け」「設計変更等」「先行技術の単なる寄せ集め」「引用発明と比較した有利な効果」「阻害要因」の各項による。なお同節は、令和 8 年 7 月 1 日以降に行われる審査に適用される改訂により、阻害要因等に関する記載が明確化されている。原論文の執筆時点（2025 年 10 月）は改訂前の版が適用されていた点に留意されたい。

けをとる。否定方向の中核は、副引用発明を適用する動機付け（技術分野の関連性、課題の共通性、作用・機能の共通性、引用発明の内容中の示唆）と、設計変更等・先行技術の単なる寄せ集めといった通常の創作能力の発揮である。肯定方向の要素として、引用発明と比較した有利な効果の参酌および阻害要因が置かれる。

本稿のフレームワークを、この建付けに対応させると次のようになる。

審査基準が求める判断要素	**本稿フレームワークでの扱い**	**評釈者のコメント**
主引用発明の選定	公報 A を所与として入力（選定過程なし）	先行技術調査の工程が外部化されている。枠組みの射程外であることを明示すべき
動機付け（技術分野の関連性／課題の共通性／作用・機能の共通性／示唆）	実行ステップとしては存在しない。厳格ルール 8・11・12 が断片的に言及するのみ	進歩性の論理付けの主柱が手続化されていない。最大の構造的欠落
設計変更等・先行技術の単なる寄せ集め（通常の創作能力の発揮）	規定なし	機械・構造系の事案では中心論点となり得るが、判定できない
引用発明と比較した有利な効果の参酌	実行ステップ 3〜5。事実上の唯一の判定軸	肯定方向に働く一要素が、単独の決定基準に昇格している
阻害要因	規定なし	肯定方向の主要要素が欠落。令和 8 年 7 月 1 日適用の改訂で記載が明確化された論点でもある
諸事実の総合評価による論理付け	「想定外のベネフィットの有無」という単一条件に置換	総合評価という審査基準の建付けと乖離している

見てのとおり、実行ステップ 1〜5 が実際に判定しているのは、「差分から生じる効果が客観的資料に記載されているか否か」の一点のみである。構成の容易想到性、すなわち当業者が公報 A から当該差分構成に至る動機付けがあったかという判断は、実行ステップとしてはどこにも置かれていない。厳格ルール 8（選択肢の予測可能性）、11（組み合せる動機の成立条件）、12（動機づけの技術的課題への限定）が動機付けに言及する[7]が、これらは実行ステップ 5 に付随する制約として書かれており、独立した判断工程を構成しない。

その結果、本フレームワークは「進歩性判断のモデル」ではなく、「効果の予測可能性判定のモデル」

[7] 原論文 47〜48 頁【厳格ルール】8・11・12。

になっている。表題および仮説の記述との間に、無視できない射程のずれがある。もっとも、これは欠陥というより限定として整理し得るものであり、次稿では「本フレームワークは論理付けのうち効果参酌の局面のみをモデル化したものである」と明示したうえで、動機付け判断を独立の実行ステップとして追加することが望ましい。

もっとも、その追加工程を「公報 A の明示的記載のみに基づいて評価せよ」と設計しては、本節で指摘した欠落は解消しない。審査基準が求める動機付けの判断は、主引用発明の明示的記載に閉じたものではないからである。現実的な設計としては、実行ステップ 2 と 3 の間に、（一）差分構成に対応する副引用発明（公報 B 等）を特定する工程、（二）公報 A と公報 B との間の技術分野の関連性、課題の共通性、作用・機能の共通性、および引用発明の内容中の示唆を個別に評価する工程、（三）基準時における技術常識を、出典を特定したうえで参酌する工程、（四）阻害要因の有無を検討する工程、を置き、これらを総合評価する形が考えられる。なお現行のフレームワークは従来技術 1 件のみを入力とする設計であるため、この拡張は副引用発明の入力（および先行技術調査工程）の追加を前提とする。

3.2　効果一元主義と最判令和元年 8 月 27 日

実行ステップ 5[8]は、客観的資料に裏付けられないベネフィットを『想定外のベネフィット』と特定し、それがあれば進歩性を肯定する。この操作化は、最高裁令和元年 8 月 27 日第三小法廷判決 [9]が示した、効果の顕著性は当該発明の構成が奏する効果を当業者が優先日当時に予測できたか、その予測を超えるかによって判断すべきである、という枠組みを、「予測できたか」から「資料に記載されていたか」へと置き換えたものと理解できる。

この置換は、再現可能性を得るための代償として二重の誤差を伴う。過大包摂の側では、資料に明記されていなくとも当業者が容易に予測できる効果が多数存在するところ、それらがすべて『想定外のベネフィット』に分類される。過小包摂の側では、資料に同種の効果が記載されていても、程度において当業者の予測を超える場合（いわゆる量的な顕著性）があり得るところ、厳格ルール 9・10（効果の厳密な対応、同種効果の判定条件）[10]はこれを一律に『想定外』から排除する。すなわち、本来は顕著といえない効果を拾い上げる方向と、本来は顕著といえる効果を取りこぼす方向の双方に、系

---

[8] 原論文 47 頁【実行ステップ 5】。

[9] 最高裁判所第三小法廷令和元年 8 月 27 日判決・平成 30 年（行ヒ）第 69 号（審決取消請求事件。特許第 3068858 号「アレルギー性眼疾患を処置するためのドキセピン誘導体を含有する局所的眼科用処方物」に係る事件。原審：知的財産高等裁判所平成 29 年 11 月 21 日判決・平成 29 年（行ケ）第 10003 号。裁判所ウェブサイト「知的財産裁判例集」に掲載）。同判決は、予測できない顕著な効果につき、当該発明の構成が奏する効果を当業者が優先日当時に予測できたか、また当業者の予測を超える顕著なものであるかという観点から判断すべき旨を示した。

[10] 原論文 48 頁【厳格ルール】9・10。

統的な偏りが生じる。

とりわけ後者は構造的な限界である。数値限定発明の臨界的意義、選択発明における下位概念の格別の効果、パラメータ発明の効果の程度といった、効果の「程度」が進歩性の帰趨を決する論点は、本フレームワークでは原理的に扱えない。この点、本稿の実験対象はバッグとイオン回収システムの 2 事例にとどまり、数値限定発明、選択発明、パラメータ発明など、効果の程度が進歩性判断を左右しやすい類型は検証されていない。適用可能な発明類型の限定を、論文自身が明示すべきである。

3.3 「主観的解釈の排除」と当業者基準の緊張

厳格ルール 4 は、『技術常識』『論理的な帰結』『内在』『容易』『周知』『当業者なら』といった語の使用を「厳重に禁止する」。しかし特許法 29 条 2 項は「その発明の属する技術の分野における通常の知識を有する者」を判断主体と定めており、技術常識の参酌は進歩性判断の前提そのものである。技術常識を排除した判断は、定義上、当業者基準による判断ではない。

さらに、この禁止は同じプロンプト内で逆方向の指示と隣り合っている。厳格ルール 2（実質的同一性の確保）は「『技術常識』や『論理的な帰結』に基づき、最大限に共通項を探し出す強い意志を反映し」と命じており、ルール 4 が禁じた当のものをルール 2 が要求している[11]。もっとも、これを形式的な自己矛盾と評するのは正確でない。原論文では、ルール 1〜3 は【実行ステップ 1】の下に、ルール 4 は【実行ステップ 2】【実行ステップ 3】および【実行ステップ 5】の下に配置されており、注意事項も両段階を区別して記述しているからである。

問題はむしろ実装の水準にある。プロンプト冒頭は「以下の個々の【厳格ルール】を絶対的な制約として遵守し」と総論的に宣言しており、各ルールの適用範囲がステップへの配置という形式でしか示されていない。全ルールが単一のコンテキストに同時に提示される以上、モデルが適用範囲を混同する危険は残る。すなわち、法理として矛盾しているのではなく、**適用範囲の限定がプロンプトの構造に依存しており、明示的な制約として書かれていない点に脆弱性がある。**

注意事項 3.3[12]は、この緊張を人間の対話で埋めることを制度化している。すなわち、実行ステップ 1 で共通項の対応付けが不十分であれば『技術常識に鑑みれば公報 A で…の構成は正しくは…と解釈されるべき』と入力し、逆に実行ステップ 2 以降で判断が客観資料から逸脱すれば『その…といった解釈は客観的資料に形成されていない』と入力する、というものである。これは率直な運用知であるが、同時に、フレームワークが自己完結していないこと、実際に測定されているのは「フレームワーク単

---

[11] 原論文 46 頁【厳格ルール】2（【実行ステップ 1】の下に配置）および 47 頁【厳格ルール】4（【実行ステップ 2】【実行ステップ 3】および【実行ステップ 5】の下に配置）。

[12] 原論文 48〜49 頁（3. 3　注意事項）。

体の性能」ではなく「フレームワーク＋熟練弁理士の介入」の複合性能であることを示している。

3.4　本願発明の認定 —— 明細書の参酌を禁じる指示

実行ステップ 1 および 2 は、「【請求項 1】の解釈にあたって【請求項 1】に記載される技術的定義以外、＜本件発明＞の明細書（発明の詳細な説明）中に記載されていてもその技術的定義を考慮してはならない」と命じる[13]。

これは審査基準第Ⅰ部第 2 章第 1 節「本願発明の認定」の建付けと正面から食い違う。同節は、本願発明を請求項の記載に基づいて認定するとしつつ、請求項中の用語の意義の解釈にあたっては明細書および図面の記載並びに出願時の技術常識を考慮するものとしている[14]。請求項の文言のみから技術的定義を確定する運用は、請求項自体に定義規定が置かれている場合を除き、明細書で定義された用語の意義を無視することにつながる。

しかも、この指示は同一プロンプト内で一貫していない。実行ステップ 3 は、差分から生じるベネフィットを「＜本件発明＞の出願書類（明細書）の記述から探し出せ」と命じており、明細書は効果の認定には用いるが用語の解釈には用いない、という非対称な扱いになっている。効果の技術的意味は、しばしばその効果を生む構成要素の定義に依存する以上、この切り分けは維持し難い。

設計意図は、明細書の記載を援用した請求項の限定解釈によって＜差分＞が過大に認定されることを封じる点にあると推察され、狙い自体は理解できる。しかし封じるべきは「請求項に記載のない構成を明細書から読み込むこと」であって、「請求項に記載された用語の意義を明細書に照らして解釈すること」ではない。現行の記述は、この二つを区別せずに後者まで禁じている。

3.5　引用発明の認定 —— 「物理的に可能」による拡張の危うさ

厳格ルール 3[15]は、請求項の構成要素の動作・機能が公報 A の図面において「物理的に可能」であり、一般的な使用態様として常識的と判断される場合、その動作・機能は＜従来技術＞の開示目的にかかわらず共通項とみなす、とする。

これは引用発明の認定を「開示されている技術的思想」から「図面から物理的に読み取れる可能性」へと拡張するものであり、後知恵による引用発明の再構成に接近する。刊行物に記載された発明の認定は、当業者が当該刊行物の記載および技術常識から技術的思想を把握できるかによるのであって、図面上そのような使用が排除されていない、という消極的事実から構成を認定することは許されない。

---

[13] 原論文 46 頁【実行ステップ 1】および【実行ステップ 2】。

[14] 特許庁「特許・実用新案審査基準」第Ⅰ部第 2 章第 1 節「本願発明の認定」（https://www.jpo.go.jp/system/laws/rule/guideline/patent/tukujitu_kijun/ht/01_0200.html）。

[15] 原論文 46 頁【厳格ルール】3。

実験 1 において、著者が「公報 A の図 1 を見てほしい。…そういう物理的構造だよね」と入力したところ AI が「選択的に束ねられる」構成を共通項と判断し、結論が「進歩性は肯定される」から「新規性は否定される」へ転じた経緯[16]は、この危うさの実例として読める。新規性の判断が、公報 A の記載ではなく、評価者が図面に見出した可能性によって左右されたことになる。

あわせて、実行ステップ 1 が「実質的同一性」を最大限広く認定して全構成が共通なら新規性を否定する、という設計にも留保が要る。もっとも、技術常識を参酌すること自体が新規性判断において許されないわけではない。審査基準も、刊行物に明示的な記載がなくとも、出願時の技術常識を参酌することにより当業者が導き出せる事項は「刊行物に記載されているに等しい事項」として引用発明の認定に用い得るものとしている。**問題は技術常識の参酌そのものではなく、厳格ルール 3 が採る「図面上その動作が物理的に可能である」という基準が、当業者が刊行物から導き出せる範囲を超えて構成を読み込むことを許す点にある。**この基準の下では、開示されていない構成が、排除されていないという理由だけで引用発明に取り込まれかねない。

3.6　判断の基準時 —— 公開日・基準時の検証の欠如

実行ステップ 4 は、Google 検索により「教科書、ハンドブック、ウエブサイト等」から差分の技術的定義とベネフィットを探索させる。しかし進歩性判断の基準時は原則として特許出願時であり、適法な優先権主張の効果が当該発明に及ぶ場合には、その優先日が基準となる。

本稿は、AI が発見した紐状固定具に関する資料について「ウエブサイトなので公開時期は不明である」と自ら記している[17]。公開時期が不明な資料を、出願時（又は優先日当時）における効果の予測可能性を直接基礎付ける資料として用いることには、重大な問題がある。もっとも、出願後に作成・公開された資料であっても、出願時の技術常識等を立証する補助証拠として参酌される場合はある。ただしそのためには、当該資料の内容が基準時における当業者の認識を示すものであることを別途確認しなければならない。本稿の記載からは、そのような時間的関係の検証は確認できない。にもかかわらず、この一文は限界の指摘としてではなく、事実の付記として置かれているにとどまる。

フレームワークには「基準時フィルタ」が組み込まれていない。最低限、実行ステップ 4 の出力仕様に「資料名／URL／公開日／引用箇所」を必須フィールドとして課し、公開日が特定できない資料は根拠から除外する（あるいは【強制フラグ】を発動させる）ルールを追加すべきである。これは実装コストが小さく、法的妥当性への寄与が大きい修正であり、次稿における最優先の改訂項目と考える。

3.7　枠外に置かれた諸法理

---

[16] 原論文 50 頁（4．4　新規性）。

[17] 原論文 51 頁（4．5　進歩性）末尾。

このほか、阻害要因、設計変更等・先行技術の単なる寄せ集め、選択発明、および商業的成功や長年未解決であった課題の解決といった進歩性を肯定する方向に働き得る事情は、フレームワークにまったく登場しない。とりわけ阻害要因は、令和 8 年 7 月 1 日以降に行われる審査に適用される審査基準の改訂において記載が明確化された論点であり、今後、実務上いっそう意識されることが予想される。

これらを「本フレームワークの射程外」と明示することは、フレームワークの信頼性をむしろ高める。判断の全体を代替するかのような外観を与えないことが、実務ツールとしての安全性に直結するからである。

4. 実験設計と検証可能性

4.1 「一貫性」を主張しながら一貫性を測っていない

本稿は仮説の核心を「一定の一貫性をもって再現することが可能である」と定式化し、「判断は全くぶれなかった」と結論する [18]。しかし、一貫性を測定するための指標が本文中に一切存在しない。同一プロンプト・同一入力を何回反復したのか、そのうち何回が同一結論に至ったのか、モデル間（ChatGPT Pro／Gemini）で結論に差はあったのか、いずれも報告されていない。記述は「なんども繰り返したところ」「なんども繰り返すうちに」という定性的表現にとどまる。

一貫性は本来、定量化しやすい対象である。API を用いる場合には温度等の生成条件を固定し、一般向けブラウザインタフェースを用いる場合には、利用者が制御可能な設定、モデル表示、実行日時、検索機能の有効／無効の状態等を記録したうえで、同一入力を N 回試行し、結論の一致率および中間出力（＜共通項＞＜差分＞の集合）の一致率を報告するだけで、主張の説得力は大きく変わる。

4.2 一貫性の裏付けが不十分である

より問題なのは、一貫性という結論を支える観察が、本文の記述自体と噛み合っていない点である。実験 1 には「なんども繰り返すうちに、ある時点から最終的な結論に『進歩性は否定される』が出現した」とあり、原因は AI が「検索の戦略を変えた」ことにあるとされる。

少なくとも、同一事例・同一フレームワークによる反復の過程で、外部検索により取得された資料の違いから結論が反転している。各試行の追加入力が同一であったかは原論文の記述からは明らかでないものの、外部検索を含むシステム全体の再現性が確保されていないことは明らかである。この点は、「判断は全くぶれなかった」という評価とは整合しない。

もっとも、変動源を実行ステップ 4 に限定できると断ずることもできない。実行ステップ 1 では構成要件の誤対応付けが生じており、これは対話による是正を要した。ログが公開されていない以上、ど

[18] 原論文 53 頁（6．1　AI の判断傾向とロジックの限界）。

のステップがどの程度安定していたかを分離することはできない。確実にいえるのは、実行ステップ4の外部検索が変動源の一つであることが観察された、という限度である。

とはいえ、この観察自体に価値がある。外部検索を組み込んだ時点でシステムは決定論的でなくなる、という知見は、フレームワークの限界としてではなく主要な結果として提示するに値する。次稿では、ステップごとの中間出力を全試行について記録し、どの工程が変動しているかを分離して報告することが望まれる。

4.3　人間の介入という交絡

実験1・2のいずれにおいても、著者は対話によりAIの理解を修正している。注意事項3.3がその手順を定型化していることは前述したとおりである。したがって、報告されている結果は介入込みの性能であり、フレームワーク自体の寄与は分離されていない。

改善は容易である。（一）介入なし条件（プロンプト投入後、一切の追加入力を行わない）を対照群として設け、（二）介入回数・介入内容をログとして記録し、（三）両条件の結論一致率を比較すればよい。これにより「フレームワークだけでどこまで到達し、人間の介入が何を追加したのか」という、本稿が本来問おうとした問いに直接答えることができる。

4.4　参照ラベルと再現性のための記載

実験2には、審査官による拒絶理由通知という外部から参照可能な判断例が存在し、AIはこれと異なる結論を維持した。ただし、**拒絶理由通知は新規性・進歩性についての法的な正解を意味するものではない。**出願人の意見書・補正、拒絶査定不服審判における審決、さらに審決取消訴訟によって覆り得るものであり、あくまで評価用の参照ラベルとして位置づけるべきである。実験1には、そうした外部参照可能な判断例すら存在せず、著者の誘導後の結論が事実上の基準として扱われている。評価枠組みとしては不安定である。

また、使用環境の記述はPCのCPU・RAMに及ぶ[19]一方、判断結果を左右する条件、すなわちモデルのバージョン・実行日・生成条件・検索ツールの有効化状態・2件の実験でChatGPTとGeminiのいずれを用いたか（および両者の結論に差があったか）が記載されていない。再現性の観点からは、パーソナルコンピューターの仕様は結果に影響しない一方、モデルのバージョンは決定的に影響する。

4.5　次段階への提案 —— 評価基盤の構築

本稿は「ルールを書いて動かしてみた」段階にある。次に必要なのは評価基盤である。具体的には、J-PlatPat等から取得可能な拒絶理由通知、拒絶査定および審決、並びに裁判所ウェブサイトから取得

[19] 原論文49頁（3．4　使用環境）。

可能な審決取消訴訟の判決を参照資料として、請求項・引用文献・判断結果を組にしたデータセットを 30〜100 件規模で構築することが考えられる。なお、J-PlatPat で取得できる包袋書類の範囲は案件により異なる点に留意を要する。

その際、ラベルの信頼度は階層化すべきである。すなわち、確定判決および確定した審決を上位、未確定の審決および拒絶査定を中位、拒絶理由通知（意見書・補正により覆り得るもの）を下位の参照ラベルとし、階層ごとにスコアを分けて報告する。そもそも法的判断を機械学習上の「正解」とみなすこと自体に限界がある以上、「参照ラベル」「比較基準」という位置づけは維持されるべきである。そのうえで、（一）フレームワーク全体のスコア、（二）厳格ルールを一つずつ除去したときのスコア変化（アブレーション）、（三）技術分野・発明類型別のスコア差、を測定する。これによって初めて、12 箇条のうちどのルールが効いており、どのルールが害をなしているかが判別可能になる。ルールの改訂を経験と対話に委ねている限り、フレームワークの改善ループは回らない。

5. プロンプト設計そのものへの技術的批評

5.1 禁止語による規制の限界

厳格ルール 4 は特定の語の使用を禁じる。しかし大規模言語モデルに対する「〜という語を使うな」という指示は、当該語の出力を抑制するにとどまり、その語が指す推論過程を抑制しない。モデルは「技術常識」と書かないまま技術常識に基づく推論を行い、表面上はルールを遵守しているように見える。著者が「生成 AI は、一貫して【実行ステップ 1】の制限事項および【厳格ルール】を遵守し続けた」と評価する際、遵守されたのが語の水準にとどまる可能性を排除できていない。

制約は、禁止ではなく出力形式によって担保するのが定石である。各判断に対し「根拠：資料名／該当箇所の引用／公開日」を必須フィールドとして課し、埋まらなければ処理を停止させれば、根拠を欠いた判断を検知しやすくなる。著者は【強制フラグ】[20]でこの発想を部分的に実装しており、設計上の優れた工夫である。これを全ルールに拡張すべきである。

ただし、必須フィールド化によって主観的補完が「構造的に排除される」わけではない点は、正確に述べておく必要がある。形式の充足を強制しても、モデルが実在しない資料名や URL、実在しない引用箇所で欄を埋めることは防げない。構造化出力の技術についても、保証されるのはスキーマへの準拠であって、出力される値の内容的な正しさではない [21]。したがって、必須フィールド化は「検知を

[20] 原論文 47 頁【強制フラグ】。

[21] 例えば OpenAI「Introducing Structured Outputs in the API」（https://openai.com/index/introducing-structured-outputs-in-the-api/）参照。構造化出力は出力がスキーマに準拠することを担保する機能であり、出力される値が事実として正しいことを担保するものではない。

容易にする措置」にとどまり、URL・引用箇所・公開日の実在確認を、AI の推論とは独立した外部検証工程として必須化しなければ、実効性をもたない。

5.2 【強制フラグ】の非対称性が生む出願人有利バイアス

【強制フラグ】は「このフラグが出力されない限り、進歩性を否定してはならない」と定める。すなわち、根拠資料が見つからないこと（＝探索の失敗）が、直ちに『想定外のベネフィット』として進歩性肯定に結びつく。デフォルトが「進歩性肯定」に設定されている。

この設計は、出願人側のドラフティング支援ツールとしては合理的だが、プロンプト冒頭の「特許庁審査官として振る舞ってください」というペルソナ設定[22]とは方向が逆である。ペルソナが求める挙動と制約が求める挙動が引っ張り合う構成になっており、モデルの挙動を不安定にする要因となり得る。実験 2 で AI が「頑なに『進歩性は肯定される』の結論を維持し続けた」ことも、この非対称なデフォルトの帰結として説明できる可能性がある。

改善としては、出力を「肯定／否定」の二値ではなく「肯定／否定／判断保留（根拠未検証）」の三値とし、探索の失敗を保留として明示することが考えられる。「調べたが見つからなかった」と「調べていない／調べられなかった」を区別しない設計は、実務上そのまま誤導につながる。

5.3 リセット指示とアンカリング

実行ステップ 1 は、全構成が共通項に該当しない場合、「【実行ステップ 2】開始時点で、【実行ステップ 1】で広義に抽出した結果を全てリセットし」と指示する[23]。しかし言語モデルに「忘れよ」と命じても、当該文脈はコンテキストウィンドウ内に残存し、後続の推論に影響する（アンカリング）。実行ステップ 1 で「最大限に共通項を探し出す強い意志」を反映させた直後に、同じ文脈で厳密な差分認定を求めることは、設計として無理がある。

技術的な解決は明快である。実行ステップ 1 と実行ステップ 2 以降を別セッション（別コンテキスト）に分離し、ステップ 1 の出力のうち＜共通項＞リストのみを機械的に受け渡せばよい。工程ごとにコンテキストを分離するパイプライン化は、判断の独立性を担保する常套手段であり、本フレームワークとの相性がよい。

5.4 実行ステップ 1 の早期終了が情報を捨てている

実行ステップ 1 は、全構成が共通項に該当した時点で処理を終了し「新規性は否定される」と回答して以降を無視する。しかし実務が知りたいのは、新規性が否定されるとしてどの構成を補正すれば進

---

[22] 原論文 46 頁【実行ステップ 1】冒頭の役割設定。

[23] 原論文 46 頁【実行ステップ 1】末尾のリセット指示。

歩性の議論に持ち込めるか、である。早期終了せず、共通項・差分の全体表を必ず出力させたうえで結論を付す設計のほうが、実務的な情報量は大きい。

5.5　構造化出力による監査可能性

現行の【最終出力フォーマット】は自然文と見出しによる出力である。API の構造化出力機能等を利用して JSON スキーマ（構成要件 ID、公報 A 中の対応箇所、共通／差分の別、対応する効果、根拠資料URL、公開日、判定、フラグ発動の有無）への準拠を担保すれば、複数回試行の機械的比較が可能となり、4.1 で指摘した一貫性の定量測定がそのまま実現する。加えて、出力の集計・監査、事務所内での品質管理への組み込みも容易になる。一般向けブラウザインタフェースでは形式の厳密な強制までは保証できないため、この点は API 利用への移行が前提となる。

6. 実務導入の当否とリスク

6.1　守秘義務と情報管理 —— 提言として最も補うべき欠落

本稿の手法は、対象出願の請求項および明細書のテキストを、ウェブブラウザ経由の商用生成 AI サービスに入力するものである。対象が未公開出願または依頼者の非公開情報を含む場合には、守秘義務および情報管理上の問題が生じる。ところが本稿には、依頼者の秘密の管理に関する記述が一切ない。実務誌における「実務導入への提言」としては、この欠落は看過し難い。

検討されるべき事項は少なくとも次のとおりである。（一）消費者向けインタフェース経由の入力について、学習利用のオプトアウト設定が有効か、法人契約か個人契約か、（二）弁理士法上の守秘義務および弁理士倫理との関係で、依頼者への説明と同意取得を要するか、（三）入力データが第三者に開示され得る態様である場合、当該技術情報の営業秘密該当性（秘密管理性）に影響しないか、（四）事務所・企業知財部としての利用規程、ログ保存、入力禁止情報の定義。

これらは技術的検証の対象ではないが、提言の実行可能性を左右する前提条件である。次稿では独立の一節を設けることを強く勧める。

6.2　注意義務と検証プロトコル

AI の出力をもとに拒絶理由への応答方針（反論か補正か）を決定した場合、その出力に基準時未検証の資料や存在しない文献が含まれていれば、弁理士と依頼者との委任・準委任関係その他の契約関係によっては、専門家としての注意義務違反が問題となり得る。なお、企業知財部員、特許事務所、外部委託先では法的根拠および契約関係が異なるため、責任の所在は一律ではない。著者は「相変わらず弁理士の関与は欠かせないだろう」と述べるが、関与の内容が示されていない。

最低限、次の検証プロトコルが必要である。（一）実行ステップ 4 が提示した各資料について、URL

および引用箇所の実在を人手で確認する。（二）その資料の用法を区別する。当該資料を、差分構成またはその効果を直接開示する先行技術として用いるのであれば、公開日が出願日（適法な優先権主張の効果が及ぶ場合は優先日）より前であることの確認が不可欠である。他方、基準時における技術常識を示す補助証拠として用いるにとどまるのであれば、公開日が基準時後であることは直ちに排除事由とはならないが、その内容が基準時における当業者の認識を示すものであることを別途確認しなければならない。（三）＜共通項＞＜差分＞の対応表を、請求項の文言と公報の記載に照らして全件確認する。（四）AI が「記載がない」と述べた事項について、記載がないことの確認は探索の網羅性に依存する以上、それ自体を結論の根拠として用いない。

6.3　審査官側での利用について

本稿は、AI が差分の抽出と客観的資料の検索を担い、審査官はより厳密な法理の考察に注力する、という役割分担を展望する [24]。方向としては首肯できる。もっとも、拒絶理由通知は、出願人に意見書提出の機会を与え（特許法 50 条）、あわせて同条に規定する期間内の補正の機会（同法 17 条の 2 第 1 項）を開く重要な手続行為であり、拒絶の理由は具体的に通知される必要がある。そのため、AI が提示した根拠を審査官が検証せずに採用することは、通知内容の適正性および説明責任の観点から問題を生じ得る。

とりわけ実行ステップ 4 のような自由なウェブ検索は、特許法 29 条 1 項 3 号における「電気通信回線を通じて公衆に利用可能となった発明」の公開日認定を厳密に要求する。ここでも基準時の問題が反復して現れる。

6.4　「一貫して誤る」というリスク

一貫性は、それ自体が品質を意味しない。同一の誤りを安定して再現するシステムは、誤りが偶発的に混入するシステムより、実務上むしろ危険である。誤りが系統的であるため、複数回の試行による相互チェックが機能しないからである。

実験 2 は、少なくとも審査官の判断とは異なる結論を、反復的に維持した事例として読める。いずれが法的に正しいかは別問題であるが、フレームワークが自らの結論を修正する契機を内蔵していない点は指摘しておく必要がある。一貫性が価値をもつのは、監査可能性の前提としてであって、正確性の代理指標としてではない。実務導入の評価軸は、一貫性ではなく「誤り方が予測可能で、かつ人間が検知できる形をとっているか」に置くべきである。

6.5　最も有望な用途 —— 判断ではなく起案の支援

[24] 原論文 53 頁（6．2（2）　審査官の役割の高度化）。

以上の限界を踏まえたとき、本フレームワークの実務価値は「進歩性あり／なし」という結論の生成にはない。価値があるのは中間生成物である。すなわち、（一）請求項の構成要件と引用文献との対応表の自動生成、（二）差分の網羅的抽出と抜け落ちの検出、（三）各差分に対応する効果が明細書に記載されているかの機械的チェック、である。

とりわけ（三）は、出願前ドラフティングの品質管理としてそのまま使える。「各構成要件について、それが奏する効果が明細書に記載されているか」という点検は、進歩性の主張の準備として直接に有用である。加えて、発明の課題解決との関係や実施可能性との関係において、場合により特許法36条4項1号・6項1号の記載要件の充実にも資する。もっとも、各構成要件ごとの効果の記載がこれらの条項の要件そのものであるわけではないから、この点検はあくまで実務上の規律であって法定要件のチェックリストではない。著者が述べた「ベネフィットの記述は『動機づけ』を可視化する」という指摘[25]は、この転回を示唆している。ただし、本願明細書の効果記載が可視化するのは、直接には＜差分＞構成の技術的意義および有利な効果であって、審査基準にいう動機付け（公報Aに公報Bを適用する契機の有無）そのものではない。両者は区別して論じる必要がある。フレームワークを「審査官の代替」ではなく「明細書起案の規律」として位置づけ直すとき、本稿の提言は最も強い説得力をもつ。

また、対話による是正が毎回必要であるならば、熟練者の時間は必ずしも削減されない。むしろ「AIを説得する」工数が加わる。恩恵が大きいのは、経験の浅い実務者に対する思考手順の教示と、対応表・差分表という定型作業の下書き生成である。「無駄な作業」の削減という著者の主張は、この限定のもとでこそ妥当する。

7. 結び

本稿の核心的な貢献は、進歩性判断の一部を、実行可能な手続として書き下した点にある。書き下したからこそ、どこが書き下せていないかが可視化された。動機付け判断の不在、本願発明の認定における明細書不参酌の指示、基準時の検証工程の欠落、ルールの適用範囲がプロンプト構造に依存していること、外部検索がもたらす非決定性は、いずれも著者がフレームワークを公開した結果として初めて検証可能になった論点である。この意味で、本稿は批判に供された時点で目的の相当部分を達している。

評釈者として次稿に期待したいのは、次の四点である。第一に、フレームワークの射程を「効果の予測可能性判定」に限定して明示し、副引用発明の特定を含む動機付け判断を独立の実行ステップとして追加すること。第二に、実行ステップ4に、根拠メタデータ（URL・公開日・引用箇所）の必須化

[25] 前掲注5に同じ。

と、資料の用法（直接の先行技術か、技術常識を示す補助証拠か）に応じた基準時の検証工程を組み込み、その実在確認を AI の推論から独立した工程とすること。第三に、拒絶理由通知・査定・審決・確定判決を信頼度に応じて階層化した参照ラベルとする小規模ベンチマークを構築し、厳格ルールのアブレーションによって各ルールの寄与を測定すること。第四に、守秘義務・情報管理に関する独立の検討を加えること。

「進歩性」という非物理的な思考領域において判断過程の言語化が重要である、という著者の認識には全面的に同意する。ただし、言語化されたものが法理として正しいかは、言語化とは別個の検証を要する。本稿が切り開いたのは、その検証を集団的に行うための入口である。

（本評釈は、公開された原論文に基づく批評であり、いかなる組織の公式見解も示すものではない。）

主要参照資料

●山崎薫「生成 AI に基づく進歩性判断のモデル化評価」パテント 2026 年 Vol.79 No.6・44〜54 頁（原稿受領 2025 年 10 月 30 日）。日本弁理士会ウェブサイト　https://jpaa-patent.info/patent/viewPdf/4836

●特許庁「特許・実用新案審査基準」第 III 部第 2 章第 2 節「進歩性」（進歩性の判断に係る基本的な考え方、動機付け、設計変更等、先行技術の単なる寄せ集め、引用発明と比較した有利な効果、阻害要因）　https://www.jpo.go.jp/system/laws/rule/guideline/patent/tukujitu_kijun/ht/03_0200.html

●特許庁「特許・実用新案審査基準」第 I 部第 2 章第 1 節「本願発明の認定」https://www.jpo.go.jp/system/laws/rule/guideline/patent/tukujitu_kijun/ht/01_0200.html

●特許庁「審査基準の改訂について」（令和 8 年 7 月 1 日以降に行われる審査に適用。第 III 部第 2 章第 2 節「進歩性」につき阻害要因等に関する記載を明確化）https://www.jpo.go.jp/system/laws/rule/guideline/patent/tukujitu_kijun/kaitei2/r8_shinsa_kijun_kaitei.html

●最高裁判所第三小法廷令和元年 8 月 27 日判決・平成 30 年（行ヒ）第 69 号（審決取消請求事件。特許第 3068858 号に係る事件。原審：知的財産高等裁判所平成 29 年 11 月 21 日判決・平成 29 年（行ケ）第 10003 号）　https://www.courts.go.jp/assets/hanrei/hanrei-pdf-88888.pdf

●特許法 29 条 1 項 3 号、同条 2 項、17 条の 2 第 1 項、36 条 4 項 1 号・6 項 1 号、50 条