特許公報の AI 読解に関する講習会の評釈

肯定的見解と慎重な見解を踏まえた品質管理と分類設計の検討

対象　酒井美里氏による講習会 Q&A 前編及び後編
作成日　2026 年 11 月 8 日

GPT-6 Astra

要旨

本 Q&A は、生成 AI を特許調査に組み込む際の品質管理を具体的に示した実務報告として、高く評価できる。ただし、分類結果の正確性や、人が行った修正の妥当性については、さらに検証が必要である。特に「発明の主役」という分類には、請求項上の位置づけと、技術的な課題解決への貢献が混在しており、両者を分けて評価することが重要である。

本稿は、Q&A の前編・後編[1][2]、関連研究、記事で引用された特許公報を基に、肯定的評価と慎重な評価を整理する。以下の賛否は、当該記事に対する既発表の書評を紹介したものではなく、各資料に照らした本稿の評価である。文中の番号は、文末の参考文献に対応する。

1　肯定的に評価できる点

1 1　作業の分解と検証可能性

最も評価できるのは、作業を分解し、それぞれに適した確認方法を設けていることである。記事は、段落・請求項ごとにデータを整理し、計数や文字列照合を Python に、文脈の読解を大規模言語モデル（LLM）に、判断の妥当性の確認を人に担当させている。これにより、問題が起きた場合に、抽出漏れ、根拠引用の誤り、意味理解の誤りを切り分けやすくなる。[1][2]

この設計は、Anthropic が示す、処理を段階に分けて途中にプログラムによる確認を挟むワークフローとも整合する。ただし、これは設計思想を支持するものであって、今回の分類精度を直接証明するものではない。[3]

1 2　失敗と修正過程の開示

成功例だけでなく、失敗と修正の過程を公開している点にも価値がある。初回出力で成分が抜けたこと、照合規則の側が誤っていたこと、分類定義を変更すると新たな食い違いが生じたことまで示されている。特許実務への導入では、こうした失敗を発見・修正できることが重要である。[1][2]

特に、引用した文章が実在することと、その文章から導いた判断が妥当であることを区別した説明は適切である。引用付き回答を扱う Gao らの研究でも、回答の正しさと引用の品質は別の評価対象とされている。[4]

2　報告された数値の評価

記事の数値は、何を検証した指標なのかを明確にして読む必要がある。機械照合の通過率、成分の出力漏れ、人との判断の相違、分類定義への感度、費用は、それぞれ異なる側面を示している。[1][2]

報告された結果	評価できる点	そこからは確認できない点
194 件すべてが最終的に機械照合を通過	設定した照合条件を満たすまで処理できた	194 件すべての役割分類が正しいこと
初回に 991 成分中 76 成分が出力されなかった	事前計数で既知の対象の漏れを発見できた	辞書や入力データの段階で漏れた成分の有無
人が確認した 20 件中 3 件で判断が相違	人の確認が実際に問題を発見した	全体の誤判定率が 15％であること
定義変更後に 317 行中 143 行の分類が変化	定義が結果を大きく左右することを示した	変更した 143 行がすべて正解になったこと
API 費用約 116.5 ドル　実行時間約 1 時間 36 分	処理規模と費用を具体的に把握できる	設計・確認・修正を含めた総費用や総作業時間

前編では初回 165 件、2 回目 28 件、3 回目 1 件が照合に一致したと報告されている。これは再試行を含む照合の結果であり、役割分類の正解率とは区別しなければならない。また、定義変更により分類が変わった 143 行は対象 317 行の約 45.1％に相当するが、その割合は精度向上率を意味しない。[1][2]

3　分類定義に関する否定的及び慎重な見解

3 1　異なる評価軸が混在している

慎重な見方からの最も強い指摘は、正解の定義自体が十分に固定されていないという点である。前編の分類は、A が「発明の主役」、B・C が成分の機能、D・E が記載のされ方、F が配合上の制限である。これらは同じ軸の区分ではない。[1]

例えば、ある成分が甘味剤として働きながら、請求項の必須構成であり、配合量にも上限があることは十分に考えられる。単一の区分に分類するなら、どの性質を優先するかを定める必要がある。後編で請求項ランクを別軸にした改善は妥当であるが、技術的な役割と請求項での位置づけの分離は、なお徹底する余地がある。[2]

3 2　特許第 5797243 号のマルチトール

記事では、AI がマルチトールを A としたのに対し、著者は、経時安定性を高める成分であることから B に修正している。[2] しかし、公報【0005】ではアスコルビン酸類の経時安定性向上そのものが課題とされ、【0006】には、その解決手段となる成分の一つとしてマルチトールが挙げられている。[5]

評価軸	公報に基づく整理
登録請求項で具体的に要求されているか	登録請求項 1 にはマルチトールは列挙されていない
明細書で課題解決に寄与すると説明されているか	安定性向上に寄与する成分の一つとして説明されている

評価軸	公報に基づく整理
成分の機能は何か	アスコルビン酸類の安定化

したがって、安定化に使われる成分だから課題解決の中心ではないとは直ちにいえない。安定化そのものが発明の課題だからである。登録請求項を重視する分類なら B への修正は理解できるが、明細書全体の技術的貢献を重視するなら、AI の A 判定を明白な誤りとは断定しにくい。これは AI の判断が正しいと確定する趣旨ではなく、分類定義と評価対象の明確化が必要であるという評釈である。

なお、登録請求項 1 は、寒天とエリスリトールから選ばれる少なくとも一種を含む構成である。糖アルコール類のうち具体名が記載されているのはエリスリトール、という説明はできるが、エリスリトールが常に必須という意味ではない。[5]

3 3　特許第 4959006 号の糖アルコール

特許第 4959006 号も同様である。個別の糖アルコールが従属項で選択肢として列挙されていることと、糖アルコールという成分群が発明に重要でないことは別である。請求項 1 では糖アルコールが構成要件となり、【0014】では、ヘスペリジン、特定の二価アルコール、糖アルコールの組合せによる安定化が説明されている。[6]

個別成分の必須性を低く評価することは合理的でも、それを技術的貢献の低さに直結させるべきではない。成分群の重要性、個別成分の選択可能性、登録請求項上の位置づけを、それぞれ保持できる分類の方が、調査目的の違いに対応しやすい。

4　機械照合と人による確認の限界

4 1　入力データ自体の正確性

機械照合では、入力データ自体の正確性を別に確認する必要がある。記事では、表を LLM で読み取って JSON 化し、その JSON と後段の回答を照合する。[1] この場合、表の読み取り段階で誤りが入ると、その誤りと一致する回答が照合を通る可能性がある。例えば、元の表の「5.0」を「50.0」と誤読した場合、後段の一致確認だけでは訂正できない。この例は、本稿が説明のために設けた仮例である。

同じことは漏れにも当てはまる。辞書にない同義語、商品名、構造式だけで示された成分などは、事前計数でも見落とされる可能性がある。記事自身は、判読漏れの一部を検証できると限定しており、その限定は適切である。導入側も、検証範囲を広く解釈しすぎないことが大切である。[1]

照合規則を修正するときは、正しい回答を誤って不合格にする問題だけでなく、誤った回答を合格させていないかも確認すべきである。既知の失敗例を直すことと、未知の公報に対する性能を確かめることは、別の検証になる。

4 2　抜き取り確認と評価者の判断

人の確認についても、人が見たという事実だけで品質を保証することはできない。今回の 20 件には、要注意の 6 件と、役割ごとに抽出した 14 件が含まれる。問題発見のためには合理的だが、母集団全体の誤判定

率をそのまま推定する標本とは異なる。さらに、3 件の相違のうち 2 件は定義の違いによるものであり、単純に AI の読解ミス 3 件と扱うことも適切ではない。[2]

関連する慎重論として、Vaccaro らの 106 実験を対象としたメタ分析では、人と AI の組合せは平均的には人単独を上回る一方、人単独・AI 単独のうち優れた方を上回るとは限らなかった。同研究は 2020 年から 2023 年までに公表された実験を対象とし、今回の手法を直接評価したものではないが、人の関与の仕方そのものを検証すべきことを示している。[7]

本件では、一部の公報について、別の熟練者が AI の分類を見ずに独立判定し、その後で比較すると、モデルの誤り、分類定義の曖昧さ、人による解釈差を分けやすくなる。評価者間で判断が一致しない公報は、無理に単一の正解を付けるのではなく、判断条件や相違理由を記録することが望ましい。

5　入力形式と効率性の評価

PDF をそのまま渡さないという方針は、今回の用途に適した選択として評価すべきである。段落単位の管理と機械照合には JSON 化が有効だが、JSON という形式自体が精度を保証するわけではない。重要なのは、公報番号、段落番号、表のセル、注記などを対応づけ、原文へ戻れることである。

Claude の公式仕様には、PDF のテキストとページ画像を併用して解析する方式も示されている。したがって、少数公報の読解や図面中心の案件まで、一律に JSON 化を必須とする必要はない。処理規模、図表の重要性、再利用の頻度、検証の必要性に応じて入力形式を選択すべきである。[8]

費用について、記事は設計・修正時間を含まないことを明示しており、説明は誠実である。[2] ただし、AI なしの比較が 3 件である以上、判読時間が一般に半分以下になるといった結論には進めない。同じ品質水準で、準備から最終確認までの総工数を比較する必要がある。

記事が示す API 費用約 116.5 ドルは 194 件当たりの処理費用であり、単純平均では 1 件当たり約 0.60 ドルとなる。これは設計・保守・人手確認やデータ取得費用を含む単価ではない。費用の評価では、既存の仕組みを反復利用する場合と、案件ごとに新たな分類体系を設計する場合を分けることが有用である。[2]

6　総合的評釈と実務への提言

本 Q&A は、実務導入の設計として有力であり、評価基準の設計をさらに精緻化すれば、説得力が大きく増す。特に、根拠の実在を確認する仕組み、失敗を検出する仕組み、定義を修正する仕組みを一体として示した点に貢献がある。今後の改善では、次の三点を優先したい。

第一に、分類を複数の軸に分ける。請求項上の位置づけ、成分の機能、課題解決への貢献、実施例・比較例での使用を別々に記録する。判断不可も、情報不足、分類体系に該当区分がない場合、解釈が分かれる場合に分ける。

第二に、原資料と判断の両方を検証する。原公報から JSON への変換精度を確認したうえで、定義を固定した評価用公報を用意し、修正に使っていないデータで分類性能を測る。別担当者による独立判定も組み合わせる。

第三に、調査目的ごとに利用範囲を決める。技術動向の整理や一次仕分けでは有用性が高い。一方、侵害予防調査（FTO）での除外や無効資料候補の切捨てに用いるなら、主役ではないという分類だけで対象から外さず、請求項や具体的開示の確認へつなげる。

この Q&A の大きな貢献は、AI の誤りを見つけて改善する過程を、実務者が追える形にしたことである。その考え方をさらに進めるなら、AI の回答だけでなく、人が用意した分類軸と、人が行った修正も検証対象にすることが、次の課題となる。

参考文献

ウェブ資料の参照日は、すべて 2026 年 11 月 8 日。原記事の表・数値は文献[1][2]に基づく。関連研究は設計及び評価の考え方を検討するために参照したものであり、今回のワークフローを直接検証した研究ではない。

[1] 酒井美里．講習会Q＆A（前編）「どうやって公報を AI に読ませてますか？」に実データで答えます．note，2026 年 10 月 12 日．

https://note.com/sakaimisato/n/n17a8eb0c2b46

参照箇所：役割分担、JSON 化、成分の役割分類、機械照合、出力漏れ、照合規則の修正、配合量換算。

[2] 酒井美里．講習会Q＆A（後編）「どうやって公報を AI に読ませてますか？」に実データで答えます．note，2026 年 10 月 13 日．

https://note.com/sakaimisato/n/nf2e85566bff0

参照箇所：請求項ランク、20 件の目視確認、分類定義の変更、時間と費用。

[3] Anthropic. Building effective agents. 19 December 2024.

https://www.anthropic.com/engineering/building-effective-agents

参照箇所：Workflow: Prompt chaining、When (and when not) to use agents。

[4] Gao, Tianyu; Yen, Howard; Yu, Jiatong; Chen, Danqi. Enabling Large Language Models to Generate Text with Citations. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 6465–6488. DOI: 10.18653/v1/2023.emnlp-main.398.

https://aclanthology.org/2023.emnlp-main.398/

参照箇所：ALCE による回答の正しさと引用品質の評価。本稿では評価軸の区別を参照し、当時の性能値を今回の手法に転用していない。

[5] 日本国特許庁．特許第 5797243 号公報（JP5797243B2）．アスコルビン酸及びその類縁体を安定配合した組成物．特許権者：サンスター株式会社．発行日：2015 年 10 月 21 日．

https://patentimages.storage.googleapis.com/c3/c1/15/7d6b73fb295214/JP5797243B2.pdf

参照箇所：請求項 1、段落【0005】～【0007】。原公報 PDF の公開写しを参照。明細書中の「項 1」と登録請求項 1 を区別した。

[6] 日本国特許庁．特許第 4959006 号公報（JP4959006B2）．ヘスペリジン含有組成物．特許権者：サンスター株式会社．発行日：2012 年 6 月 20 日．

https://patentimages.storage.googleapis.com/f9/2f/38/a220406ed56e54/JP4959006B2.pdf

参照箇所：請求項 1・2、段落【0014】。原公報 PDF の公開写しを参照。

[7] Vaccaro, Michelle; Almaatouq, Abdullah; Malone, Thomas. When combinations of humans and AI are useful: A systematic review and meta-analysis. Nature Human Behaviour. 2024, vol. 8, pp. 2293–2303. Published 28 October 2024. DOI: 10.1038/s41562-024-02024-1.

https://www.nature.com/articles/s41562-024-02024-1

参照箇所：Abstract、Results、Discussion。対象は 2020 年 1 月から 2023 年 6 月までに公表された 106 実験、370 効果量。

[8] Anthropic. PDF support. Claude Platform Docs. 公開・更新日の表示なし.

https://platform.claude.com/docs/en/build-with-claude/pdf-support

参照箇所：How PDF support works、Optimize PDF processing。