誰が科学に異議を唱えられるのか？――生成AIと科学的精査の民主化

Who Gets to Contest Science? Generative AI and the Democratization of Scientific Scrutiny

George Kujo（九条丈二）

Independent Researcher | Tokyo, Japan

ORCID: 0009-0006-6864-0145
george.kujo.irjp@gmail.com

日本語訳 | v1.0 | 2026年9月6日

英語版 v1.0 の全文日本語訳。参考文献は原語表記。

要旨

生成AIは、執筆、コーディング、分析への障壁を下げることで、科学的知識の生産を民主化しうる技術として論じられることが増えている。一方、既存の評価上の役割の外で科学的主張を精査する際の障壁も下げうるという、並行する可能性には十分な注意が払われていない。本稿は、この変化を概念的に論じる。科学的資料へのアクセスと、それらを吟味する能力とを区別し、生成AIが科学的知識の生産と科学的精査の二重の民主化に寄与しうると主張する。統計的解釈、コードの検査、文献比較、引用の確認、方法論的推論を支援することで、AIは評価能力の一部を正式な査読者や制度的に承認された専門家の外へ広げうる。そこには「能力と承認の不一致」が生じる。妥当である可能性のある科学的問題を発見できる人と、その批判が制度によって承認される人とは一致しない場合がある。しかし、精査の拡大は品質管理の問題も生む。同じシステムが、見せかけの精査、幻覚に基づく異議、技術的には高度に見えても検証できない主張を生み出すコストも下げるからである。したがって制度的な課題は、より踏み込んだ審査や対応に進めるべき異議を、そうでない異議からどう区別するかにある。私は、科学的問題の発見への参加を比較的開かれたものとしつつ、制度的な措置に至る前に、証拠の来歴、再現可能性、証拠の具体性、専門家による裁定への要件を段階的に強める、層状のガバナンス原則を論じる。生成AIは、誰が科学的主張を生み出せるかだけでなく、誰がそれらに実際に異議を唱えられるかを変えることによっても、科学を民主化しうる。

キーワード： 生成AI；科学的精査；オープンサイエンス；評価能力；制度的承認；科学ガバナンス

1. 序論

生成AIは、科学的活動の中に急速に組み込まれつつある。大規模言語モデル（LLM）と関連システムは、文献の検索と統合、コーディング、データ分析、原稿作成、仮説生成、さらに科学的発見のますます複雑な構成要素を支援できる（Reddy & Shojaee, 2025; Zhang et al., 2025）。近年の研究は、これらのツールが個々の作業の遂行方法を変えるだけでなく、科学的生産そのもののパターンを変え始めていることを示唆する。異なる技術的時代にわたるAIツールを扱った研究で、Hao et al. (2026) は4,100万件を超える科学論文を分析し、AIツールの採用が個々の科学者による論文発表数と被引用数の大幅な増加と関連する一方、科学的活動は既存の研究領域へいっそう集中していたと報告した。より広くは、生成AIは、それまで必要な専門技能の一部を持たなかった利用者にも認知的負荷の高い課題への取り組みを可能にするため、知識労働を民主化しうる技術とされている（Daepp et al., 2026）。そのため、AIと科学をめぐる議論の多くは、ますます馴染み深い問い、すなわち**誰が科学的知識を生み出せるようになるのか**に焦点を当ててきた。

しかし、科学のシステムは主張の生産だけでなく、その批判にも依存する。公表された知見は、解釈され、先行文献と照合され、分析上の弱点を調べられ、可能であれば再現され、理由があれば異議を唱えられ、成立しなければ訂正されなければならない。生成AIに関する既存の議論は、研究生産性、科学的発見、原稿作成、そして近年では査読の内部でのAI利用を広く検討してきた。例えば、学術査読におけるAIに関する189件の資料を対象とした2026年のスコーピングレビューは、編集段階のトリアージや査読者支援から査読の自動生成までの応用を記録している（Nabavi et al., 2026）。これらの展開は重要であるが、主として**すでに確立された評価上の役割の内部**におけるAI利用に関わる。生成AIが、そのような正式な役割の外で科学的主張を精査する実践的能力を誰が持つかを変えうるという別の可能性には、それほど注意が向けられていない。

この可能性は、オープンサイエンスや開かれた参加に還元できない。オープンサイエンスはすでに、科学出版物、データ、ソフトウェア、基盤、研究過程のアクセス可能性と再利用可能性を高めようとしてきた。市民科学やその他の参加型モデルも、社会の多様な主体が知識生産に参加する機会を広げてきた（UNESCO, 2021）。オープン査読にも、批判を招待された査読者だけに委ねず、より広い共同体の成員が評価的コメントを寄せられる「開かれた参加」の形態が含まれる（Ross-Hellauer, 2017）。PubPeerのような出版後のプラットフォームも、原稿の採択で精査が終わる必要はないことを示している（PubPeer, n.d.）。それでも、参加の機会は、参加する実践的能力と同じではない。自由に閲覧できる回帰表も、統計の知識がなければ吟味は難しい。公開されたソースコードも、プログラミング能力がなければ評価は難しい。利用できる補足付録も、その分析の論理を再構成できない読者には、実質的にアクセス不能なままでありうる。市民科学の検証実務も、幅広い参加と分散した貢献の品質確認が両立しうることを示している（Baker et al., 2021）。**したがって、形式的な開放性が、実効的な評価への参加を必ずしも生み出すわけではない。**

生成AIは、この第二の層に介入しうる。専門用語を言い換え、統計手続きを説明し、コードを生成・解釈し、競合する文献群を要約し、主張と引用を特定し、頑健性の確認を提案し、利用者による分析手続きの再構成を支援できる。AIによる平易化が、科学的資料に対する非専門家の読解を改善しうるという実験的証拠はすでに存在する（Markowitz, 2024）。もっとも、読解だけでは科学的専門性と同等とは到底いえない。この支援は、科学的主張に実質的に関与するための認知的・技術的障壁を低くしうる。この区別は重要である。オープンアクセスは、ある人が論文に到達できるかを変える。AI支援は、到達

した後にその人が実際に何をできるかを変えうる。原理的には、科学的主張の生産への障壁を下げる同じ技術変化が、その主張を吟味する際の障壁の一部も下げうるのである。

私は後者の可能性を**AI支援による分散型精査**と呼ぶ。これはAI支援査読と区別されるべきである。AI支援査読は、すでに確立された評価上の立場を占める主体によるAI利用を指す。例えば、編集者によるトリアージへのAI利用や、査読者による原稿の点検へのLLM利用である。AI支援による分散型精査は、当該分野の外にいる研究者、実務家、独立研究者、ジャーナリスト、患者団体や公益団体、市民科学者、技術的能力を持つ一般市民という、より広い人間の主体が、公表された科学的主張を理解、検証、再現し、異議を唱え、その他の形で吟味するためにAIを利用することに関わる。新しいのは、部外者が科学を批判できることではない。それは以前から行われてきた。また、従来の査読の外に出版後査読が存在することでもない。ここで問題となる変化は、そうした主体が科学の専門的資料にどれほど深く、どれほど大規模に関与できるかを制限してきた能力上の障壁の一部を、生成AIが低くしうることである。

しかし、評価能力がより広く分布しても、科学的権威が自動的に同じように分布するわけではない。科学的批判は従来、その信頼性の一部を制度的役割や地位のシグナルから得てきた。分野の専門性、学術資格、所属機関、編集者による選定、同業者の承認、確立された実績などである。これらのシグナルは不完全だが、社会認識論における長年の承認問題の解決を助ける。専門性を独力で評価できない人も、誰の科学的判断を重く受け止めるべきかを決めなければならない。近年の研究も、非専門家が認識的な信頼に値するかを評価する際の、資格と関連指標の重要性と限界の両方を示し続けている（Mirkin & Sojka, 2026）。したがって、生成AIは非対称性を生みうる。科学的主張を吟味する実践的能力が、正当な科学的批判が従来承認されてきた制度的役割を超えて広がる可能性がある。私はこれを専門性の消滅ではなく、起こりうる**能力と承認の不一致**と捉える。

このような展開は、批判の増加が必ず科学の信頼性を高めることを意味しない。AIは妥当な精査のコストを下げうるが、表面的な異議、捏造された引用に基づく批判、方法論の形だけの模倣、過度に確信的な見せかけの精査のコストも下げうる。非専門家によるLLM利用から得られた証拠は、AIシステムへのアクセスだけでは証拠に基づく評価が保証されないことをすでに示唆している。利用者は低品質の出力を得ることがあり、利用の仕方を改善する簡単な介入を行っても、相当な品質上の欠陥は解消されない（Rebitschek et al., 2025）。出版後査読自体も、制度側に相当な処理負担を生じさせてきた（Caron et al., 2025）。したがって中心となるガバナンス上の問題は、すべての異議を等しく妥当と認める方法ではなく、**信頼に値する精査**を、単に低コストな精査から区別する方法である。本稿はこの問題を三つの問いによって展開する。生成AIは科学的主張の生産と精査に必要な能力上の障壁をどのように変えるのか。AI支援による分散型精査は、オープンサイエンス、市民科学、従来の査読、出版後査読とどう異なるのか。そして評価能力が、科学的評価に従来付随してきた制度的承認よりも広く分布すると、何が生じるのか。私は、生成AIが、第二の、十分に検討されていない意味で科学を民主化しうると主張する。それは科学的主張の生産に参加できる人を増やすだけでなく、その主張に実質的に異議を唱えられる人を増やすことである。**そこから生じる問いは、誰が科学を生み出せるかだけでなく、誰が科学に異議を唱えられるのかである。**

議論は、概念的役割に応じて番号を付した五つの関連命題からなる。P1は二重の民主化というテーゼを述べる。すなわちAIは、科学的生産と精査の双方について能力上の障壁を下げる。P2は、資料へのアクセスと、それを吟味する能力とを区別する。P3はP1の精査側の機構を特定し、作業コストの低下が、割り当てられた評価上の役割を超えて実質的な評価をどう広げうるかを示す。これはそのテーゼの系であり、独立した民主化の主張ではない。P4は、この能力と制度的承認との間に生じうる隔たりを特定する。

P5は、信頼に値する分散型精査のための証拠上の要件を特定する。以下ではまずアクセスと能力の区別（P2、第2節）を確立し、次いで包括的なテーゼ（P1、第3節）、精査側の系（P3、第4節）、制度的含意（P4とP5、第5節と第6節）を展開する。

2. アクセスから能力へ

科学の民主化は歴史的に、主としてアクセスと参加の拡大によって追求されてきた。オープンアクセス出版は科学文献への価格障壁を下げる。オープンデータとオープンコードの取り組みは、研究資料を点検・再利用できるようにする。オープンサイエンスはより広く、科学の過程と成果のアクセス可能性、透明性、再利用可能性を高めようとする（UNESCO, 2021; Wilkinson et al., 2016）。市民科学の取り組みは、データ収集から解釈、場合によっては研究設計までの活動に一般市民を関与させることで、専門研究者の外へ参加を拡大する（UNESCO, 2021）。オープン査読と出版後査読も同様に、科学的評価に参加しうる主体の範囲を広げる（Ross-Hellauer, 2017）。

これらの展開は、科学への制度的なアクセス可能性を大きく変えるものである。しかし、アクセスを有効に活用するために必要な知識、技能、時間、技術的能力から生じる、もう一群の障壁を取り除くわけではない。研究論文が無料で利用できても、解釈は難しいままかもしれない。データセットがダウンロードできても、その構造や分析に必要な統計ソフトウェアに不慣れな人には実際には利用できないかもしれない。ソースコードが公開されていても、プログラミング言語、計算環境、組み込まれた分析上の仮定を理解できない読者には不透明なままでありうる。したがって、科学的対象を公開することは、それを点検する機会を増やすが、そのために必要な能力を必ずしも分配しない。

この区別がとりわけ重要なのは、参加がすでに科学の開放性に関する既存のモデルへ組み込まれているためである。Ross-Hellauer (2017) によるオープン査読のシステマティックレビューは、より広い共同体が査読に貢献できるという「開かれた参加」を、オープン査読システムに繰り返し見られる特徴の一つとして特定した。そのため、従来招待される査読者集団の外にいる人が科学的成果にコメントできる可能性は、生成AIによって初めて生まれたものではない。公的批判、市民参加、出版後コメントも新しい現象ではない。重要な問いはむしろ、技術変化が、すでに資料への形式的アクセスを持つ人々にとって**実際に到達可能な参加の深さ**を変えるかどうかである。

そこで私は、**アクセス**と**能力**を区別する。アクセスの拡大は、科学的情報、資料、評価過程をより広い人々が利用できるかに関わる。能力上の障壁の低下は、その資料を用いて意味のある科学的作業を行うための認知的・技術的要件が緩和されるかに関わる。この区別は分析上のものであり、絶対的なものではない。アクセスはしばしば能力の前提条件であり、能力自体も教育、分野での経験、制度的支援、ソフトウェア、データ基盤、時間に依存しうる。それでも両者は等しくない。あるシステムは高度に開かれていても、その開放性を活用する技能を持つ限られた集団だけが実際に利用できる状態にありうる。

生成AIが重要なのは、形式的アクセスと実際の利用との間に介入できるからである。不慣れな統計手法に直面した読者は、推定量、その仮定、典型的な失敗の仕方について説明を求められる。不慣れな言語で書かれたコードは、翻訳、注釈付け、デバッグ、再構成ができる。補足付録を要約し、本論文の主張と結び付けることができる。確認のために引用を抽出できる。代替仕様を提案できる。専門用語を、より理解しやすい言葉に置き換えられる。これらの機能は、判断、領域知識、検証の必要性をなくすわけではないが、評価を始めるために必要な中間的理解を得るコストを下げうる。

科学コミュニケーションの証拠は、この機構の限定的な形を示す。Markowitz (2024) は、GPTによる科学的資料の平易化が、非専門家による科学情報の処理を容易にし、科学者に対する認識を変えうることを示した。この知見を、言語モデルが非専門家の読者を専門家へ変える証拠と解釈してはならない。それが示すのは、より限定的だが理論上重要なことである。科学の専門的情報と非専門家の読者を隔てる認知的障壁の少なくとも一部は、技術によって変えられる。この結果は研究評価ではなく理解に関するものだが、専門的資料の実際の理解可能性が固定されていないことを示す。

これが科学的精査にとって重要なのは、評価が専門家による単一の判断行為だけで成り立つことはまれだからである。評価は多数の小さな操作からなる。論文の中心的主張を特定すること、分析がその主張へ至る道筋を再構成すること、比較すべき文献を探すこと、引用先が帰属された命題を支持するか確認すること、統計モデルを理解すること、提供されたコードを実行すること、分析上の仮定を変更すること、見かけ上の不整合がより深い調査に値するか判断することなどである。従来、その多くは、専門的な技術技能を自ら持つか、その技能を持つ協力者にアクセスする必要があった。生成AIは、こうした中間段階の一部をますます媒介できるようになっている。

したがって、この変化は専門性の消滅ではなく、**精査に必要な能力の閾値**の変化として理解する方が適切である。三つの模式的事例を考えよう。以前は疫学論文の要旨を読むことはできても回帰仕様を評価できなかったジャーナリストが、今ではモデルの体系的な説明を得て、専門家の追加確認が必要な問いを特定できるかもしれない。学術研究の外にいる臨床家が、関連するプログラミング言語を以前から深く学んでいなくても、提供された分析コードを点検し、出力を再現し、簡単な代替仕様を試せるかもしれない。独立研究者が、従来ならはるかに多くの時間や専門的支援を必要とした規模で、数十の引用された主張を追跡し、研究間で補足資料を比較し、分析を再構成できるかもしれない。いずれの場合も、AIが分野内の権威を与えるわけではない。評価に先立ち、評価を支える作業のコスト構造を変えるのである。

この区別は、**AI支援による分散型精査**を市民科学の別名として扱うべきでない理由も明確にする。市民科学は通常、科学的知識の生産への参加に関わる。ただし、その形態は大きく異なり、解釈や共同設計を含む場合もある。分散型精査は、すでに科学的記録に入った、あるいは入ろうとする主張の評価に関わる。また、それはオープン査読と同義でもない。オープン査読は、誰が参加できるか、何が可視化されるか、査読過程をどう組織するかを変える。分散型精査は、それらの参加者が関与した際に、技術的に何をできる可能性があるかに関わる。一方は制度的な機会の構造を記述し、他方は変化する能力の構造を記述する。この区別は、単純な二次元モデルとして表せる。**能力を伴わないアクセス**は、形式的には開かれていても、大多数の人には実際に吟味することが難しい科学を表す。**アクセスを伴わない能力**は、関連する専門性を持っていても、非公開の論文、利用できないデータ、独占的なコード、アクセス不能な補足資料を点検できない場合を表す。

能力を伴うアクセスは、科学的対象と、それを吟味する実践的手段の双方が利用できるため、分散型精査に最も強い条件を与える。生成AIはアクセスの問題を解決せず、最後の条件を単独で作り出すわけでもない。その固有の意義は、形式的開放性が達成された後にも残る認知的・技術的コストの一部を減らし、個人を第一の条件から第三の条件へ移行させうる点にある。この枠組みは、先に導入したアクセスと能力の区別をP2として確立する。

P2 — アクセスと能力の区別。 科学的情報へのオープンアクセスは、その情報を解釈、検証し、異議を唱える実践的能力とは異なる。

この区別は、第4節でP3として展開する精査側の系の前提を与える。生成AIが、専門的解釈、コードの再構成、文献比較、引用検証、探索的再分析などの作業に必要な能力の閾値を下げうるなら、科学的評価能力は、従来同程度のコストでそれを行使できた人々よりも広い集団へ分布しうる。

これは、その結果として生じる評価が正しいことを確立しない。能力の閾値を下げることは、ツールの適切な利用と不適切な利用の双方を増やしうる。生成AIは統計モデルを正しくも誤っても説明できる。分析過程を忠実に再構成することも、微妙な誤りを持ち込むこともできる。本当の引用の不一致を特定することも、存在しない不一致を幻覚として作り出すこともできる。したがって、ここで展開した区別は、**精査を行う能力に関するものであり、その結果の妥当性に関するものではない**。この区別は後の議論で中心になる。科学的主張に異議を唱える能力を民主化すると、その異議自体を評価するという新たな問題が生まれるからである。

オープンサイエンスは、科学コミュニケーションの中心的なボトルネックとしてアクセスに取り組む（UNESCO, 2021）。生成AIは今、第二のボトルネック、すなわち科学的資料へのアクセスとそれを吟味する能力とを隔てる技術的・認知的閾値を弱めうる。そうであれば、AIが科学にもたらす帰結は、論文を書ける人がどれだけ増えたかを数えるだけでは理解できない。それを点検、再現し、疑問を投げかけ、異議を唱えられる人がどれだけ増えるかも含めなければならない。

3. 生成AIと科学的生産

生成AIが科学に及ぼした最も目に見える影響は、研究の生産過程への統合である。大規模言語モデルと関連システムは現在、文献統合、仮説生成、コーディング、実験設計、データ分析、解釈、原稿作成、科学コミュニケーションなど、研究サイクルの広い範囲にわたる活動を支えている。科学的発見のためのAIに関する近年のレビューは、特定の課題を支援する段階から、科学的過程の複数の段階を遂行できる、より統合されたシステムへの進展を記述している（Reddy & Shojaee, 2025）。他の研究も同様に、大規模言語モデル（LLM）を、執筆や情報検索に限られる道具ではなく、科学的方法の生産性を高めうる構成要素として捉えている（Zhang et al., 2025）。

この拡大は、科学の自動化と自律性に関する、より広い研究文献を促してきた。現代のシステムは補助者としてだけでなく、分析者として、さらに野心的な構成では半自律的・自律的な科学エージェントとして提案されている。Reddy and Shojaee (2025) は科学的推論、シミュレーション、実験の進歩を述べる一方、統合された自律的研究への障害が残ることを指摘する。Zhang et al. (2025) も、現在の支援と、根本的な発見に貢献するシステムの展望とを区別する。これらの展開を踏まえても、自律的な科学判断を現在のシステムへどこまで委ねられるかは未決のままである。

帰結は自動化に限られない。生成AIは、歴史的に相当の訓練や協働を要した課題の遂行に必要な専門労働を減らしうる。プログラミング支援によって、ソフトウェアの専門知識が限られた研究者も分析の作業工程を構築・変更できるかもしれない。言語モデルは文献統合を加速し、分野間の用語の翻訳を助ける。仮説生成ツールは大量の候補アイデアを生み出せる一方、AI支援分析はデータの探索や代替仕様の検証に必要な時間を減らしうる。これらの能力は、AIを知識労働の民主化技術として論じる議論の拡大に寄与してきた。例えばDaepp et al. (2026) は、生成AIが複雑度の高い知識労働の課題へのアクセスを広げるか、そして誰がその能力から利益を得る立場にあるかを明示的に検討している。

それでも民主化という主張には限定が必要である。作業コストの低下は、データ、基盤、計算資源、分野の知識、指導、研究資金、制度的ネットワークへのアクセスの不平等をなくさない。Daepp et al.

(2026) は、AI支援による知識労働の便益に社会的・地理的格差を見いだしている。また、技術的支援は 高品質な科学的推論を保証しない。評価、再現可能性、人間による検証は、AI支援による発見に引き続き制約を課す（Reddy & Shojaee, 2025）。

ただし本稿の議論にとって、AIが自律的な科学者になるかを決着させる必要はない。より控えめな命題で十分である。生成AIは科学的生産の複数の構成要素に関わる認知的・技術的・時間的コストを下げうる。それらの構成要素を最終的に人間の研究者が統合するにせよ、自律システムが統合するにせよ、多くの研究課題を遂行するための閾値は変わりつつある。この観察が、包括的な二重の民主化のテーゼであるP1を導く。

P1 ― 二重の民主化。 生成AIは、科学的主張の生産と科学的精査の双方について、能力上の障壁を下げうる。

この命題の生産側は、すでに文献において比較的よく展開されている。精査側はそうではない。既存研究の多くは、AIが仮説を生成できるか、分析を行えるか、論文を書けるか、査読者を支援できるかを問う。これらは重要な問いだが、同じ技術的能力から生じる、構造上異なる帰結を脇に置いている。研究者の分析構築を助けるまさにその道具が、別の人による分析の再構成も助けられる。コードをその作成者へ説明するモデルは、同じコードを批判者へも説明できる。原稿作成のための文献統合を加速するシステムは、出版後の引用確認も加速できる。分析仕様を提案するツールは、公表結果に異議を唱えるための代替仕様も提案できる。したがって同じ作業コストの低下は、二つの方向に働きうる。**生産：** 問いやデータセットから科学的主張へ。**精査：** 科学的主張から、その仮定、証拠、コード、引用、代替説明へ遡る。

AIと科学に関する既存の議論は、主として第一の方向に集中してきた。本稿の残りは第二の方向に集中する。

4. 生成AIと科学的精査

4.1. 制度化された精査としての正式な査読

科学的批判は従来、最も目に見える形では査読を通じて制度化されてきた。出版前に編集者は、原稿の妥当性、意義、方法論、解釈、提示方法を評価することが期待される査読者を選ぶ。したがって査読者は二つの立場を同時に占める。科学的主張の精査を期待されるという評価上の立場と、雑誌からその精査を正式に認可されているという制度上の立場である（Ross-Hellauer, 2017）。

これら二つの機能は分析上区別できる。査読者は、招待を受けただけで原稿を評価する技術的能力を得るわけではない。むしろ雑誌は、分野の専門性、発表歴、専門家としての評判、その他のシグナルを用い、その能力を持つと推定される人を特定する。したがって査読は、選定の仕組みを通じて**評価能力**と**制度的承認**を結び付けている。

この仕組みには重要な利点がある。関連する専門性を持つと期待される人々に評価を集中し、編集者が事務的に扱える範囲の判断を提供する。しかし同時に、出版前評価へ正式に参加する人数と範囲を制約する。査読者不足、作業負担、分野の細分化、利益相反、不完全な査読者選定は、長く認識されてきた制度上の限界である（Nabavi et al., 2026）。少数の査読者では、現代の研究論文に含まれるすべての仮定、引用、分析上の判断、データセット、コードの各行を点検することはできない。したがって、出

版の決定を科学的評価の完了と理解することはできない。それは、より長い精査過程の中にある一つの制度的な確認地点を表す。

4.2. AI支援査読

生成AIは、この確立された評価システムにますます入り込んでいる。査読者は、原稿の要約、潜在的な弱点の特定、コメントの整理、表現の改善、統計手法の点検、査読報告の一部の生成に言語モデルを利用しうる。編集者や出版社も、AI支援によるトリアージ、査読者選定、技術的確認、自動評価を試みている。

これらの実践に関する文献は急速に拡大している。Nabavi et al. (2026) は、2025年8月までの学術査読におけるAIに関する189件の資料を特定し、編集段階のトリアージ、査読者支援、自律的査読生成を扱った。その統合的検討は、効率上の便益とともに、機密性、バイアス、不十分な領域固有の推論に関わるリスクを示している。

これらの懸念は、重要なガバナンス研究を生み出してきた。このレビューはまた、出版社の方針の不統一と、自律的評価の限界を記述している（Nabavi et al., 2026）。したがって、流暢で査読らしい文章それ自体は、その背後にある判断の質を確立しない。しかし本稿の議論にとって、AI支援査読は、ある重要な意味で制度的には従来の枠組みを維持する。それは査読者が使える道具を変えるが、誰を査読者とみなすかを必ずしも変えない。基本構造は次のままである。

雑誌 → 承認された査読者 → AI支援 → 評価的判断。

AIが過程に入る前に、査読者はすでに制度的閾値を越えている。AI支援査読は、確立された評価上の役割の内部でのAI利用に関わる。AI支援による分散型精査は、その役割を超えた評価能力の拡大に関わる。

4.3. 出版後の精査

科学的評価は出版後にも行われる。編集者への書簡、論評、追試、訂正、ソーシャルメディアでの議論、機関による調査、専用の出版後査読システム、PubPeerのようなプラットフォームは、従来の査読が終わった後にも、公表された主張に異議を唱える経路を提供する（Hardwicke et al., 2022; PubPeer, n.d.）。

出版後査読が重要なのは、まさに出版前査読が必然的に不完全だからである。他の研究者が追試を試み、基礎となる画像やデータを点検し、異なる状況に手法を適用し、あるいは論文をより広い文献と比較して初めて、誤りが見えることがある。オンラインのプラットフォームはさらに、当初の出版雑誌から独立して、より迅速に批判を行うことを可能にする。PubPeerは、分散した出版後精査の力と制度的帰結の双方を示す。Caron et al. (2025) は、コメントが研究不正に関する手続きにつながってきたこと、そしてその量と速度が制度側の対応に負担を与えうることを述べている。したがって出版後査読は、雑誌が当初選んだ小集団を超えて科学的精査が分散しうることを、すでに実証している。

それでも分布は不均等なままである。PubPeerや同様のシステムへの参加は、すべての参加者が、分析の再構成、統計モデルの検討、コードの点検、引用の追跡、技術的に意味のある不整合の特定を等しく行えることを意味しない。開かれた参加は、批判者になりうる人々を増やすが、第2節で述べた能力上の障壁を除去しない。生成AIは、この関係を変えうる。

4.4. AI支援による分散型精査

私は**AI支援による分散型精査**という語を、正式に割り当てられた評価上の役割の外で科学的主張を精査する人間の主体が直面する技術的・認知的コストの一部を、生成AIが減らす科学的評価を指して用いる。

その定義上の特徴は、AI利用自体ではない。AI支援による原稿査読は、独自の文献群を形成するほどにはすでに広がっている。また、公衆の参加も定義上の特徴ではない。それはオープン査読、市民科学、書簡、ブログ、追試の共同体、出版後査読を通じて、生成AI以前から存在していた。固有の特徴は、次の三条件の相互作用である。

1. **科学的主張を外部から検討できること。**

2. **評価者が正式な査読上の役割への任命に依存しないこと。**

3. **以前ならより多くの技術的専門性、時間、専門家の支援を要した作業を、AIが媒介すること。**

これらの条件の下では、評価過程の一部を遂行する実際のコストが下がるため、精査はより分散しうる。

この機構は、より限定された技術的応用にすでに見られる。Alvarez et al. (2024) は、LLMと引用文から得た例を用いた科学的主張の検証を実証し、人手で作成した主張と証拠の組により訓練した従来のファインチューニング済みシステムに近い性能を得た。Edelman and Skolnick (2025) は、科学的主張を大量に検証するオープンソースシステムValsciを開発した。その評価では、基盤モデルの出力に比べ誤分類を減らし、捏造引用を回避し、学部生研究者1名による手作業の検証よりも大幅に速く動作した。これらのシステムは通常、研究者、出版社、査読者、研究公正の専門職のための道具として論じられる。しかし、その基礎にある機能は、本来的にそれらの制度的役割に結び付いているわけではない。引用検証ツールは、利用者が査読者に任命されているかを知る必要がない。

コードを解釈するツールは、分析過程を説明する前に学術機関への所属を要求しない。言語モデルは、主張とその引用先を比較する前に編集者の認可を要求しない。**評価の機能**と**評価上の役割**のこの分離が、AI支援による分散型精査の中心にある。

引用確認を考えよう。論文には数十、数百の参考文献が含まれうる。各文献が、その文献に帰属された命題を実際に支持するかを判断することは、概念的には単純でも多くの労力を要する。抽出、検索、要約、主張と証拠の比較を自動化すれば、そのコストを大幅に減らせる。従来なら査読者や強い動機を持つ批判者が選択的にしか行わなかった課題を、より大規模に遂行しやすくなる。

同じ論理は分析コードにも当てはまる。公開コードは従来、関連するプログラミング言語と計算環境をすでに理解できる読者に恩恵を与えてきた。今では生成モデルが、関数を一行ずつ説明し、言語間で翻訳し、依存関係を特定し、欠けた段階を再構成し、変更した仕様を生成できる。これは正しい解釈を保証しないが、分析の吟味を始めるために事前に必要となるプログラミングの専門知識を減らす。

統計的解釈も一例である。論文の結論が不慣れな推定量に依存すると気づいた読者は、以前なら先へ進む前に統計家を必要としたかもしれない。AIシステムは仮定の初歩的説明を提供し、妥当性を脅かす一般的な問題を特定し、診断のための問いを提案し、感度分析のコードを生成できる。読者はなお専門家の確認を必要としうるが、専門家の支援が必要になる時点を、より後の段階へ移せる。

文献比較も同様に変わりうる。個人の批判者が、論文が数十の先行研究を正確に表現しているかを判断するには、従来かなりのコストがかかった。AI支援による検索、抽出、要約、比較は、その作業負担を

圧縮できる。したがって、著者が文献をより速く統合できるようにする同じ基盤が、批判者によるその統合の吟味も速くできる。

これらの例は、科学のためのAIをめぐる現在の議論の多くにある、重要な非対称性を明らかにする。科学用AIツールは通常、その遂行課題、すなわち文献レビュー、コーディング、統計分析、画像点検、引用確認、仮説生成によって分類される。しかし、その多くは**方向に関して中立的**である。コード生成は生産を支援し、コードの再構成は精査を支援しうる。文献統合は序論を支え、文献比較はその序論の正確さを確かめうる。統計モデリングは結果を生み、モデルの説明と代替仕様はその結果を吟味しうる。主張の検証は投稿前の著者を支え、同じ能力が出版後の批判者を支えうる。その技術的能力は、科学的主張のどちらか一方の側に本来的に属するわけではない。これによって、分散型精査と二重の民主化モデルとの関係が定まる。生成AIは、研究者のための新しい道具を作るだけではない。既存の研究用の道具の一部を、それにアクセスできる誰もが、より低い実効的な技能・時間の閾値で利用できるようにする。

対象となる人々は、専門家の代わりになろうとする人だけで構成される必要はない。異なる主体が異なる形の精査に貢献しうる。

臨床家は、公表されたモデルの仮定が現場の実態と衝突すると気づくかもしれない。患者の権利擁護者は、報告されたアウトカムと、基礎となる測定が実際に捉えているものとの不一致を特定するかもしれない。ジャーナリストは、中心的主張を引用の連鎖に沿って追跡するかもしれない。別分野の研究者は、論文の直接の分野の外では異例な方法論的仮定に気づくかもしれない。独立研究者は、公開分析を再現したり、複数の資料にわたって公表された主張を比較したりするかもしれない。技術に通じた一般市民は、専門家による検討に値する不整合を発見するかもしれない。AIは、こうした観察を直感にとどめず、構造化され、証拠と結び付いた異議へ進められる主体を増やしうる。この移行は、精査の段階として表せる。

観察「この結果は意外に思える。」→ **吟味**「どのような仮定、データ、分析がこの結果を生んだのか。」→ **再構成**「証拠から結論へ至る、報告された道筋を再現できるか。」→ **異議提起**「代替仕様、比較、出典確認を経ても、この主張は成立するか。」→ **検討に開かれた批判**「他者が独立に点検できる証拠と手続きを伴って、この異議を述べられるか。」

生成AIは、この段階のいくつかの移行でコストを減らせる。そのすべてで専門性の必要をなくすとは考えにくい。その意義はむしろ、より多くの主体が、以前より先の段階へ進めるようになることにある。これによりP3は、P1の精査側の構成要素の系として特定される。

P3 — 分散型精査。 生成AIは、評価課題の技術的・認知的・時間的コストを下げることによって、科学的精査の実質的な構成要素を遂行できる人々を、正式な査読上の役割や従来の専門家共同体の外へ拡大しうる。

この命題は意図的に、同等の専門性ではなく**精査の構成要素**を扱う。この区別は重要なカテゴリー上の誤りを防ぐ。感度分析を実行できることは、そのすべての含意を解釈する能力の証明にはならない。異常な引用を見つけても、論文の結論が誤りだとは確定しない。もっともらしい方法論的異議を作っても、その異議の妥当性は確立しない。したがって、AI支援による分散型精査は、批判の科学的妥当性を確立する能力よりも、**批判の候補**を生み出す能力を容易に再分配する。批判のコストが下がるにつれ、この違いはますます重要になる。

方法論的異議の作成が、かつて数時間の専門的作業を要し、今では数分のAI支援による調査で済むなら、作成できる異議の数は大幅に増えるかもしれない。その一部は真の誤りを特定する。他は誤解、不適切なモデルの仮定、捏造証拠、AIの幻覚を反映する。その結果、科学が直面する問題は、批判の不足から、信頼に値する批判を雑音から区別する仕組みの不足へ変わりうる。

これは全く前例のないことではない。PubPeerはすでに、科学に対する公的批判の量と速度の増加が、制度的負担をどう生みうるかを示している。生成AIは、さらに別の拡大機構を加えうる。コメントの機会を持つ人が増えるだけでなく、技術的に精緻な異議を作成する道具を備えた人が増えるのである。したがって重要な帰結は、科学的権威が不要になることではない。**科学が従来評価能力を見いだすと想定してきた制度的な場所の外にも、その能力がますます存在しうること**である。そこから次の問題が生じる。

正式に任命された査読者が異議を提起するとき、雑誌はすでに、その異議を受け取り、重み付けする仕組みを提供している。編集者が提起する場合も制度的権威は同様に明確である。匿名のPubPeer投稿者、独立研究者、ジャーナリスト、臨床家、患者の権利擁護者、技術的能力を持つ部外者が、公表結果に対して再現可能な異議を提示する場合、その批判をどう扱うかを決める仕組みは、はるかに未確立である。したがって生成AIは、一つの問題の一部を解決すると同時に、別の問題を顕在化させうる。AIは、次の問いを問うコストを下げられる。

「この主張に異議を唱えられるか。」

しかし、次の問いには答えない。

「その異議は、いつ科学的に信頼に値するものとして承認されるべきか。」

次節では、分散した評価能力から、この科学的承認の問題へ議論を移す。

5. 分散した評価能力と科学的承認

科学的精査には、異議を形成する能力以上のものが必要である。異議が認識的な重みを獲得するための、何らかの仕組みも必要になる。評価能力が制度的役割から分離されると、この問題がいっそう明確になる。

科学の諸制度は歴史的に、この承認問題を解くために地位のシグナルへ依存してきた。学術資格、発表実績、分野の専門性、所属機関、専門職への任命、編集者による選定、同業者の承認は、いずれも不完全だが社会的に利用可能な専門性の指標となる。雑誌の編集者は、すべての査読候補者の能力全体を独力で把握し直すことはできない。臨床家は、臨床指針の根拠となるすべての科学者の専門性を自ら検証できない。一般市民は、専門家のあらゆる主張を支える研究全体を再現できない。そのため科学共同体は、認識的に信頼するに値すると考えられる判断を行う人を特定する社会的仕組みに依存する（Rolin, 2025）。

これらの仕組みは恣意的ではない。専門性は、一部には分野での継続的経験、暗黙知、方法論的能力、専門家共同体への参加を通じて成り立つ。認識的信頼に関する近年の研究が強調するのは、まさにこの点である。例えばMirkin and Sojka (2026) は、資格、実績、議論における遂行能力といった、よく使われる目印では、非専門家がある人の専門性の有無を直接判定することはできないと論じる。それらはむしろ、経験と認識的な信頼性の間接的指標として働く。Keren (2025) も専門性と認識的権威を区別し、特定の問いに関する権威を、専門性だけから推論することはできないと強調する。したがって、専

門性の制度的承認は重要な認識的機能を果たす。誰の判断に最初に注意を向けるべきかを決めるコストを減らすのである。

問題は、承認と能力が完全には一致しないことである。承認された専門家も誤る。専門家は、その経験が適切に関係する領域の外で発言しうる。有資格の研究者も、専門外の手法を誤解し、誤りを見逃し、利益相反を開示せず、信頼できない証言を行うことがある。近年の社会認識論研究は、真の専門性を欠く擬似専門家と、正当な資格を持ちながら体系的に信頼できない証言をする専門家とを区別している（Croce & Marsili, 2025）。そのため資格は有用なシグナルではあるが、特定の科学的主張や批判の妥当性を保証しない。逆の状況もありうる。科学の諸制度が通常、権威ある批判を期待する役割を占めていなくても、妥当で再現可能な問題を特定する人はいるかもしれない。

この可能性は生成AI以前から存在する。公開・匿名の出版後コメントの投稿者が提起した懸念が、後に正式な研究公正の手続きに入った事例がある（Caron et al., 2025）。PubPeerは公開・匿名の出版後コメントを認めることで、この可能性の一部を制度化している。ただしPubPeer自体はコメントの科学的正しさを認証せず、読者が評価することを期待する（PubPeer, n.d.）。したがってこのプラットフォームは、批判する機会と、その批判の制度的認証とを分けている。

分析の再構成、引用の追跡、コードの点検、文献比較、感度の確認を行う技術的コストが下がれば、承認された評価上の役割の外にいるより多くの人が、注意を向けるに値するほど技術的に実質のある異議を作成できるかもしれない。そこから生じる課題は、部外者を信頼すべきかどうかだけではない。従来の代理指標である

承認された地位 → 推定される評価能力

という関係と、特定の評価課題を遂行する実践的能力との結び付きが弱まるとき、科学の諸制度は批判をどう評価すべきか、ということである。私はこれを**能力と承認の不一致**と呼ぶ。すなわち、科学的評価の実質的な構成要素を遂行する能力が、正当な評価者を特定するために通常用いられる制度的承認よりも広く分布している状態である。この概念は、能力が専門性と同等であることを意味しない。また、制度的承認が不要になったことも意味しない。両者の関係の変化を特定するものである。

生成AIが広く普及する前は、一部の批判は十分に高コストであったため、それを作成する能力は専門的訓練や制度内の位置としばしば強く相関していた。計算分析の再現には、相当なプログラミングの専門知識が必要かもしれなかった。長い引用の連鎖の確認には、手作業による検索に数日を要したかもしれなかった。代替的な統計仕様の探索には、統計家へのアクセスが必要かもしれなかった。これらのコストは不完全ながらも、技術的に精緻な批判を通常のコメントから隔てる非公式な障壁として機能していた。そのコストが下がるにつれ、技術的に精緻な批判が存在するというだけで伝わるシグナルは弱くなる。

詳細な方法論的異議は、経験豊富な専門家が作成したものかもしれない。AI支援を慎重に用いる隣接分野の研究者によるものかもしれない。AIを責任を持って使い、一段階ずつ検証する初心者によるものかもしれない。あるいは、表面的なプロンプトから生成された、流暢だが根本的に誤った異議かもしれない。批判の表面的な形が、その背後の能力について明らかにすることは、ますます少なくなる。このため、純粋な資格主義も、内容の純粋な平等主義も、十分な対応にはならない。純粋な資格主義は、主に批判者の地位に従って信頼性を割り当てる。資格と分野での経験は専門性に関する情報を含むため、これは今も効率的で、しばしば合理的である。しかし偽陰性を生む。技術的に妥当な批判が、その出所に従来型の権威がないために軽視されうる。

内容の純粋な平等主義は、その実質が独立に評価されるまで、すべての批判を一応同等に扱う。これは地位による排除を避けるが、深刻な規模拡大の問題を生む。AIがもっともらしい技術的異議を生成するコストを下げれば、科学の諸制度がすべての異議を専門家による全面的な裁定に付すことは、現実的にはできない。したがってガバナンス上の問題は、**批判が低コストになった条件下での信頼性の配分**である。一つの対応は、批判のうち、独立に点検できる性質へより大きな重みを置くことである。これらの性質は、その特定の批判の信頼性について証拠を提供する。このアプローチは、次だけを問うのではない。

誰が主張しているのか。

次のことも問う。

- 批判は特定可能な証拠に結び付いているか。

- 引用された出典を取得して確認できるか。

- 分析の手順は開示されているか。

- 別の評価者が報告された不一致を再現できるか。

- 仮定は明示されているか。

- 明白な誤解を訂正した後にも批判は成立するか。

- 不確実性は認められているか。

- 批判者は、どのような証拠がその異議を反証するかを特定できるか。

- 批判は反対証拠に応答しているか。

これらの基準は、評価の一部を個人の地位から、実証可能な認識的遂行へ移す。

それらは専門性に取って代わらない。実際、多くの問いは専門家の暗黙知なしには解決できない。再現可能な計算であっても、実質的な問いには無関係かもしれない。統計的に正しい代替仕様も、科学的には不適切かもしれない。分野での経験によって、汎用AIシステムや外部の批判者には見えない制約が明らかになることがある。したがって、そのような指標は、従来の専門性の指標に代わるのではなく、それと並んで働く。結果として得られるモデルは、置換的ではなく加算的である。

地位に基づく信頼性 + 独立に点検可能な証拠 → 科学的異議の評価

地位は、見込まれる能力について事前情報を与える。遂行実績は、特定の批判の質について証拠を与える。両者の相対的な重みは、事例ごとに異なりうる。

結果が「おかしく見える」と述べるだけの匿名の主張は、科学的対応の根拠をほとんど与えない。重複画像を特定し、図の正確な位置を示し、比較手続きを記録し、独立に点検可能な証拠を提供する匿名投稿者は、異なる事例である。逆に、高い資格を持つ科学者であっても、証拠なしに関連分野の外で批判するなら、制度的地位だけで無制限の認識的重みを得るべきではない。この区別は、生成AIが遂行実績のシグナルを同時に強めも弱めもするため、AI支援による精査にとりわけ関係する。生成AIは、実行可能なコード、引用表、代替モデル仕様、監査記録、構造化された比較、再現可能な作業工程という、透明な裏付け資料の作成を容易にすることで、そのシグナルを強める。しかしAIは、厳密さの外観も作り出せるため、そのシグナルを弱める。本当の理解がなくても、流暢な方法論的表現、多数の参考文献、

コード、統計用語を生成できる。かつて技術的能力の高コストなシグナルとして働いたものは、偽装しやすくなりうる。

その結果、重要なシグナルは再び移る。技術的洗練それ自体から、**検証可能性**へである。批判は、技術的に高度に見えるというだけではなく、その証拠の連鎖を独立に点検できるからこそ、より大きな重みを与えられるべきである。

この区別は、AI支援による分散型精査が提起する問いへの、一つの答えを与える。科学の諸制度は、批判者が発言するに足る社会的地位を持つかを事前に決定する必要はない。また、すべての批判を等しく扱う必要もない。代わりに、異議を提起する権利と、その異議が制度的な注意を要求するために必要な証拠上の負担とを区別できる。ここから第四の命題が導かれる。

P4 — 能力と承認の不一致。 科学的評価能力の普及は、制度的承認の同程度の普及を必ずしも伴わない。AIが正式な評価上の役割の外での実質的精査のコストを下げるにつれ、科学の諸制度は、専門性の指標と、独立に点検可能な証拠に基づく遂行の指標とを組み合わせて批判を評価する仕組みを、ますます必要とするようになる。

匿名の懸念が正式な過程へ入ることは、承認が可能であることを示すが、不一致が消えたことを証明しない。ここでいう不一致は、外部から提起された懸念が、その証拠上の重みに関する説明責任を伴う決定へ至るまでの移行に関わる。コメントを受け取ること、調査を開始すること、誤りを確立すること、応答を求めることは、それぞれ別の制度的行為である。Caron et al. (2025) が記録するのは処理上の負担であり、それらの行為を結ぶ閾値が一般的に十分であることではない。残る問いは、どの具体的異議が調査に値するのか、誰がそれを評価して応答しなければならないのか、そして結果を批判者の地位ではなく証拠にどう帰属させるかである。既存の経路は個別事例を解決できても、規模が拡大したときのこれらの問いを未解決のまま残しうる。したがってP4は、承認の仕組みが一切存在しないことではなく、外部の能力が適切に制度へ受け入れられる条件を扱う。

科学的異議は、制度の外から来たという理由で妥当にも不当にもならない。その制度的な重みは、その異議自体が精査に耐えるかに依存しなければならない。したがって能力と承認の不一致は、品質管理の問いへつながる。制度は、信頼に値する外部の精査を、技術的に高度に見えるだけの批判から、どう区別できるか。第6節では、批判が低コストになるにつれこの区別がなぜ難しくなるのか、そして何を必要とするかを検討する。

6. リスクとガバナンス——批判が低コストになるとき

生成AIは、批判者による誤りの特定と、制度側による異議の確認の双方を支援できる。第4.4節で述べた方向に関する中立性は、両側のこれらの技術的操作に当てはまる。しかしそれは、批判の候補を生成する総コストと、それについて正当化された制度的決定へ至る総コストとが、同程度に低下することを意味しない。

その違いは、各到達点が必要とする作業から生じる。異議の候補の作成は、もっともらしい主張、裏付けとなるコード、あるいは代替分析案を出したところで終わりうる。その異議が訂正その他の制度的措置に値するかを決めるには、さらに関連性を確認し、領域固有の曖昧さを解消し、著者の応答を考慮し、決定の責任の所在を定める必要がある。AIはこれらを支援できるが、出力を生むこと自体では、その認識上・手続き上の要件を満たしたことにはならない。文脈に即した判断を機械的な確認へ還元できない場合には、暗黙知が依然として必要である。制度的措置が問題となる場合には、応答の機会と、説明責

任を伴う決定が依然として必要である。既存の出版倫理の手続きも、懸念を提起することと、その帰結を決めることをすでに区別している（Committee on Publication Ethics [COPE], 2008; Wager & Kleinert, 2021）。

したがって、これらの残存する裁定作業が相当の比重を占める場合、**批判の生成コストは、その検証と裁定の総コストよりも速く低下しうる**。これは条件付きの非対称性であり、AIが生成だけに便益を与えるという主張ではない。容易に確認できる数値上の不一致なら、検証も同じ程度、あるいはそれ以上に低コストになりうる。規模拡大の問題は、増え続ける異議の候補に対し、同じ速さでは加速しない文脈的評価と、説明責任を伴う決定が必要になるときに生じる。

6.1. 見せかけの精査

私は**見せかけの精査**という語を、科学的評価の外形を備えていながら、十分に信頼できる証拠上・分析上の基礎を欠く批判を指して用いる。見せかけの精査は、意図的な欺瞞である必要はない。純粋な誤解からも生じうる。利用者が論文の弱点を特定するよう言語モデルに求め、実際には適切に使用されている推定量に対する、技術的にはもっともらしい異議を受け取るかもしれない。モデルが、研究設計と両立しない仮定に基づく感度分析を推奨するかもしれない。実際には異なる推定対象や母集団を扱う二つの論文の間に、矛盾があると指摘するかもしれない。正当な方法論的論争を、一方の立場が決定的に確立されたかのように記述するかもしれない。あるいは、公表された出典が実際には何も支持していない異議を裏付ける引用を、作り出すかもしれない。生成AIの修辞的な質の高さは、これらの失敗の帰結をとりわけ重大にする。弱い批判が弱く見えるとは限らない。

エビデンス統合における生成AIの研究は、有用な警告を与える。19研究のシステマティックレビューは、複数の研究課題にわたって相当な誤りを見いだした。生成AIシステムは文献検索において関連研究の68–96%を見落とした。誤り率の中央値は、誤った除外判断で28%、データ抽出で14%、バイアスリスク評価で27%であった。著者らは、現時点の証拠は、人間の関与や監督なしにエビデンス統合へ生成AIを用いることを支持しないと結論した（Clark et al., 2025）。これらの知見は科学的批判を直接扱うものではなく、エビデンス統合に関するものだが、その含意は移しうる。科学論文を批判するのに必要な操作の多く、すなわち文献検索、証拠比較、データ抽出、方法論的分類は、それ自体がAIの誤りに脆弱である。したがってAI支援による異議の存在は、精査が行われたことの証拠にはなる。

しかし、その精査が適切に遂行されたことの証拠にはならない。

6.2. 自動化バイアスと認識的依存

第二のリスクは、AIが誤りを生むことではなく、人間の評価者がそれを見抜けないことである。自動化バイアスは広く、自動化された推奨や出力に過度に依存する傾向を指す。人間とAIの協働に関する研究は、過信、注意の制約、AIリテラシー、専門性、課題の検証に必要な負担、説明の複雑さを、その依存に影響する要因として特定している（Romeo & Conti, 2026）。生成AIを用いた実験研究も、誤ったAI支援を与えられた利用者の成績が、支援なしの利用者を下回りうることを示す。警告による介入は、この影響を軽減できても除去はできない（Wingerter et al., 2025）。これが重要なのは、AI支援による分散型精査が、従来の専門家によるソフトウェア利用とは構造的に異なる、人間とAIの関係に依存するためである。従来の統計ソフトウェアは通常、利用者が明示的に指定した手続きを実行する。それに対し生成モデルは、手続きを提案し、それが適切である理由を説明し、コードを生成し、結果を解釈し、批判を形成できる。

そのため同じシステムが、分析と、その分析を信頼することの正当化の双方を提供しうる。ここに**認識的依存**の可能性が生じる。批判者は科学的主張を独立に評価しているように見えても、異議の選択、その背後の技術的推論、証拠の解釈を、同じAIシステムに依存しているのである。危険なのは、単にモデルが誤りうることではない。どこが誤っているかを見抜くために必要な、独立した能力を利用者が欠いているかもしれないことである。したがって、結果として生じる評価能力の拡大は浅いものにとどまりうる。人は、分析操作を**実行する**能力を得ても、その操作がいつ適切かを独立に判断する能力までは得ないかもしれない。この区別は、本稿の議論にとりわけ重要である。能力の閾値を下げることは、その閾値より上にある専門性の勾配をなくすことと同じではない。

6.3. 幻覚に基づく批判

引用の幻覚は、この問題を最も明瞭な形で示す。生成システムは、正当に見えても実在する出版物に対応しない参考文献を作り、著者名や題名を損ない、あるいは実在する参考文献を、その文献が支持しない主張に結び付けうる。幻覚による参考文献は、もはやLLMの振る舞いを示す実験室内の実演に限られない。Sakai et al. (2026) は、2024–2025年のAssociation for Computational Linguistics、その北米支部、およびEmpirical Methods in Natural Language Processingの各会議録において、少なくとも一つの幻覚による引用を含む論文を約300件特定した。同じ機構は批判にも働きうる。批判者は、権威あるように見えて実在しない反証文献を用いて、公表された主張に異議を唱えるかもしれない。

モデルは、実在する論文が述べていない方法論的立場を、その論文に帰属させるかもしれない。そのため文献比較は、多数の出典に裏付けられているように見えて、捏造された証拠や、意味の上で対応しない証拠に依拠している場合がある。これは特有の再帰的問題を生む。AIは、科学論文中の幻覚による参考文献を検出するために用いられるかもしれない。同時にAIは、その論文に対する批判の中でも、幻覚による参考文献を生成しうる。こうして科学的評価には、もう一つの層が加わる。

主張 → 主張への批判 → 批判の検証

AIによるコスト低下は各層で生じうるが、誤りの生成も各層で生じうる。

6.4. 批判の氾濫

規模が掛け合わされると、問題はさらに深刻になる。従来の批判は歴史的に高コストであった。詳細な方法論的応答を書くこと、分析を再現すること、数十の参考文献を追跡することには、相当の時間が必要である。これらのコストは参加を制限するが、量も制約する。生成AIは、その量の制約を弱める。意欲のある一人の利用者が、論文に対する詳細な異議を数十件生成できるかもしれない。組織化された集団は、数百の出版物を精査できるかもしれない。自動エージェントは、査読に似た評価を継続的に生み出せる可能性がある。その増加は必ずしも悪意によるものではない。多くは、科学を改善しようとする正当な試みを反映するかもしれない。しかし、科学の諸制度が持つ検証能力は有限である。編集者の時間は限られている。著者が応答する能力は限られている。研究公正部門の調査資源は限られている。統計査読者と各領域の専門家は、依然として希少である。

批判を生産する限界費用がゼロに近づく一方、その裁定の限界費用が高いままなら、制度は**批判の氾濫という問題**に直面しうる。

出版後査読には、すでにその初期的な類例がある。PubPeerは公表研究における重大な問題の発見を可能にしてきたが、公正性への懸念の可能性を提起するコメントの件数、頻度、速度は、どの申し立てが

調査に値するかを判断しなければならない機関に、相当な事務負担も生み出してきた（Caron et al., 2025）。生成AIは、技術的に精緻な異議を作るコストを、その裁定コストほどには下げずに低下させる ことで、この既存の非対称性を増幅しうる。これはスパムとの類比ではない。科学的批判を無価値と推定すべきではないからである。重要なのは構造上の点である。提出コストが大幅に下がっても裁定コストが下がらないとき、選別は中心的なガバナンス問題になる。したがって、問題は批判の多さそのものではない。**批判の生成能力と批判の検証能力との不一致**である。

6.5. 来歴と監査可能性

一つの対応は、AI支援による批判の監査可能性を高めることである。生成AIに依存する批判は、理想的には、批判者にもAIシステムにも信頼を置かずに他者が結果を点検できるよう、証拠へ至る経路を十分に示すべきである。ここから、いくつかの望ましい性質が示唆される。

- 特定可能な出典文書。

- 取得可能な引用箇所や主張の所在。

- 開示された分析コード。

- 再現可能な変換。

- 明示されたモデルの仮定。

- 該当する場合の、版が管理されたデータとソフトウェア。

- 検索・取得した証拠と、AIが生成した解釈との明確な分離。

- 関連する場合の、実質的なAI支援の開示。

- AIが生成した誤りが発見された場合の訂正の仕組み。

これらの性質は**来歴（provenance）**という観点から理解できる。発見可能性、アクセス可能性、相互運用可能性、再利用可能性からなるFAIR原則は、再利用を支える要件に来歴を含む。World Wide Web Consortium（W3C）のPROV枠組みは、デジタル対象の生成に関わる実体、活動、主体を記述する（Wilkinson et al., 2016; Groth & Moreau, 2013）。中心となる問いは、AIを使用したかどうかだけではない。別の評価者が、その批判の由来、裏付けとなる証拠、行われた変換、生成的推論に依存する部分を特定できるかどうかである。すべてのプロンプトややり取りの完全開示を求めることは、実際的でなく、認識上も有益でない場合があるため、この点はとりわけ重要である。長い会話記録があっても、批判が再現可能になるとは限らない。

より重要なのは、実質的な証拠の連鎖の来歴である。異議が引用の不一致に関するものなら、元の主張と引用先を特定できるべきである。再分析に関するものなら、データ、コード、分析仕様を点検できるべきである。文献比較に関するものなら、採用過程と基礎となる研究を追跡できるべきである。したがってガバナンスの目標は、**見せるための透明性ではなく、監査可能性**であるべきである。

6.6. 技術的洗練から検証可能性へ

生成AIは技術的に洗練された表現の生産を容易にし、認識的な質のシグナルとしての価値を弱める。複雑な統計用語、多数の引用、整ったコードが重みを持つのは、その証拠上の基礎が独立した確認に耐える限りにおいてである。したがって検証可能性、来歴、再現可能性は、技術的な外観だけよりも多くの情報を与えるようになる。これが第5節の信頼性に関する含意である。表面的な欠陥を減らしても、説得力のある提示に隠れた実質的な誤りは残り、あるいは増幅しうる。

6.7. 分散型精査のガバナンス原則

以上のリスクから、以下でP5として述べる要件が導かれる。異議提起へのアクセスは開かれたままであるべきだが、制度的な注意は、証拠の具体性、再現可能性、関連性に依存すべきである。AI支援の開示によって異議の正しさは確立しない。技術的確認だけでは文脈上・方法論上の問いが解消しない場合、専門家による裁定が引き続き必要である。これらの要件は、異常の可能性の検出と、誤りや不正の確立とを区別する。第7節では、それらの制度的実装を一つの層状構造として展開する。

6.8. 精査の品質問題

ここから第五の命題が導かれる。

P5 — 精査の品質問題。 生成AIが既存の評価上の役割を超えて科学的批判を生産するコストを下げるにつれ、制度的承認は、AI支援や技術的な外観だけに依存することはできなくなる。したがって、分散型精査の信頼性は、来歴、再現可能性、証拠の具体性、適切な専門家による裁定に、ますます依存する。

この命題は、科学の民主化の中心的な逆説を捉えている。障壁の低下は、より多くの誤りの可能性に気づけるようにするため、有益である。障壁の低下は、存在しない誤りの申し立ても増やせるようにするため、危険である。解決策は、以前の障壁が雑音を除いていたというだけの理由で、その障壁を復活させることではない。また、参加自体が認識上の改善を保証すると想定することでもない。ガバナンス上の問題は、異議を唱える拡大した能力を維持しながら、その異議が科学的な重みを得る仕組みの質を高めることである。

批判は低コストになりつつある。信頼に値する批判は、そうではない。

より正確には、批判を生産できる速度に合わせてそれを検証し、優先順位を付け、裁定する能力が、ますます希少な資源になるのである。

7. 科学の諸制度への含意

二重の民主化モデルの含意は、個々の研究者による生成AIの利用にとどまらない。評価能力がより広く分布するなら、科学の諸制度は、既存の専門家の役割の内部でAI利用を規制するだけでは解けない問題に直面する。制度的な問いは、より広いものとなる。

重要である可能性のある問題を特定する能力が、それを特定する人に従来与えられてきた権威より広く普及するとき、科学はどう応答すべきか。

その答えは、批判の周囲に従来の資格上の障壁を再建することであるべきではない。また、制度がすべての異議を認識的に同等と扱うことであるべきでもない。より妥当な方向は、**異議提起へのアクセス**と**制度的な注意を受ける資格**とを分けるシステムである。科学的異議は、誰もが提起できてよい。しかし

制度的措置の重大性が高まるほど、証拠上の裏付け、再現可能性、検証にも、より高い水準を要求すべきである。この区別は、雑誌、出版後査読、研究機関、オープンサイエンス政策、従来型ではない評価者の役割に含意を持つ。

7.1. 雑誌――査読者の身元から検証可能な異議へ

科学雑誌は歴史的に、主として承認された役割を通じて批判を組織してきた。編集者が査読者を選ぶ。査読者が原稿を評価する。著者が応答する。読者は後に書簡、論評、出版後批判を投稿できる場合があるが、これらの仕組みは通常、出版前査読に対して二次的な位置を占める（Hardwicke et al., 2022）。生成AIは、この構造をますます不十分なものにしうる。査読者に選ばれなかった読者でも、引用の誤りを特定し、分析を再現し、表と基礎データとの不整合を見つけ、方法論的仮定に裏付けがないことを発見するかもしれない。したがって、雑誌にとって重要な問いは、次だけではなくなるべきである。

誰が批判しているのか。

次の問いも、ますます重要になるべきである。

その批判を独立に確認できるか。

これは、査読者の専門性が重要でなくなることを意味しない。多くの争点、とりわけ評価が暗黙知、争われている方法論的基準、実質的解釈に依存する場合には、専門家の判断が引き続き不可欠である。しかし専門性と検証可能性は、異なる機能を果たす。専門性は、制度が難しい主張を評価するのを助ける。検証可能性は、その段階へ進める価値が批判にあるかを制度が判断するのを助ける。そのため雑誌は、正式な査読者と外部の批判者という二分法ではなく、構造化された証拠上のカテゴリーによって出版後の異議を扱うことから、利益を得られるだろう。例えば、数値上の不整合の主張には、該当する表、計算、再現可能な導出を求められる。引用に対する異議には、争点となる記述、引用先、正確な不一致を求められる。再分析には、コード、データの来歴、分析仕様、結果を再現するための十分な情報を求められる。概念的・方法論的異議には、問題とする仮定と関連文献の明示を求められる。

こうした要件は、批判が正しいかを決めるものではない。さらに注意を向けるに値するほど、批判が十分に特定されているかを決めるものである。そこから得られる原則は、次のとおりである。

異議を提起する障壁は低くし、制度的措置をより重大な段階へ進めるための証拠上の閾値は高くする。

7.2. 出版後査読――規模拡大を見据えた設計

出版後査読は、従来の査読上の役割の外から生じる評価を可視化する経路を提供するため、分散型精査の下でとりわけ重要である（PubPeer, n.d.）。そのため、その重要性は増すかもしれない。同様に、そのガバナンス上の負担も増しうる。AIが潜在的問題の特定、形成、提出のコストを下げれば、出版後のシステムは、既存の編集・研究公正の仕組みが実際に裁定できる量を超える批判を受け取るかもしれない。適切な対応は、出版後の批判を抑制することではない。それでは精査の拡大の主要な便益の一つを犠牲にする。代わりに、システムは主張の水準を区別する必要があるかもしれない。有用な構造は、次のものを分ける。

観察 ／ 証拠に結び付いた異常 ／ 実証された誤り ／ 不正の申し立て

これらのカテゴリーは、互いに置き換えられない。説明のつかない不一致は、論文が無効であると示さなくても、注意を向けるに値しうる。再現可能な分析上の誤りは、不正を意味せずとも、結果を損ない

うる。疑わしいパターンは、それ自体で意図を確立しなくても、調査を正当化しうる。生成AIは、基礎となる証拠が十分に確立される前に修辞的に強い申し立てを生み出せるため、この区別をいっそう重要にする。したがって、出版後査読のシステムは、非難よりも具体性を評価すべきである。再現可能な誤りを特定する短い批判は、数十の推測的懸念を含む長大なAI生成査読より価値が高いかもしれない。注意の単位は、批判の量や技術的洗練ではなく、ますます**検証可能な異議**となるべきである。

7.3. 研究機関——調査の前のトリアージ

大学、病院、研究所、資金提供機関、研究公正部門は、関連する課題に直面する。これらの機関は、外部から提出されたすべての批判を、正式な不正申し立てであるかのように調査することはできない。しかし、承認された専門職ネットワークの外から来たというだけで異議を退ければ、本当の問題を見逃すかもしれない。したがって分散型精査は、よりよいトリアージの必要性を生む。可能な制度的順序の一つは、次のとおりである。

異議の受領 → 証拠の十分性の確認 → 実行可能な場合の技術的再現 → 領域専門家による検討 → 正当化される場合に限った研究公正上の調査

この順序は重要である。技術的異常を指摘する行為が、不正行為の告発と混同されることを防ぐ。また、制度的地位が、異常に注意を向けてもらうための唯一の入口になることを防ぐ。このモデルの下では、部外者が不正を確立する必要はない。初期段階の負担は、もっと限定される。特定の科学的懸念を独立に検討できるだけの証拠を提供することである。その後、懸念が技術的検証に耐えるか、追加措置が正当化されるかを決める責任を機関が負う。批判の生成能力が正式な調査能力より速く拡大するにつれ、この分業はいっそう重要になりうる。

7.4. オープンサイエンス

含意はオープンサイエンスにも及ぶ。オープンサイエンスとFAIRは、アクセス可能で再利用可能な研究資料の目標を明示している（UNESCO, 2021; Wilkinson et al., 2016）。出版物、データ、コードへのアクセスを可能にすることは引き続き不可欠だが、アクセスだけでは、より広い主体が科学的主張を実質的に検討できるとは保証されない。したがって、本稿で先に展開したアクセスと能力の区別は、個人の能力だけでなく、科学の基盤の水準でも重要である。

生成AIは、アクセスが与えられた後にも残る技術的・認知的障壁の一部を減らしうる。しかし精査への貢献は、科学的対象が、主張の追跡、計算の再構成、分析の再現、特定可能な証拠への異論の結び付けを可能にする形で利用できるかに依存する。したがって開放性は、アクセスと再利用だけでなく、科学的主張を実際に追跡、吟味、再現し、異議を唱える能力も支えるべきである。

これは透明性、再現可能性、再利用という既存の目標に取って代わるものではない。むしろ分散型精査は、これらの目標に追加的な制度上の意義を与える。それらは、確立された評価上の役割の外から生じた批判が、批判者の地位によって受容・棄却されるのでなく、どれほど容易に独立した検討に付されるかを決めるのである。

7.5. 従来型ではない評価者と科学上の評判

分散型精査は、独立研究者、実務家、患者研究者、技術専門家、その他、従来の評価上の役割を占めずとも関連する経験を持つ人々にとって重要である。AI支援は、彼らが実質的な懸念を点検可能な異議へ変えるのを助けうる。それは同等の領域専門性を与えるわけでも、すべての異議が制度的受容に値すると確立するわけでもない。

承認の問題は、資格の廃止ではなく、その十分性に関わる。所属機関、分野での地位、以前からの専門性は、特に裁定が暗黙知に依存する場合、引き続き情報を与える。しかし、特定の異議が検討に値するかを、それらが事前に決める必要はない。同様に、AIによって容易に作れるというだけで、技術的に洗練された提示が検証の代わりになることはできない。批判者に一般的な専門家としての地位を与えなくても、精査に耐えた批判に対して条件付きで承認を与えることはできる。証拠に結び付いた異議を受け入れ、実証された価値に従って次の段階へ進めるシステムは、専門性の役割を維持しながら、より広い参加の便益を保つ。

7.6. 層状のガバナンス構造

これらの含意を合わせると、AI支援による分散型精査のための層状ガバナンスモデルが支持される。その手続き上の要素には先例がある。公表論文のデータ捏造が疑われる場合のCOPEのフローチャートは、読者の懸念から評価、著者への連絡、応答に応じた追加措置へ進む（COPE, 2008）。その後のCLUE勧告は、公表された記録に対する雑誌の責任と、機関による不正調査とを区別し、不正に限られない懸念も明示的に扱う（Wager & Kleinert, 2021）。以下の層は、研究公正のトリアージを新たに発明したと主張するものではない。その目的は、既存の公正性に関する指針もすでに認める、不正を申し立てない通常の分析上・解釈上の異議を含め、AI支援による異議のより広い受け入れに、こうした手続きを接続することである。貢献は、新しい不正調査手順ではなく、拡大する評価能力と、制度的受容のための証拠上の閾値との関係にある。

第1層：開かれた異議提起

科学的異議は、批判者の制度的地位にかかわらず、比較的容易に提起できるべきである。この段階の目的は、裁定ではなく異常の検出である。

第2層：証拠との結び付け

制度的な注意を求める批判は、関連する主張、証拠、データ、分析、出典を、確認できるだけの精度で特定すべきである。

第3層：再現可能性と技術的検証

批判が計算、引用、データセット、分析結果に関わる場合、実行可能であれば、その異議を独立に再現すべきである。

第4層：領域専門家による裁定

機械的には解消できない問いは、適切な実質的・方法論的専門性を持つ人によって評価されるべきである。

第5層：制度的措置

訂正、懸念表明、撤回、反論、棄却、研究公正上の調査は、それ以前の水準で次の段階へ進むことが正当化されて初めて行われるべきである。

市民科学は、層状の検証を通じて分散した情報提供を管理する、隣接領域の先例を与える。Baker et al. (2021) は、生態学的な市民科学の259事業を調べ、そのうち142事業で検証に関する情報を見いだした。彼らは専門家、共同体、自動化によるアプローチを記述し、問題が示された記録を専門家へ送る階層的検証を提案する。科学的批判は生物種の記録の検証とは異なるが、制度上の原則は参考になる。幅広い参加が、すべての段階で一様な検証を意味する必要はない。この構造は、参加と裁定を分離する。幅広い参加が最も価値を持つ地点、すなわち問題の可能性を発見する地点で、それを認める。批判の帰結が大きくなるにつれ、より選択的になる。この原則は、次のように表せる。

発見の段階では開放性を、措置の段階では厳密さを。

7.7. 査読から、より広い検証の生態系へ

従来の査読は、より広い**検証の生態系**の中で、専門家の判断を担う中心的な制度であり続ける。読者と独立した分析者は公表結果を検討し、計算システムは異常検出を支援し、領域専門家は曖昧な知見を評価し、編集者と機関は正当化された措置を決める。生成AIは、これらの主体間における能力の分布を変える。問題は、彼らの共同の働きが、拡大した精査を信頼できる科学的訂正へ変えるかどうかである。批判の潜在量が説明責任を伴う裁定より速く増える場合、より広い異議提起へのアクセスには、証拠の優先順位付けと争点の解決の仕組みが伴わなければならない。層状構造は、すべての評価課題を正式な査読者に割り当てることなく、それらの要件を位置付ける。

7.8. 適用条件と限界

本稿は、ある機構とその制度的含意を概念的に説明するものであり、その普及度や正味の効果の推定ではない。その適用範囲は、科学的資料への利用可能なアクセス、課題に適したAI支援、少なくとも自ら報告する証拠上の手順を確認できる利用者を必要とする。作業コストが低いことは、アクセスが公平であること、すべての分野で評価が改善すること、機関が受け取る異議のうち妥当なものの割合が増えることを確立しない。

P3には、参加の増加を超える経験的含意がある。特定された課題とアクセス条件の下で、AI支援は、割り当てられた評価上の役割の外にいる主体が、独立に妥当性を確認された精査の構成要素をより多く、またはより低いコストで生み出せるようにするはずである。比較では課題の難しさと利用可能な資料を一定に保ち、異議をその出所とは独立に評価すべきである。証拠に基づく遂行が改善せず、批判がより流暢になっただけでは、P3を支持しない。明示した条件にわたって、妥当性を確認された遂行の改善やコスト低下が繰り返し得られないなら、その状況における提案機構に反する証拠となる。

承認された評価上の役割の外から来た、独立に妥当性を確認された異議が、確立された評価者による同程度に関連性があり十分に裏付けられた異議よりも、制度的検討を受けるまでに時間がかかる、あるいは実質的な検討が少ないなら、P4は支持を得る。この比較では、出所の地位を、証拠、関連性、手続き上の要件の違いから区別しなければならない。これらの集団間で一貫して同等かつ適時で、証拠に即した受け入れが行われているなら、調査対象の機関における、提案された不一致に反する証拠となる。

P5は、経験的な根拠と規範的な提言を組み合わせている。来歴、再現可能性、証拠の具体性が、AI利用の開示や技術的な外観を超えて妥当な異議と不当な異議の区別に役立ち、技術的確認だけでは不十分な場合に適切な専門家の裁定が決定を改善するなら、その経験的根拠は支持を得る。これらの特徴が判別上の価値を加えない、あるいはその負担が決定を改善せずに妥当な異議を体系的に排除するなら、提案した実装には改訂が必要である。また、規範的提言は、開かれた異議提起と説明責任を伴う制度的対応に置かれる価値にも依存する。その価値への関与は、観察結果だけでは決着しない。

最後に、コストの非対称性は条件付きである。検証と説明責任を伴う裁定が、批判の候補と少なくとも同じ速さで規模を拡大するなら、予測されたボトルネックは生じない。この結果は規模拡大の議論に異議を与えるが、分散型精査という、より広い可能性を反駁するものではない。

8. 結論

生成AIは、科学的主張の生産だけでなく、その吟味への障壁も下げうる。アクセスと能力の区別は、それがなぜ重要かを説明する。論文、データセット、分析手法へのアクセスは、それ自体では、それを評価する能力を与えない。AI支援は、アクセスを達成した後にも残る技術的コストの一部を減らしうる。これが**二重の民主化**の精査側の構成要素であり、P3では、割り当てられた役割を超えた実質的な評価課題の拡大として特定された。

外部の批判を受け取り、重み付けする制度の経路が評価能力の拡大に追い付かない場合、その拡大は**能力と承認の不一致**を生む。既存の出版後の仕組みは、そのような承認が可能であることを示す一方、閾値、責任、規模について問いを残している。能力上の障壁を下げることは、もっともらしいが不当な異議への障壁も下げる。そこから生じる**精査の品質問題**は、制度的対応に値する批判をどう特定するかに関わる。

統括的な原則は、**発見の段階では開放性を、措置の段階では厳密さを**である。批判は、特定の主張を明らかにし、それを点検可能な証拠に結び付け、実行可能な場合は技術的検証に耐え、必要な場合には適切な領域専門家の裁定を受けるべきである。そのうえで、精査に耐えた批判に条件付きの承認を与えられる。これは専門性の役割を保ちながら、従来型ではない評価者が科学の訂正に貢献することを可能にする。

アクセスと一部の能力上の障壁が低下すると、実践上のボトルネックは、検証、優先順位付け、裁定へますます移るかもしれない。そのため科学の諸制度は、異なる配分問題に直面しうる。批判を可能にする方法だけでなく、どの異議が追加の注意に値するか、そしてより重大な措置を取る前にどの程度の証拠を求めるべきかを特定する方法である。したがって、AIが媒介する環境の中で、科学の中心的な問いは、次から離れていくかもしれない。

誰に批判する資格があるか。

そして、ますます次へ向かうかもしれない。

科学が対応すべきものとなる前に、批判は何を示さなければならないか。

生成AIは、誰が科学的主張を生み出せるかを変えることだけによって、科学を民主化するのではない。誰がその主張に異議を唱えられるかも変えうる。その移行が科学の自己訂正を改善するかどうかは、異議提起への開放性を維持しつつ、検証に耐える証拠に信頼性を依存させることが、諸制度にできるかどうかによって決まる。

参考文献

Alvarez, C., Bennett, M., & Wang, L. L. (2024). Zero-shot Scientific Claim Verification Using LLMs and Citation Text. In *Proceedings of the Fourth Workshop on Scholarly Document Processing (SDP 2024)* (pp. 269–276). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.sdp-1.25

Baker, E., Drury, J. P., Judge, J., Roy, D. B., Smith, G. C., & Stephens, P. A. (2021). The Verification of Ecological Citizen Science Data: Current Approaches and Future Possibilities. *Citizen Science: Theory and Practice, 6*(1), Article 12. https://doi.org/10.5334/cstp.351

Caron, M. M., Lye, C. T., Bierer, B. E., & Barnes, M. (2025). The PubPeer conundrum: Administrative challenges in research misconduct proceedings. *Accountability in Research, 32*(8), 1369–1387. https://doi.org/10.1080/08989621.2024.2390007

Clark, J., Barton, B., Albarqouni, L., Byambasuren, O., Jowsey, T., Keogh, J., Liang, T., Moro, C., O'Neill, H., & Jones, M. (2025). Generative artificial intelligence use in evidence synthesis: A systematic review. *Research Synthesis Methods, 16*(4), 601–619. https://doi.org/10.1017/rsm.2025.16

Committee on Publication Ethics. (2008). *What to do if you suspect fabricated data: (b) Suspected fabricated data in a published article* [Flowchart]. https://content.ifzg.hr/poveznice/Ethical_codex_for_Editors-COPE-english.pdf

Croce, M., & Marsili, N. (2025). Misplaced Trust in Expertise: Pseudo-Experts and Unreliable Experts. *Social Epistemology, 39*(6), 667–682. https://doi.org/10.1080/02691728.2025.2491104

Daepp, M. I. G., Tomlinson, K., Counts, S., & Suri, S. (2026). AI and the democratization of knowledge work. *Nature Computational Science, 6*(6), 558–564. https://doi.org/10.1038/s43588-026-00985-z

Edelman, B., & Skolnick, J. (2025). Valsci: an open-source, self-hostable literature review utility for automated large-batch scientific claim verification using large language models. *BMC Bioinformatics, 26*(1), Article 140. https://doi.org/10.1186/s12859-025-06159-4

Groth, P., & Moreau, L. (Eds.). (2013). *PROV-Overview*. W3C Working Group Note. https://www.w3.org/TR/2013/NOTE-prov-overview-20130430/

Hao, Q., Xu, F., Li, Y., & Evans, J. (2026). Artificial intelligence tools expand scientists' impact but contract science's focus. *Nature, 649*(8099), 1237–1243. https://doi.org/10.1038/s41586-025-09922-y

Hardwicke, T. E., Thibault, R. T., Kosie, J. E., Tzavella, L., Bendixen, T., Handcock, S. A., Köneke, V. E., & Ioannidis, J. P. A. (2022). Post-publication critique at top-ranked journals across scientific disciplines: a cross-sectional assessment of policies and practice. *Royal Society Open Science, 9*(8), Article 220139. https://doi.org/10.1098/rsos.220139

Keren, A. (2025). Expert Authority and Its Assessment. *Social Epistemology, 39*(6), 612–625. https://doi.org/10.1080/02691728.2025.2472780

Markowitz, D. M. (2024). From complexity to clarity: How AI enhances perceptions of scientists and the public's understanding of science. *PNAS Nexus, 3*(9), Article pgae387. https://doi.org/10.1093/pnasnexus/pgae387

Mirkin, J., & Sojka, M. M. (2026). From Expertise to Trustworthiness: What Laypeople Can Really Assess about Experts. *Episteme*, 1–18. https://doi.org/10.1017/epi.2026.10118

Nabavi, A., Safari, F., Shmoury, A. H., Tabet, S., Perdomo-Luna, C., & Celi, L. A. (2026). Artificial intelligence in scholarly peer review: a scoping review of applications, risks, and governance challenges. *International Journal of Medical Informatics, 214*, Article 106418. https://doi.org/10.1016/j.ijmedinf.2026.106418

PubPeer. (n.d.). *Frequently asked questions*. Retrieved September 5, 2026, from https://pubpeer.com/static/faq

Rebitschek, F. G., Carella, A., Kohlrausch-Pazin, S., Zitzmann, M., Steckelberg, A., & Wilhelm, C. (2025). Evaluating evidence-based health information from generative AI using a cross-sectional study with laypeople seeking screening information. *npj Digital Medicine, 8*(1), Article 343. https://doi.org/10.1038/s41746-025-01752-6

Reddy, C. K., & Shojaee, P. (2025). Towards Scientific Discovery with Generative AI: Progress, Opportunities, and Challenges. *Proceedings of the AAAI Conference on Artificial Intelligence, 39*(27), 28601–28609. https://doi.org/10.1609/aaai.v39i27.35084

Rolin, K. (2025). Expert trustworthiness and the value-free ideal of science. *European Journal for Philosophy of Science, 15*(4), Article 77. https://doi.org/10.1007/s13194-025-00704-x

Romeo, G., & Conti, D. (2026). Exploring automation bias in human–AI collaboration: a review and implications for explainable AI. *AI & Society*, *41*(1), 259–278. https://doi.org/10.1007/s00146-025-02422-7

Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research, 6*, Article 588. https://doi.org/10.12688/f1000research.11369.2

Sakai, Y., Kamigaito, H., & Watanabe, T. (2026). HalluCitation Matters: Revealing the Impact of Hallucinated References with 300 Hallucinated Papers in ACL Conferences. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 47295–47376). Association for Computational Linguistics. https://doi.org/10.18653/v1/2026.acl-long.2189

UNESCO. (2021). *UNESCO recommendation on open science*. https://www.unesco.org/en/legal-affairs/recommendation-open-science

Wager, E., & Kleinert, S. (2021). Cooperation & Liaison between Universities & Editors (CLUE): Recommendations on best practice. *Research Integrity and Peer Review, 6*, Article 6. https://doi.org/10.1186/s41073-021-00109-3

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*(1), Article 160018. https://doi.org/10.1038/sdata.2016.18

Wingerter, T. L., Straub, T., & Schweitzer, S. (2025). Mitigating Automation Bias in Generative AI Through Nudges: A Cognitive Reflection Test Study. *Procedia Computer Science, 270*, 2106–2114. https://doi.org/10.1016/j.procs.2025.09.331

Zhang, Y., Khan, S. A., Mahmud, A., Yang, H., Lavin, A., Levin, M., Frey, J., Dunnmon, J., Evans, J., Bundy, A., Dzeroski, S., Tegner, J., & Zenil, H. (2025). Exploring the role of large language models in the scientific method: from hypothesis to discovery. *npj Artificial Intelligence, 1*(1), Article 14. https://doi.org/10.1038/s44387-025-00019-5

宣言事項

資金提供

本稿の執筆に際して、資金、助成金、その他の支援は受けていない。

利益相反

著者に、開示すべき関連する金銭的・非金銭的利益相反はない。

データ利用可能性

本稿は概念論文であり、データセットの生成・分析は行っていない。

生成AIの利用

草稿作成、改稿、文献探索、原稿統合に生成AIツールを利用した。AIによる提案と書誌情報は原稿準備中に確認した。議論、出典の解釈、最終稿に対する責任は著者が負う。