UTIE Instruments Inc.

UTIE Research Institute

Whitepaper Series 2026-2

AI Slop の強要により成果物の質的劣化はどう引き起こされるか

― シュプリンガー・ネイチャー社の査読プロセス事例とした分析 ―

発行日: 2026 年 2 月 23 日

---

*本報告書で取り上げたシュプリンガー・ネイチャー（Springer Nature）社における事案の詳細については当社の公式アーカイブにて公開しているので事前に参照されたい。

1.エグゼクティブ・サマリー

本報告書は、現代の学術出版システムにおいて急速に蔓延している「LLM（大規模言語モデル）による非開示の査読（AI 査読）」が、科学的成果物の質的劣化をいかにして引き起こしているか、そのメカニズムを解明することを目的とする。シュプリンガー・ネイチャー社（以下、対象企業）の傘下ジャーナルにおけるインシデントの AI 生成物フォレンジック解析を通じて、本稿は単なる一企業の不正告発にとどまらず、非対称な権力構造下で発生する権力による強制的なモデル崩壊（Coercive Model Collapse）の概念を提示する。実証解析の結果、対象企業の査読プロセスにおいて、AI が生成した粗悪な出力（AI Slop）が人間の著者に強要されることで、①無難な指摘による「科学の多様性の縮減（テンプレート飽和）」と、②人間のチェック機能（Human-in-the-loop）の形骸化による致命的な事実誤認（ハルシネーション）の流通という事態が現場で引き起こされている事実が確認された。対象企業の「人間の専門家による厳格な査読」という宣言は事実上形骸化しており、高額な論文掲載料（APC）を徴収しながら、実態は AI のハルシネーションと AI テンプレートによる知の再帰的な汚染サイクルを駆動させる装置へと変質している。本報告書は、この危機が学術出版システム全体に内在する構造的欠陥であると断ずるとともに、HR（人事）や教育・審査分野といった社会全般の評価プロセスにも波及する重大な脅威であると警鐘を鳴らす。結論として、既存の「査読済み」ブランドに依存しない「ゼロトラスト・サイエンス」への移行と、独立した第三者機関による AI フォレンジックを用いた新たな保護枠組みの構築を提言する。

2.イントロダクション

本稿では、計算機科学におけるモデル崩壊（Shumailov et al., 2024）の概念を拡張し、権力構造を介した対象物の劣化を「権力による強制的なモデル崩壊（Coercive Model Collapse）」として定義する。これは、次の 2 つの種類に大別できる。一つ目は多様性の強制的な刈り取り（テンプレート飽和）、すなわち今回の事例においては AI 査読者の無難で尤もらしい指摘により、著者が自らの独創性を削り落とし、分布の尾部が失われることを指す。2 つ目は制度的なハルシネーションの追認（Human-in-the-loop の形骸化）であり、本事例においては論文テーマへの根本的な事実誤認（例：「GPT-5 は存在しない」「4o ショックやハッシュタグキャンペーンは実在しない」等）に対して、権力側が「専門的知見である」と強弁し、誤謬を強権的にファクトとして流通させる事態としてあらわれた。Naddaf (2026) が示すように、研究者の過半数が既にポリシーに反して AI を査読に使用しており、また Chawla (2025) が報告するように AI 検知ツールはこの事態を捕捉できていないという現実である。すなわち、Coercive Model Collapse は単一の出版社の例外的な不祥事ではなく、学術出版システム全体において同時進行的に発生している構造的危機である。

3. 対象企業における不正と隠蔽の事実関係

対象企業に投稿された論文（2025 年 11 月提出）の査読プロセスにおいて、以下の重大な研究倫理違反が確認された。

(1) LLM の知識期限に起因するハルシネーション

2025 年 11 月に行われた AI を専門とする学術雑誌の査読者レポート（Reviewer 2）において、「2025 年 8 月の出来事はまだ起こっていない」「GPT-5 は存在しない」という、人間の認知能力では到底あり得ない数々の主張が提示された。これは、学習期限が GPT-5 リリース以前でかつ内部検索機能を作動させなかった LLM 特有のハルシネーションである。この査読プロセスにおいて、同社査読者と編集部は「査読レポートとしての体裁の良さ」のみを評価し、致命的なミスを検知できない状態に陥っており、人間のチェックによる安全性の確保(Human in the loop)が完全に失われていることを示している。

(2) 管理部門（Laura Whitehead 氏と研究倫理部門）による組織的隠蔽

対象企業の編集部、サポートセンター、管理職および Research Integrity 部門は、著者の度重な

る調査の要請に対し、徹底的な問題の無視、隠ぺいを選択した。これは、出版規範委員会(COPE)のコアプラクティス違反にあたる。出版規範委員会(COPE)が正式に動いたのち、編集部は調査結果を発表した。本社 Research Integrity 部門も 2026 年 2 月に調査報告を出したが、「AI 生成の証拠はない」「人間の正当な判断である」「プロセスに違反なし」と編集部の報告を追認した。特筆すべきは、管理部門にとっては、これが AI の出力を見抜けない無能や過失（認知機能の劣化）ではなく、「事態を把握した上で、AI Slop（粗悪な生成物）を著者に押し付ける方が、企業にとって都合が良かった」という、極めて悪意ある経営的判断の表れであるという点だ。

もし対象企業がここで AI 査読の実態を認めてしまえば、以下の 3 つの致命的危機（ドミノ倒し）を引き起こすことになる。そして、これらは集団訴訟のリスクとなりうる。

1. **明確なポリシー違反**: 同社が公式に掲げる「査読における AI 使用禁止」が完全に形骸化していることの自白。
2. **重大な守秘義務違反**: 著者の未公開データ（投稿論文）を、査読者が無断で外部の LLM プロバイダに送信・入力しているという法的・倫理的逸脱の露呈。
3. **APC（論文掲載料）ビジネスモデルの崩壊**: 「人間の専門家による厳格で高度な判断」という付加価値を大義名分として数千ドルの高額な掲載料を徴収する、同社の搾取的収益構造の否定。

したがって、管理部門にとっての至上命題は真理の究明ではなく、「自社のビジネスモデルの防衛」であった。そのために彼らが採った手法は、無視、暴言、アカデミック・ガスライティング(後述)を組織ぐるみで著者に浴びせ、精神的に追い詰めて告発を諦めさせることであった。

(3) 編集部と本社によるガスライティング

著者が上記(1)の「AI 特有の論理破綻」を指摘したにもかかわらず、担当編集者である Da Tao(深セン大学)および Yina Liu(Discover シリーズ)は 2 か月に渡る慎重な調査ののち、不正行為の継続を決断した。「GPT-5 が存在しない」という論文テーマの実在に関する致命的な事実誤認を「非常に些細な要素（very minor element）」「非常に些細な側面（very minor aspect）」などと一蹴した。「架空の研究と明記すれば価値ある貢献」架空と断言したデータについて「IRB 承認」を要求するという AI のハルシネーションについても、「どのジャーナルでもこのようなミスは時々起こる」などと正当化し、この査読者(すなわち AI)は「注意深い精読と領域の専門知識（careful reading and domain expertise）を反映している」「基礎的なトレーニングを受けていれば誰でも同じ結論に達する」「このような品質の原稿を受理することのほうが、リジェクトするよりもレビュープロセスの誠実性についてはるかに深刻な懸念を引

き起こす（raise far more serious concerns about the integrity of the review process）と報告した。これらは、編集部自身が述べたように、2 か月に渡る慎重な調査によって導かれた結論であり、研究公正倫理部門もこの結論を追認している。すなわち、編集部と本社は AI のハルシネーションを鵜呑みにすることが基礎的トレーニングを受けたあらゆる研究者の共通的結論であると報告し、著者の能力不足へ論点をすり替えたほか、人間による実証論文の掲載は科学誠実性に「不適切」と退ける等の明確なガスライティングを行った。

---

3. 質的劣化のメカニズム：AI Slop はいかにして科学を破壊するか

対象企業の事例は、偶発的な事故ではなく、Naito (2025) が提唱する「AISP（AI Selection Pressure」理論に基づく構造的欠陥である。査読プロセスにおいて、AI 査読者による改訂強要が発生すると、以下の 3 つの段階を経て科学的成果の質的劣化を引き起こす。

3.1 テンプレートの飽和（Template Saturation）

計算機モデルにおける Early Model Collapse では、分布の尾部（tail）が最初に失われる。即ち、少数派のデータポイントが消去され、分布の中心部は一見維持されるため、崩壊は外部から視認されにくい。Coercive Model Collapse の文脈において、これに対応するのが「テンプレート飽和による科学の多様性縮小」である。AI 査読者は、「サンプル数を増やせ」「もっと広範な変数を制御せよ」といった、どの論文にも適用可能だが具体性を欠く要求を乱発する。査読者や編集者は認知コスト削減のため、このもっともらしい指摘をそのまま採用する。これにより、査読は論文を改善するプロセスからどの論文も当てはまる無難なテンプレの押し付けへと退化する。

3.2 権力による強制的なモデル崩壊（Coercive Model Collapse）

採択権を握られた著者は、AI のハルシネーションや的外れなテンプレ指摘に対しても、正当な反論（Rebuttal）を行うことができない*。結果として、研究者は掲載されるために自らの独創的な論理や画期的なデータを削り落とし、自らの手で「AI が次に学習しやすい、均質化されたテンプレ文章」へと論文を書き換えることを強いられる。これは、分布の尾部に位置する新規性の高い研究、少数派の方法論、学際的な知見が優先的に排除されることを意味する。現時点でこの崩壊を技術的に停止させる手段は存在しない。
計算機モデルにおける Late Model Collapse では、分布が単一のモードに収束し、元の分布とは似ても似つかぬものに変質する。Coercive Model Collapse の文脈では、これに対応するのが人間によるチェックの安全性（Human in the Loop）の完全な喪失である。この崩壊では、AI のハルシネーションがそのまま科学的事実として流通する。対象企業の事例では、「GPT-5 は存在し

ない」という、AI を専門とするジャーナルの査読であれば決してあり得ない水準の事実誤認が発生した。より広範には、「フィクションであると明記すれば実証論文として価値ある貢献になる可能性」等の、人間の専門家(実際は AI 査読者)が論文の実在する研究対象を否定した上で「それでも代替的価値がある」と提案するという、もはや現代科学そのものを否定するレベルの誤った指摘までもが現場で発生した。これらはすべて AI のハルシネーションに起因するものであり、人間の査読者が犯す誤りとは質的に異なる。

3.3 再帰的汚染サイクル（Recursive Pollution Cycle）の完成

上記のプロセスを何度か経て、AI の要求に合わせて著者が改訂した論文が、『Nature』等のブランドの下で「査読済み論文」として出版される。それが次世代 AI の学習データとして摂取されることで、科学の多様性（ばらつき）は収縮し、平均化されたテンプレートや嘘は科学的ファクトとして定着していく。「ゆっくりとした多様性の縮小」も長期的には問題になる可能性が高いが、間違いなく現時点で深刻な問題なのは「短期的かつ致命的な AI のミス」がそのまま現場で通ることである。前者は科学の基盤を徐々に侵食するが、後者は即座に危機的な結果をもたらす。医学、工学、薬学などの領域で、AI のハルシネーションが査読を通過した論文が「査読済み」の権威を以て実務に適用されれば、その影響は直接的に人命に及ぶことは想像に難くない。この AI 査読による再帰的汚染サイクルは、著者(人間)による学習と AI によるチェックへの先回り的な対策を経て、2 種類の問題の両方を増幅させるフィードバックループとして機能する。

*正当な反論（Rebuttal）を行うことができない

対象企業は、表面上は「Speak Up（内部告発・相談窓口）」や「不服申し立て（Appeal）」のプロセスを整備し、著者の正当な反論に対して透明かつ協力的であるかのように表明している。しかし実態は、著者がこれらの正規ルートを用いて AI 査読等の明らかな不正を報告しても、以下の「防衛プロトコル」によって完全に無力化される。

a. 外部圧力（COPE）の必須化と門前払い

著者が単独で明確な証拠（AI ハルシネーションのログ等）を提示しても、研究公正部門はケース番号すら発行せず、調査拒否を貫く。COPE（出版倫理委員会）等の外部機関からの公式な介入があって初めて調査が開始される。

b. 内部告発（Speak Up）システムの機能不全

コンプライアンス窓口に重大な倫理違反を報告しても、その内容は独立した第三者ではなく、告発対象である担当編集部（不正の当事者）へそのまま差し戻される。これにより、組織的な

不正の問題すら、著者による単なる「判定への不満（routine appeal）」へと強制的にリフレーミングされ、無意味な追加データの要求や数週間の遅延によって著者の戦意喪失が図られる。

c. 管理部門の形骸化と循環論法

最も異常なのは、調査における論理の破綻である。パブリッシャー側や研究公正部門の管理職は客観的な鑑識（フォレンジック）や独立した調査は行わず、「担当編集部が違反はないと回答したため、プロセスは正当である」と当事者の自己申告をそのまま追認する。さらに編集部側は、AI 特有の事実誤認という証拠を提示されても、「当社の査読における AI 利用には整備されたポリシーがある。ゆえにこの査読が AI 丸投げであるという証拠はない」「査読プロセスと結果は正当なものであったが、AI ポリシー違反があったかについての調査は続ける」などという、ルールの存在を以て物理的証拠を否定する非科学的な循環論法や、調査の意味すら分かっていないかのような支離滅裂な主張を展開する。

結論として、著者がどれほど科学的・客観的な証拠（正当な反論）を提示しても、事実の否定によって握り潰される。これが、権力構造下において著者が AI Slop を強要され、泣き寝入りせざるを得ない（つまり Coercive Model Collapse を受け入れるしかない）最大の要因である。この査読現場でのプロトコルはアシスタントエディタレベルにまで徹底されており、異議申し立ては、新規投稿よりも後回しにされ決定まで数週間を要すること、そして*「大半のケースで元の決定が維持される（original decision will be upheld）」*ことが強調されていた。すなわち、反論しても無駄であり、時間の浪費になるだけだ、と暗にメッセージを送ることで、著者の戦意を喪失させ、「いったん編集部が正規に報告した結果は、仮に明らかな問題があったとしても覆らないのだ」とそのまま受け入れさせるよう誘導を行っている。

---

4. 結論と実務的な是正へ

対象企業の査読システムは、AI による嘘（ハルシネーション）を真実として上書きする装置へと変貌している。本事例は、AI 査読の実務的な黙認が、単なる質の低下にとどまらず、アカデミック・ハラスメントやデータ捏造の強要へと直結していることを示した。既往研究と本稿が提起した Coercive Model Collapse の構造は、この危機が単一の出版社の不祥事ではなく、学術出版システム全体に内在する構造的危機であることを示唆している。よって、良識ある科学コミュニティおよび関係機関は、以下の「ゼロトラスト・サイエンス」*への移行を実行すべきである。

*ゼロトラスト・サイエンス

すべてのプレプリントを「信頼できるプレプリント」として指定することは、現実的にも望ましいものでもない。当社はその前提を共有した上で、実社会への影響が大きい文脈において限定的な担保を発動する必要性を認識している。とりわけ研究成果が、研究資金配分、医療判断、規制評価、査読プロセスに関する紛争に用いられる場合には、関係主体は以下を積極的に活用すべきである。

- 査読プロセスのシステムログ開示（AI 利用痕跡・プロンプト履歴を含む）
- 独立第三者による監査（AI 生成物フォレンジックを含む）

出版時点で一律の保証を付与するのではなく、品質担保メカニズムはイベント型とし、現実世界への影響とリスクの大きさに応じて段階的に強化されるべきである。このアプローチはプレプリント文化の開放性を維持しつつ、低品質な AI 生成コンテンツが重大な意思決定領域において不当に影響を持つことを防ぐ。

4-1「査読済み」ブランドの無効化

密室で AI と不誠実な編集者による「検閲・改ざん」を受けた査読済み論文よりも、信頼できる著者や組織による生データ(プレプリント)を一次情報として高く評価する基準へ移行し、ポリシー違反の AI 査読が当然となっている学術出版社やいまだに AI 査読に関するポリシーや運用が不透明な学術出版社による「査読済み」の権威は信頼すべきでない。査読は品質保証制度ではなく弱い品質シグナルにすぎない。

4-2 第三者機関による「AI 生成物フォレンジック」サービスの積極的活用

UTIE Instruments Inc.等が提供する、人間の専門家による AI 生成物フォレンジックを用い、論文の内容に「低品質な AI 生成論文の痕跡（時系列の矛盾やテンプレ的収束、ハルシネーション）」が含まれていないかを技術的・客観的に監査するプロトコルを、公的研究費の配分基準等に組み込むべきである。

4-3 権力による強制的な順応からの救済と倫理的保護枠組みの構築

本事例が示した危機は、AI そのものの欠陥というよりは、「権力を持つ側が AI に仕事を丸投げし、その結果（ハルシネーションやテンプレートへの順応）を立場の弱い者に強制的に同化させる構造」である。この「権力による強制的なモデル崩壊（Coercive Model Collapse）」から人間を保護するため、早急に以下の枠組みを構築する必要がある。

4-3-1 学術界における「第三者機関（COPE 等）」による介入とガイドライン策定

自らの利益（高額な APC モデルと AI によるコスト削減）を最優先する出版社内部の「研究公

正部門」には、自浄作用も客観的な調査能力も存在しないことが明らかになった。したがって、不当な AI 丸投げ査読に対する異議申し立ての受理や、査読プロセスのシステムログ(AI 使用の痕跡やプロンプト履歴等）の開示要求については、COPE（出版倫理委員会）のような独立した第三者機関が率先して管理・介入する仕組みへと移行すべきである。著者らが企業との紛争で多大な労力を払わずにすむためにも「AI 丸投げ査読に対する透明性・救済ガイドライン」の策定が急務である。

4-3-2 社会全般（HR・教育・審査分野）における「不当な AI 評価」からの保護

この非対称な権力構造による低品質な AI 生成物の押し付けは、学術出版にとどまる問題ではない。現在、全世界において同時進行的に採用面接、社内の人事評価、助成金や融資の審査など、人生を左右するあらゆる評価プロセスにおいて、評価者がコスト削減のために AI へ判定を丸投げする事態が増えつつある。その結果、独自の強みや画期的なポテンシャルを持つ人材が、AI の学習データ（過去の平均値）から外れているという理由だけで「不適格」とされ、AI のスクリーニングによって排除されることがおこり得る。評価対象者が、ブラックボックス化された AI の判定（およびそれの妥当性をチェックする能力がない人間や組織）に対して、判定根拠の開示要求や人間の専門家による再審査を要求できるセーフティネットを、社会全体として早急に構築しなければならないと当社は考えている。

参考文献

Chawla, D. S. (2025). “‘A serious problem’: peer reviews created using AI can avoid detection.” Nature.

Naddaf, M. (2026). “More than half of researchers now use AI for peer review — often against guidance.” Nature.

Naito, H. (2025). “AI Selection Pressure: Template Saturation and the Reshaping of Human Discernment.” Zenodo. https://doi.org/10.5281/zenodo.17644956

Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., & Anderson, R. (2024). “AI models collapse when trained on recursively generated data.” Nature, 631, 755–759.