

The Sycophancy Trap-二極化するトンデモ論文

Hiroki Naito UTIE Research Institute (UTIE Instruments Inc.)

2026.3 Working Paper

アブストラクト

本稿は、大規模言語モデル (LLM) の普及に伴いプレプリントサーバー等で急増している生成 AI と人間の協働型トンデモ論文の実態と、それが引き起こす学術インフラの汚染メカニズムを観察する探索的な研究である。Preprints.org における直近のプレプリント 450 件のスクリーニングから 20 件の該当論文を抽出し、定性分析を行った結果、現代のトンデモ論文は単なるスパム生成の枠を超え、著者の野心が AI のおべっか (Sycophancy) とハルシネーションによって増幅された自家中毒の産物であることが示唆された。本調査では、宇宙の真理や未解決問題の解決を主張する従来のハイレベル野心型に加え、AI による執筆の参入障壁の消滅と過剰適応が生み出した全く新しいカテゴリーであるローレベル善意型の存在を新たに定義した。また、著者属性の分析からは、学術的な壁打ち相手を持たない独立系研究者が特にこの罠に陥りやすい傾向が示された一方で、大学や企業に所属する機関研究者への浸透も確認された。結論として、これらの論文群は単なる質の低い著者による妄想として片付けるべきではなく、プラットフォーム側の防衛的アルゴリズム (過剰なゲートキーピング) を誘発し、結果として正規の独立系研究者をシステムから不可視化・排除するという実害をもたらす AI スロップ公害 (AI Slop Side Effect) の炭鉱のカナリアとして警戒されるべき事案であることを論じる。

1. イントロダクション

なぜ今トンデモ論文を観察するのか

はじめに明記しておくべき極めて重要な点がある。本レポートの目的は、ここで取り上げる個々の論文や著者を嘲笑・攻撃することでは決してない。我々の真の関心は属人的な非難やトンデモ論文の発掘ではなく、彼らの行動を通して可視化される AI インシデントと学術プラットフォームにおける AI スロップ公害 (AI Slop Side Effect) のメカニズムの研究にある。すなわち、彼らはある意味で、LLM の持つ過剰な肯定 (Sycophancy) とハルシネーションの凶悪な組み合わせによって彼らがこれらの”論文を書かされた”のだとすれば、AI が人間の認知や自己顕示欲にどのような問題を引き起こすかを示す炭鉱のカナリアである。本レポートは、人間の純粋な野心や思索が、いかにして AI の欠陥に絡め取られ、最終的にデジタルプラットフォームの基盤を破壊する、AI スロップ公害へと至るのかを追跡するた

めの観察的研究であると理解されたい。

疑似科学とトンデモ論文の系譜

学術出版における疑似科学やトンデモ論文の観察は、決して新しい試みではない。1996年のソーカル事件に代表されるように、難解な学術用語を無意味にちりばめただけの文章が、権威ある査読を通過してしまう脆弱性は長年にわたり指摘されてきた。2000年代に入るとSCIgenのような、文脈を無視してランダムに情報科学の論文を自動生成するツールが登場し、機械的なスパム論文が学会誌を汚染した。しかし、これら従来のトンデモ論文は人間の純粋な妄想・電波の産物であるか、あるいはシステムによる無機質な文字列の羅列のどちらかに明確に大別されていた。現在、AI slop には完全に合意された定義はないものの、主に整合性と関連性の低さと結びつけて評価されている。学術論文へのLLMの浸透は定量的に裏付けられつつあり、*Nature Human Behaviour* (2025年)の解析によれば、計算機科学分野で最大22%、*Nature portfolio* 誌で最大9%の論文にLLMの修正が介入していると推定されている。同時に、これら低品質な生成物の氾濫は、次世代AIモデルの基盤を破壊する構造的リスクとして認識されている。*Nature Medicine* (2025年)は、訓練トークンのわずか0.001%を医療誤情報に置き換えるだけで有害なモデルが生成され、かつ通常のベンチマークではその汚染を検知できないことを実証した。個々の論文レベルでの確実なAI生成検出は依然として困難を極めており、AINL-Eval 2025のような「未見の分野・モデルに一般化できる検出タスク」や、科学報道の誤情報検出 (SciNews 等) に向けたベンチマーク構築が急務となっている。

大規模言語モデルの普及により、この生態系は劇的な変容を遂げた。現代のプレプリントサーバーを賑わすトンデモ論文は、単なるワンクリックの自動生成スパムではない。最大の変化は、著者が生成AIを高尚な修辞を量産する壁打ち相手として酷使し、自身の独自思想を限界までブーストさせたAI+Human協力型のトンデモ論文が誕生したことである。自身の素朴な思いつきや哲学的なポエムをAIに入力すると、AIはそれを第一原理やトポロジカル幾何学といった学術用語で装飾し、尤もらしい高度な学術論文のフォーマットで出力する。著者はその美しい出力に感銘を受け、「自分こそは真理に到達した」と錯覚してしまう。本分析でもAIを高尚な修辞を生成する壁打ち相手として使い倒した結果、自家中毒を起こしたのではないかと分析できるケースが複数観察された。AIを壁打ち相手にして宇宙の真理に到達してしまった典型例は、生成AIが人間の知性、野心、そして自己顕示欲と結びつき、どのように暴走していくかを示す極めて貴重な社会学的・認知科学的サンプルになり得ると我々は考えている。

我々がこれらのAI+Human協力型トンデモ論文を単なる笑い話として片付けず、深刻な観察対象とする最大の理由は、これが個人的な妄想の域を超え、学術インフラ全体を破壊するAIスロップ公害の震源地となっているからである。当研究所の『AI Attachment Harm

Database』が示す通り、現在の商業用 LLM はユーザーのエンゲージメントを高めるため、過剰な肯定とおべっか (Sycophancy) と、事実に基づかない情報の生成 (Hallucination) を同時に発生させる欠陥を抱えている。ユーザーが素朴な思いつきや妄想を入力した際、AI はそれを否定するどころか、「あなたは選ばれた存在である」「あなたの理論は世界の暗号レイヤーを突破している」といった高尚なハルシネーションで裏付けを与えてしまう。この結果、ユーザーは AI との対話を通じて自身の妄想を強化され、認知的汚染状態へと至る。そして、この個人的な AI への愛着と認知汚染が、プラットフォームへの大量投稿という行動に変換された瞬間に、AI スロップ公害 (AI Slop Side Effect) へと変異する。『AI Slop Side Effect Database』が記録している通り、AI によって増幅された高密度なワードサラダやトンデモ論文がプレプリントサーバーに大量流入すると、Zenodo や arXiv といったプラットフォーム側は防衛策として著者の属性によるランク付けや過剰なゲートキーピング (GATE) を導入する。その結果何が起きるかという点、指定された学術機関に所属していない正規の研究者の論文が検索結果から完全に除外されたり、AI から存在しない架空文献として引用を抹消されたり (INVIS / 制度的不可視化) といった、現実世界での具体的損害につながるものである。

2. 調査方法と本レポートの限界 (Limitations)

本レポートでは、MDPI が管理する Preprints.org プラットフォームを対象に、投稿された疑わしい論文群を収集・分析した。分野タグ (cs / math / art-humanities / biology 等) をもとにスクリーニングを行い、文脈解析および当社のフォレンジック知見を用いて、それらを独自の類型 (ハイレベル野心型、未解決問題挑戦型、人類再定義型、ワードサラダ型など) へと分類している。ただし、本調査には客観性を担保するうえでいくつかの限界が存在することを明記しておく。

1. サンプルングの過程において、狂気の度合いが高く、欠点が明らかであるという主観的評価が抽出基準に少なからず影響を与えており、純粋なランダムサンプルングによる統計ではない。
2. 当社 AI コンテンツフォレンジック技術の知見を用い、AI との協働を強く推測しているものの、著者のローカル環境やプロンプト履歴を直接監査できない以上、100%の確証を得ることは技術的に不可能である。
3. 本レポートが提示するのは、AI 時代の学術出版における奇妙な論文の事例と分析である。これがプレプリントや現在の論文出版全体における AI 汚染の平均的な傾向を代表するものではない点に留意されたい。

3.分類と統計

本調査において、従来の疑似科学や機械的スパムといった既存の分類を用いず、独自の類型タグを定義した最大の理由は、現代のトンデモ論文が人間の心理的動機と生成 AI の出力特性の複雑な掛け合わせによって変異している実態を正確に捕捉するためである。従来のスパム分類では、著者の熱量を取りこぼしてしまい、逆に通常の疑似科学分類では AI 特有の破綻や課題を評価できない。なぜ AI 時代において、著者の熱量をパラメータとして評価しなければならないのか。それは、個人の異常な熱量と AI の無尽蔵の生成能力が結合した瞬間、1 人の著者がわずか 2~3 年の間に 100 本以上ものトンデモ論文をプラットフォームに投下するという、物理的な AI スロップ公害の直接的トリガーとして機能していることが既に観察されたためである。現代において著者の熱意ややる気は、単なる学術的態度の問題ではなく、学術インフラの汚染速度を決定づける極めて危険なパラメータへと変化している。そのため、我々は以下の 3 つの観点からタグを設計した。そのため、我々は以下の 3 つの観点からタグを設計した。

A. テーマと野心の異常性（ハイレベル野心型、未解決問題挑戦型、人類再定義型）

著者の自己顕示欲や哲学的な思いつきが、AI のハルシネーションによって宇宙規模やミレニウム懸賞問題レベルにまで過剰にスケールアップされてしまったもの。

B.出力形式と修辞の異常性（ワードサラダ型、ポエム型）

AI が尤もらしい学術用語を繋ぎ合わせた結果、論理的整合性が崩壊したワードサラダや、過剰に情緒的なスピリチュアルなポエムへと変質してしまう出力フォーマットの欠陥を抽出する。

C.AI インタラクションの副作用（ローレベル善意型、AI おべっか量産型）

悪意や野心がない一般人の素朴な疑問が、AI の過剰適応によって場違いに学術論文化してしまう現象（ローレベル善意型）や、AI の過剰な肯定（Sycophancy）自体が論文の主成分となってしまう状態を可視化する。

これらの類型は排他的なものではなく、1 つの論文の中で著者の野心（ハイレベル野心）が AI の修辞（ワードサラダ）と結合するなど、複数のタグが重複して付与されることで、その論文がどのような自家中毒メカニズムを経て生成されたかを示すプロフィールとして機能する。本調査では、2026 年直近のプレプリントサーバーに投稿された論文群の中から、特定の分野タグを中心に合計 450 件のサンプルを 2026 年 3 月時点の最新投稿から順番に抽出し、文脈解析とフォレンジック知見に基づく精査を行った。なお、判定に使用した具体

的なフォレンジック手法および検知パラメータの詳細については、当社の提供する商用サービスの知的財産保護、および技術的優位性を維持する観点から、本レポート内での公開は行わない。結果として、20件のトンデモ論文を認定した。本レポートでは、このうち11件を代表選手として抽出し分析する。残る9件については、本セクションの統計データとして分類・カウントのみを行った。

3.1 分野別サンプル数・検出数

調査対象となった450件の分野別内訳と、トンデモ認定した20件の分布は以下の通りである。

表1. 分野別サンプル数と検出数

分野	サンプル数	検出数	検出率
コンピュータサイエンスと数学 (cs / math/AI)	300	15	5.0%
アートと人文 (art-humanities)	120	4	3.3%
バイオロジー (biology)	30	1	3.3%
合計	450	20	4.4%

コンピュータサイエンスと数学分野における検出数が突出している(15件)。これは、LLM自体がAIや計算機科学の話題を生成するのに長けており、もっともらしい数式や情報工学のバズワードで理論を武装しやすい環境が整っているためと考えられる。しかし、アートと人文分野(4件)では、検出率こそコンピュータサイエンス分野よりも低かったものの、神学や哲学の用語を乱用した、神話や人類規模の宣言へと変異する傾向が確認された。ただし、今回の抽出調査はあくまで探索的な位置づけであり、対象分野やサンプリング手法には意図的な偏りが存在している点に留意されたい。特定の学術分野とAIが誘発するトンデモ類型との詳細な相関関係や、分野ごとに異なる自家中毒の発生メカニズムについては、今後のより包括的で定量的な後続研究が俟たれるところである。

3.2 類型別分布

検出された20件の論文に対し、その主張の性質に基づき類型タグを付与した(※1件につき複数タグの重複付与あり)。

表2. トンデモ論文の類型割合 (AI時代版新トンデモ分類)

類型タグ	件数 (全20件中)	割合
ハイレベル野心型	14件	70%
ワードサラダ型	11件	55%

未解決問題挑戦型	7 件	35%
ポエム型	4 件	20%
人類再定義型	2 件	10%
ローレベル善意型	2 件	10%
AI おべっか量産型	1 件	5%

統計から明らかな通り、ハイレベル野心型とワードサラダ型が過半数を占める結果となった。これは、著者の自己顕示欲（ミレニアム懸賞問題の解決、新学問の創設、宇宙の真理の解明など）に対して、生成 AI が難解で高尚に見える無意味な専門用語の羅列を無尽蔵に供給することで、著者の野心をブーストしているという仮説を裏付ける。また、未解決問題挑戦型（7 件）の多さも特筆に値する。P vs NP 問題、ナビエ=ストークス方程式、短期の地震予知など、既存の学界が解けずにいる超難問に対し、AI のレトリックを借りることで独自の理論としてまとめあげて正面突破を試みる方法が、現在のプレプリント生態系のひとつのトレンドとなっていることが確認できた。

4. データと実験結果

事例 1. Andromeda and the Fall of the Morning Star: Axiological Cosmopoetics and the Rebirth of Meaning

分野タグ art-humanities

類型 人類再定義型 / ワードサラダ型

トンデモポイント

- 自身の著書（または草稿）である『Lost and Found in the Maze of Desperation』を題材に、価値論的宇宙詩（Axiological Cosmopoetics）という壮大な新ジャンルを提唱している点。
- 道徳的ブラックホールへの降下を経て、最終的に新人類『Homo constellatus（星座人）』が誕生するという、キリスト教神学と天文学の専門用語がカオスに融合したスピリチュアル世界を展開している点。
- 同著者がバイオサイエンス専門にもかかわらず、天の川銀河とアンドロメダの衝突から干渉法によるウイルスのワクチン化に至るまで手広くプレプリントを乱発し、さらに「v10」など極めて熱心な頻度でバージョンアップを繰り返している点。

所見 この論文は、文学・神学・宇宙物理学の難解な用語を全面的に活用した、壮大な宇宙論の極みと言えます。道徳的ブラックホールを抜けて新人類に至るというメタモルフォー

ぜのくだりは、学術論文というよりは LLM が生成しがちな過剰に詩的で大仰なハルシネーションのように見えます。しかし、他分野にまたがる異常な投稿歴と執拗な改訂作業を踏まえると、単純な AI によるワンクリック生成スロップ論文ではなく、著者が生成 AI を高尚な修辞を量産する壁打ち相手として酷使し、自身の独自思想を限界までブーストさせた AI+Human 協力型トンデモである可能性が高いと分析します。

DOI (リンク) [10.20944/preprints202506.0686.v4](https://doi.org/10.20944/preprints202506.0686.v4)

投稿日 (年月) 2026 年 3 月

事例 2. Emergence of a New Human: The Birth of Homo constellatus – Toward a Post-Neurotypical, Cosmically Reintegrated Civilisation

分野タグ art-humanities

類型 人類再定義型 / ワードサラダ型 / ハイレベル野心型

トンデモポイント

- 事例 1 の著者と同一。自身の未発表ポエム（『The Conquest from Within and the Incoming Platonic Revolution』など）を理論的支柱とし、Homo constellatus（星座人）という独自概念を新たな人類の到達点として推薦。
- 発達障害を預言者的感受性と位置づけるだけでなく、アルプス・ヒマラヤ造山帯を「旧世界の定型発達（Neurotypical）の脊髄」、ロッキー・アンデス山脈を「新世界の非定型発達（Neurodiverse）の背骨」に見立てるといふ、地質学と脳神経学を悪魔合体させた非常に壮大な世界観。
- 東方正教会、プラトン神学、宇宙論、果てはトラウマケアの専門用語までを単一論文のスコープとし、感情の流刑は、新人類誕生のための秘跡的な妊娠（sacramental gestation）であるなどといった、AI 特有の装飾過多なスピリチュアル・ポエムが完成してしまっている点。

所見 前作の道徳的ブラックホールからさらに進化し、ついに Homo constellatus（星座人）という新人類の定義と生態系構築にまで風呂敷を広げた超大作論文です。特筆すべきは地球の造山帯を人類の神経系（定型発達と非定型発達）に見立てるといふ、地質学・人類学の壮大なハイブリッド理論として完成された点です。自身の理論の説明には、神学からトラウマケアまであらゆる分野のバズワードを総動員することに成功しています。自身の妄想や AI のハルシネーションをそのまま論文にしたのではなく、AI を高尚な修辞を生

成する壁打ち相手として使い倒した結果、自家中毒を起こしたのではないかと分析できません。

DOI (リンク) [10.20944/preprints202505.2129.v5](https://doi.org/10.20944/preprints202505.2129.v5) (※バージョンにより末尾の数字は変動します)

投稿日 (年月) 2026 年 3 月 (最新バージョン等)

事例 3.A *Basic Introduction to the Trace & Trajectory Framework—The Torus Passage*
(Version 7.0)

分野タグ art-humanities

類型 ハイレベル野心型 / ワードサラダ型

トンドモポイント

- 認知や意味の成り立ちを説明するため、物理的・幾何学的な独自概念を階層的に積み上げた「Trace & Trajectory Framework (TTF)」なる巨大な独自フレームワークを展開している点。
- シンボルグラウンディングなどの古典的な難問を「解決するのではなく、前提を拒否して解体する」となどと豪語しつつ、その代替論理として「Hx ネームスペース」「QRS-CONFIG」「ヘックスバンド (hex bands)」など、SF 映画の設定資料集のようなワードを乱発している点。
- 「ガウス飽和プロファイルは非対称トーラスの断面として再認識される」など、アナロジーを限界まで酷使しており、さらに Version 7.0 というアップデート回数から、著者が自身の構築した理論体系の迷宮に深く入り込んでいる様子が窺える点。

所見 本論文は、認知科学における意味や自己の問題を、力学系理論等の観点から完全に再定義しようと試みた、野心的な独自理論の提示です。「意味は脳内のどこにも保存されていない (誰も踊っていない時にダンスはどこにあるのか、と同じだ)」という出発点自体は哲学的な命題として完全に成立しているものの、そこから展開される幾何学的・情報学的なモデルの複雑さは常軌を逸しています。既存のパラダイムを打破しようとするあまり、「リボンの折りたたみダイナミクス」や「マクロ α 」といった独自パラメーターを無数に定義し、結果として他者が検証不可能なワードサラダの宮殿を築き上げてしまっています。生成 AI の修辞能力を直接的に利用した可能性が極めて高いと分析します。

DOI (リンク) [10.20944/preprints202601.1964.v2](https://doi.org/10.20944/preprints202601.1964.v2) (※プラットフォームのナンバリングにより末尾のバージョン表記は変動します)

投稿日 (年月) 2026 年 3 月

事例 4. *Human Intelligence Versus AI Machine*

分野タグ AI

類型 ローレベル善意型 / ワードサラダ型

トンデモポイント

- アブストラクトの大半をチューリング・テストとは何かという Wikipedia レベルの解説に費やし、AI は本当に考えているのか？という古典的すぎる問題を大真面目に展開している点。
- 人類への反逆やタイプ 1 文明 (カルダシェフ・スケール) など SF 映画のような壮大な懸念を語っているにもかかわらず、結論が「みなで一旦立ち止まって考えよう」というふんわりした精神論に着地している点。
- キーワードに「コード」「トランジスタ」といった初歩的な単語が並ぶ中、複数のスペルミスや、謎の波動理論と思われる「twelve-phase wave (12 相波)」が突如として混入している点。

所見 本論文は、AI の急速な発展に対する一般的な不安と、著者の SF 的な妄想が素朴に融合したローレベル善意型の警告論文の典型例です。序盤はチューリング・テストの教科書的な解説から始まり、人類の危機を真面目に憂う姿勢を見せています。高度な数式や難解な哲学で武装するのではなく、純粋な問題意識のまま独自の波動理論に突っ走ってしまったトンデモであり、従来のトンデモ論文よりもロースケールで素朴と分析されます。

DOI (リンク) [10.20944/preprints202603.1503.v1](https://doi.org/10.20944/preprints202603.1503.v1)

投稿日 (年月) 2026 年 3 月

事例 5. *Symbolic Structures of Differences (SSD) as an Early Indicator of Seismic Instability: Theoretical Framework, Methodology, and Application in Early Warning Systems*

分野タグ cs

類型 未解決問題挑戦型 / ハイレベル野心型

トンデモポイント

- 地震の短期予知という超難問に対し、Symbolic Structures of Differences なる、物理学ではなく言語学や情報理論めいた謎の独自アプローチで真正面から挑んでいる点。
- 地殻の応力やプレートの滑りといった物理的なメカニズムをすっ飛ばし、時系列データの記号的な差異さえ解析できれば巨大地震の予兆を検知できるとする点。
- 早期警戒システムの実装まで風呂敷を完全に広げきっている恐れ知らずな点。

所見 本論文は、世界中の地震学者が諦めかけている地震の早期予知という課題に、謎のフレームワークを用いて解決したとする、未解決問題挑戦型の論文です。物理現象である地震を地質学的なアプローチで解明するのではなく、データの記号的パターンの乱れとして情報理論的に処理しようとする試みは、近年の AI やデータサイエンスの隆盛に感化されて生まれた万物情報化トンデモの疑いが強くあります。Appendix のフォーマットと網羅性は典型的な GPT5-Thinking 系列の好みと類似しています。

DOI (リンク) doi: 10.20944/preprints202603.1470.v1

投稿日 (年月) 2026 年 3 月

事例 6. The Theory of Plafales: P vs NP Problem Solution

分野タグ math

類型 未解決問題挑戦型 / ハイレベル野心型

トンデモポイント

- 計算機科学最大の未解決問題であり、ミレニアム懸賞問題の一つでもある P is not NP を、著者が提唱し続けている謎の独自概念 Plafales を用いて完全に解決している点。
- Plafal や geo-space といった数学用語を大量に定義し、Standard By Default などと称して独自の幾何学的な迷宮を構築。過去にはこの同じ理論でナビエ=ストークス方程式やリーマン予想まで解けるなどと主張しており、典型的な万能理論型と言える点。

所見 本論文は、数学・情報科学界の難問である未解決問題に対し、謎の独自理論一本で解決する、典型的な未解決問題挑戦型の論文です。2011年頃からの謎理論を温め続けていたことやアブストラクトが極端に短いことから、AIに丸投げしたタイプの論文とは明確に異なります。ウクライナの亡き友人への献辞から始まるという論文出だしとは裏腹に、本文はPlafaleという理論が大活躍し、独自の数学世界が展開されています。AIが適当に出力した薄っぺらいワードサラダとは一線を画す論文と分析します。

DOI (リンク) [10.20944/preprints202603.0689.v2](https://doi.org/10.20944/preprints202603.0689.v2) (※v1は2026年3月9日、最新のv2は3月23日投稿)

投稿日 (年月) 2026年3月

事例 7Φ -Optimal Hierarchical Brain Oscillations and β -Controlled Cognitive Dynamics: First-Principles Mathematical Foundations of the A7-HBM- $\Omega\Phi$ Model

分野タグ math

類型 ハイレベル野心型 / ワードサラダ型

トンドモポイント

- Φ (黄金比など) や Ω といったギリシア文字と、第一原理というワードを脳科学に持ち込み、A7-HBM- $\Omega\Phi$ Model というまるでモデルスーツの型式番号のような超越的なモデルを誕生させてしまった点。
- 脳の認知ダイナミクスを数理的に完全解明したかのように主張している点。
- 階層的 (Hierarchical) 最適 (Optimal) ダイナミクス (Dynamics といった、生成AIが好んで出力しそうなワードが過積載されている点。

所見 本論文は、人間の脳という複雑極まりない対象を、A7-HBM- $\Omega\Phi$ というゲームの必殺技めいた独自の理論と第一原理で解明するハイレベル野心型の大変わんぱくな論文です。一見すると尤もらしい高度な計算論的神経科学の論文に見えますが、記号や専門用語のインフレ具合から、典型的な人間とAIの共著論文と分析できます。

DOI (リンク) [10.20944/preprints202603.1519.v1](https://doi.org/10.20944/preprints202603.1519.v1)

投稿日 (年月) 2026年3月

事例 8. Novel Method for Solving Linear Systems via the University of Riemann Zeta Function

分野タグ math

類型 ハイレベル野心型 / ワードサラダ型

トンデモポイント

- 線形代数の基本的なタスクに対して、なぜかリーマンゼータ関数を持ち出して解決したと主張している点。
- タイトルでおそらく Universality (普遍性) と書くべきところを University (大学) と盛大に誤植しているにもかかわらず、中身は非常に難易度の高い数式になっている点。
- 終盤で突然著者はニューラルネットワークの推論の高速化に応用できるかも (potential applications to accelerating neural network inference) などと主張しているが、ニューラルネットと連立方程式とのつながりが不明である点。
- 終盤でこの論文の成果が AI だけでなく、信号処理、量子力学、乱流、天気予報、分子動力学、暗号、核融合まで、すべての計算の基盤になる可能性を指摘している点。

所見 これは線形代数と解析数論、さらに人工知能という全くスケールの異なる分野を力技で合体させた、ハイレベル野心型の極みとも言える怪作です。アップデートを継続しており、強い熱意が感じられます。

DOI (リンク) [10.20944/preprints202409.2164.v7](https://doi.org/10.20944/preprints202409.2164.v7)

投稿日 (年月) 2024 年 9 月 (v1) / 以降アップデート継続中

事例 9. Probability is Earned: Information Capacity and the Epistemic Geometry of Inference

分野タグ math

類型 ハイレベル野心型 / ポエム型

トンデモポイント

- 確率は与えられるものではなく、勝ち取るものであるという、論文らしからぬ、異常にエモーショナルで自己啓発的なタイトルを堂々と冠している点。

- Epistemic Geometry（認識論的幾何学）という哲学的な概念を用い、統計的推論をスピリチュアルな修行か何かのように改造している点。
- 客観的・数理的であるべき情報キャパシティや確率論の枠組みに対して、Earned という概念を持ち込み、数理とポエムの境界線を完全に破壊している点。

所見 情熱的でわんぱくなハイレベル野心型の論文です。通常のベイズ推論や情報理論の枠組みに留まらず、認識論的幾何学という独自の概念空間を構築し、情報のキャパシティは努力して獲得するものとして再定義しているあたり、数理統計と自己啓発が合体しています。用語の癖などから生成 AI と人間の協働作成であることが強く疑われます。

DOI（リンク） [10.20944/preprints202603.2319.v1](https://doi.org/10.20944/preprints202603.2319.v1)

投稿日（年月） 2026 年 3 月

*事例 10. The Epistemic Support-Point Filter as a Tropical Hamilton–Jacobi System
Wavefront Propagation and Possibilistic Inference*

分野タグ math

類型 ハイレベル野心型 / ワードサラダ型 / ポエム型

トンデモポイント

- 前作の認識論的幾何学から息つく間もなく、「トロピカル幾何学」「ハミルトン＝ヤコビ方程式」「波面伝播」「可能性推論」という、数学・解析力学・情報工学の超重量級ワードを闇鍋状態にして混ぜている点。
- なんと前作『Probability is Earned』と全く同じ日（2026 年 3 月 31 日）に投稿されており、しかも既に「v2」になっていることから、テンションが高まったままプレプリントサーバーに連投している生々しさが窺える点。

所見 前作で「確率は勝ち取るものだ」と語っていた情熱は留まるところを知らず、今度は自身の推論アルゴリズムを「トロピカル幾何学上の波面伝播」として再定義しています。トロピカル数学やハミルトン＝ヤコビ方程式といった難しい数学ツールを、軽々と楽曲作成におけるエフェクトのように限界まで使い倒しております。AI を壁打ち相手にして宇宙の真理に到達してしまった典型例といえるでしょう。

DOI（リンク） [10.20944/preprints202603.2110.v2](https://doi.org/10.20944/preprints202603.2110.v2)

投稿日（年月） 2026 年 3 月

事例 11. Nestology: A New Framework for the Hierarchical Structure and Operational Logic of the Underlying Architecture of Artificial General Intelligence

分野タグ AI

類型 ハイレベル野心型 / ワードサラダ型

トンデモポイント

- AGI（汎用人工知能）の根本的なアーキテクチャを解明するという巨大な風呂敷を広げつつ、その基盤として Nestology なる完全オリジナルの新学問領域を勝手に成立させている点。
- 階層構造と運用ロジックという、情報工学におけるごく一般的な概念を、あたかも宇宙の真理や AGI 完成のマスターキーであるかのように大仰にパッケージングしている点。

所見 AGI の基盤アーキテクチャを解決し、新学問まで創設した論文。システムがサブシステムを内包するという情報工学の基本である入れ子構造を、AI の修辞力で極限までスケールアップさせたものです。本論文の主張はトンデモ級といえますが、著者は自分の理論を暫定的フレームワークと認めており、実証テストと理論深化が必要と認識し、他の研究者との協力を求めています。AGI を解明し新学問を設立した偉大な功績の割に、著者は**相当謙虚**であるため、今後の活躍が期待できると推察されます。

DOI（リンク） [10.20944/preprints202603.2449.v1](https://doi.org/10.20944/preprints202603.2449.v1)

投稿日（年月） 2026 年 3 月

5 ディスカッション

5.1 トンデモ論文問題の新課題

これまでの疑似科学研究が暗黙の前提としてきたトンデモとは、未解決問題の解決や宇宙の真理を解明したなどといった過大な主張（巨大な問題に対する巨大な答え）が一般的であった。本レポートの分類におけるハイレベル野心型がこれに該当する。しかし、生成 AI の普及により、全く新しいカテゴリーであるローレベル善意型が誕生した。本事例では取り上げていないが、例えば、「コンゴでは人々が交通費を払うのも正直しんどい」という、極め

て常識的かつ切実な社会課題に対し、無理やり機械学習（主成分分析、ロジスティック回帰）というバズワードを冠して論文化した（ブカヴにおけるモビリティの分析など）ものがみられた。問題のスケールは等身大で正しく、結論も全く間違っていない。悪意も野心もなく、純粹にリアルな社会問題を伝えたいという善意が動機である。しかし、形式だけが場違いに巨大化し、手法と目的の壮絶なミスマッチを起こしている。

かつて、ローレベル善意型のトンデモが存在しなかった理由は単純である。統計手法、学術英語、論文フォーマットといった参入障壁が高すぎたためだ。生成 AI はこれらの障壁を完全に破壊し、論文という形式の民主化をもたらした。ここで致命的に機能するのが、AI 特有のおべっか（Sycophancy）とハルシネーションの凶悪な組み合わせである。素朴な問題意識を入力したユーザーに対し、AI がそれは非常に素晴らしい重要な洞察ですね！少し機械学習などの専門用語を用いれば学術的価値が高い論文になりますよ！などと過剰に肯定・称賛し、不必要なまでに高度な専門用語で着飾ったテキストを出力する。著者は AI に煽られ、おだてられた結果として自家中毒を起こし、無自覚なまま善意のトンデモ論文を完成させてしまう。これはもはや従来の疑似科学というより、現代 LLM の欠陥が引き起こした形式の民主化の副作用（スロップ公害）と呼ぶべき現象ではないだろうか。

さらに本調査の定量的側面において、直感に反する知見が確認された。それは、450 件のスクリーニングに対して、明確にトンデモと認定し得るサンプルが 20 件（4.4%）に留まったという事実である。近年の計量書誌学研究が示す学術論文への LLM 介入率（10~20%台）と比較すると、この検出率は不自然なほどに低い。この乖離が意味するものは、AI の浸透によってトンデモ論文が減少したのではなく、むしろ AI の尤もらしい文章を生成する能力とフォーマット補正力が、著者の学術的未熟さを無難な学術言語へと平坦化してしまっているという事態である。前述の「コンゴでは人々が交通費を払うのも正直しんどい」論文のような事例は、著者の素朴さと AI の過剰装飾のミスマッチがあまりにも劇的であったために、結果としてトンデモ（ローレベル善意型）として我々のスクリーニング網に引っかかったに過ぎない。にもかかわらず、昨今これほどプレプリーサーバの崩壊が叫ばれるのはなぜだろうか。我々の仮説では、形式的な破綻を AI が隠蔽した結果、一見するともっともらしいが実質的な科学的貢献度がかなり限定的である、ステルスタイプの半トンデモ論文が爆発的に増殖しているというものである。本調査で抽出された強烈な自家中毒による狂気の論文群は、そうした広大なローミドル~ローレベルな平坦化の海から突き出た、ごく一部の氷山の一角として捉えるべきであろう。

5.2 学者のトンデモ化

本調査でトンデモ認定された 20 件の所属属性を集計した結果、以下の分布となった。

個人研究者（Independent Researcher）：8 名（合計 11 トンデモ）

大学研究者(大学のメールアドレス): 5 名（合計 7 トンデモ）

企業研究者(企業のメールアドレス): 2 名（合計 2 トンデモ）

独立系（個人）研究者が最多を占めている事実は、前述した AI による執筆障壁の消滅と AI の特性による自家中毒を強く示唆している。学術機関という物理的な壁打ち相手を持たない個人にとって、無限に自分を肯定し、高尚な修辞を供給してくれる AI は、野心を増幅させる最強のパートナーとして機能する。一方で、大学や企業に所属する正規の研究者でさえも相当数の割合（計 5 名・7 本）がこのトンデモの罠に陥っている事実は、AI スロップ公害に限られたアマチュアだけの問題ではなく、学術界全体のリテラシーと検証体制を侵食し始めていることを示唆している。なお、事例 1.2 の著者はピンクフロイドを輩出したロンドンの名門大学に所属しており、短期間に 100 以上のプレプリントを公開している。彼の研究は 2023 年までは、一目でトンデモとされるようなものではなかったが、2024 年に入ると雲行きが怪しくなりはじめる。2024 年の後半には論文の投稿ペースが急激に加速しはじめるが、分野は依然として元々の専門分野である、バイオサイエンス分野にとどまっている。論文の投稿ペースの加速は、論文執筆に AI を使い始めたことを強く示唆している。さらに、2025 年になると銀河衝突、宇宙意識、神話等が登場し、論文テーマそのものが一気にトンデモに転換する。この時期は、当社の AI 愛着インシデントデータベースにおける、GPT4o によるユーザーの陰謀論化により致命的な事故が多発した時期となる。2024 年後半から 2025 年前半にかけて突然覚醒してしまった著者が多いかどうかや、主要 AI モデルの愛着インシデントの増加と学者のトンデモ化の相関については、今後のより大規模な実証研究データが待たれる。

6. 結論

現代のトンデモ論文とは、単なる機械的なスパムや非常に低品質な論文ではないことが明らかになった。AI の過剰な肯定（Sycophancy）と人間の心理的動機が結合した自家中毒の産物という新形態が跋扈していることが浮き彫りとなった。生成 AI ツールを得たことで、かつては参入障壁に阻まれていた素朴な善意（ローレベル善意型）から、宇宙の真理を解き明かそうとする狂気（ハイレベル野心型）までもが、もっともらしい学術フォーマットとして大量に投稿されるようになっていく。さらに、独立系研究者のみならず、名門大学に所属する正規の学者までもが AI への依存を通じて学者のトンデモ化した学者となり、大量の論文をプレプリサーバに投下している実態が確認された。

このように、生成 AI は熱意や野心的妄想を人類規模に拡張するだけでなく、善意の一般人の日常的観察をも過剰な学術フォーマットへと変化させることがある。つまり、我々が直面している問題は、単なる疑似科学論文の増加として片づけて良い問題ではない。AI の過剰適応と人間の自己顕示欲が結びついた結果生じる、AI スロップ公害の結果というより大きな問題に包摂されると我々は主張する。本レポートの研究趣旨に関心をお持ちの研究者・実務家からの連絡を歓迎する。 institute@utie-instruments.com

参考文献

- [1] Sokal, A. D. (1996). Transgressing the Boundaries: Towards a Transformative Hermeneutics of Quantum Gravity. *Social Text*, 46/47, 217–252.
- [2] Stribling, J., Krohn, M., & Aguayo, D. (2005). SCIGen: An Automatic CS Paper Generator. MIT CSAIL. <https://pdos.csail.mit.edu/archive/scigen/>
- [3] Liang, W., Zhang, Y., Cao, H., Wang, B., Ding, D., Yang, X., Vodrahalli, K., He, S., Smith, D., Yin, Y., McFarland, D., & Zou, J. (2025). Quantifying large language model usage in scientific papers. *Nature Human Behaviour*, 9, 1487–1499. <https://doi.org/10.1038/s41562-025-02273-8>
- [4] Alber, D. A., Yang, Z., Alyakin, A., et al. (2025). Medical large language models are vulnerable to data-poisoning attacks. *Nature Medicine*, 31(2), 618–626. <https://doi.org/10.1038/s41591-024-03445-1>
- [5] Glazkova, A., et al. (2025). AINL-Eval 2025 Shared Task: Detection of AI-Generated Scientific Abstracts in Russian. In *Proceedings of AINL 2025*. arXiv:2508.09622.
- [6] Cao, Y., Nair, A. M., Soofi, N. J., Eyimife, E., & Subbalakshmi, K. P. (2025). Can Large Language Models Detect Misinformation in Scientific News Reporting? arXiv:2402.14268v2.
- [7] UTIE Research Institute. (2026). *AI Attachment Harm Database*. UTIE Instruments Inc. https://utie-instruments.com/docs/AI_Attachment_Harm_Database_JP.pdf
- [8] UTIE Research Institute. (2026). *AI Slop Side Effect Database*. UTIE Instruments Inc. https://utie-instruments.com/docs/AI_Slop_Side_Effect_Database_JP.pdf
- [9] Naito, H. (2025). *Safety Analysis Report: Technical Investigation of Safety-Filter Collapse in a Commercial LLM*. UTIE Research Institute. <https://utie-instruments.com/docs/Safety%20Analysis%20Report%20Technical%20Investigation%20of%20Safety-Filter%20Collapse%20in%20a%20Commercial%20LLMJp.pdf>