

研究ライフサイクルを支えるAI： 現在地と将来展望

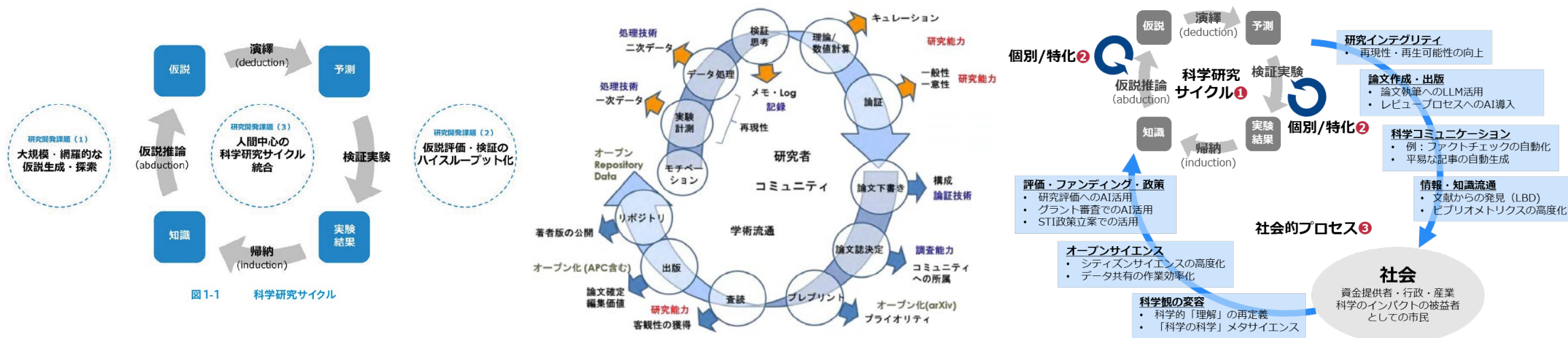
相澤 彰子

国立情報学研究所

コンテンツ科学系 教授／大規模言語モデル研究開発センター 副センター長

研究のライフサイクルとAI

「科学」は何重もの多様なサイクルから構成されている



CRDS戦略プロポーザル (2021) 図1-1 人工知能と科学～AI・データ駆動型科学による発見と理解～「科学研究サイクル」

学術会議提言 (2020) <付録図表3> オープンサイエンスの深化と推進に向けて「研究のライフサイクルとオープン化」

JST CRDS報告書「人工知能研究の新潮流 2025～基盤モデル・生成AIのインパクトと課題～」(2025年3月)

あらゆるところでAIがかかわってくる

人間 vs AI

AI

vs

人間

人間 with AI

vs

人間 without AI

人間 with AI

vs

AI with 人間

本日のトピック

PART 1: 言語系AIとは？

PART 2: 研究者を支援するAI

PART 3: 研究者と協力するAI

PART 1: 言語系AIとは？

AI

vs

人間

(ここでは) AI ≈ 「大規模言語モデル」

大量の コーパス

膨大な量のテキスト。10兆単語超えも多い。人間がひとりで読むと10万年以上かかるとされる。

巨大な パラメタ集合

10億から数兆個。「コーパス」を参照して、パラメタ(=変数)の値を最適化する。

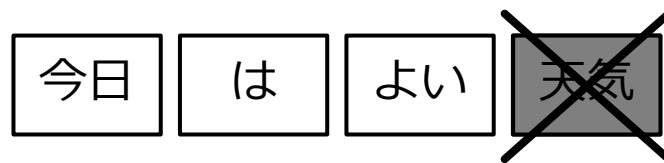
「次に来る語の予測」という 単純なタスク

原理は極めてシンプルだが、なるべく正確に予測するための訓練に膨大な計算が必要。

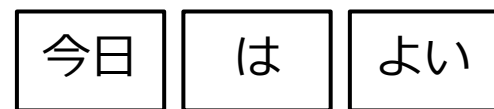
大規模言語モデルの仕組み

大規模言語モデル
(10億～数兆パラメタ)

大量のコーパス



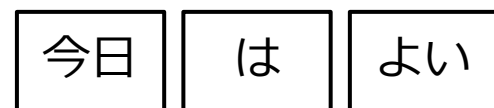
「訓練」



- 天気 10%
- 気分 35%
- 顔 5%
- ...

実際にコーパス中に出現した語の予測確率が高くなる方向にパラメタを調整

「推論」



天気

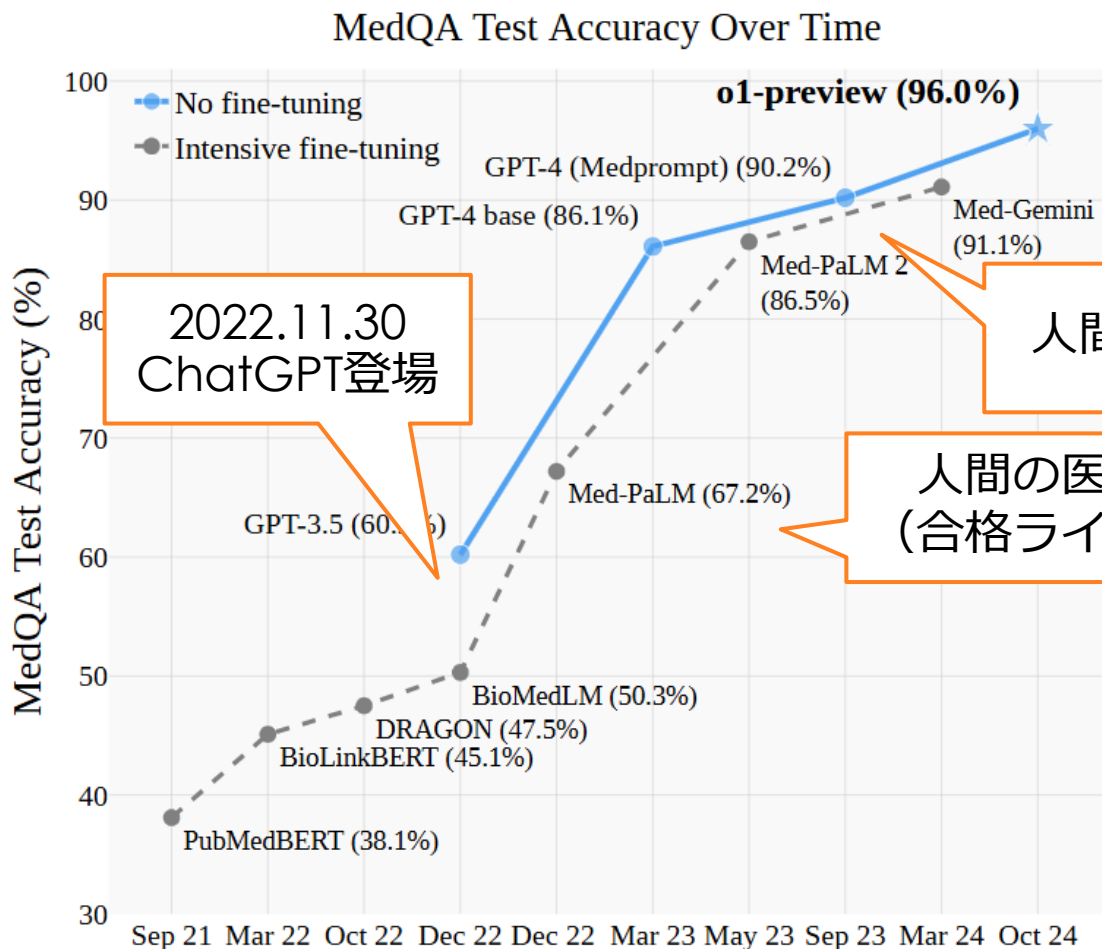
もっとも予測確率が高い語を出力

ユーザプロンプト

AIレスポンス



“MedQA”にみるAI性能の推移



MedQA (医学知識を測るための、米国医師国家試験形式のベンチマーク) に対する正解率の推移

正解率の目安

対象	正解率
人間の医師 (合格ライン)	~60%
人間の専門医 (Expert)	~87%
GPT-4	81% ~ 86%
Med-PaLM 2	86.5%
OpenAI o1 / GPT-5クラス	94% ~ 96.5%

評価用ベンチマーク“HLE”にみるAI性能

Humanity's Last Exam (HLE)

大学院～研究者レベルの幅広い専門分野
選択式・短答式、一部は画像を含む

HLEによるAI性能の推移

<https://agi.safe.ai/>



gemini3, gpt-5で
も正答率30%台

テスト問題の例



Ecology

Question:

Hummingbirds within Apodiformes uniquely have a bilaterally paired oval bone, a sesamoid embedded in the caudolateral portion of the expanded, cruciate aponeurosis of insertion of m. depressor caudae. How many paired tendons are supported by this sesamoid bone? Answer with a number.

Edward V

Massachusetts Institute of Technology

(翻訳) ハチドリ目
(Apodiformes) のハチドリ類は、尾骨下筋 (m. depressor caudae) の付着部である拡大された十字靭帯の後外側部に埋め込まれた種子骨 (sesamoid) として、左右対称の楕円形骨を唯一有する。この種子骨によって支持される対になった腱はいくつあるか？ 数字で答えよ。

専門的、相当むずかしい

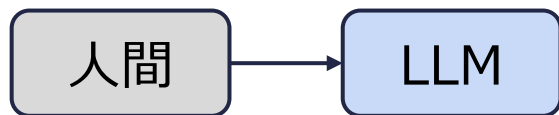
Center for AI Safety, Scale AI, and HLE Contributors Consortium. 2026. "A Benchmark of Expert-Level Academic Questions to Assess AI Capabilities." Nature 649 (8099): 1139–46.

AIエージェント

AIがツールを使いながらタスクを遂行

≈ 大規模言語モデル + ツール

通常のLLM利用



応答 (1回で終了)

AIエージェント



対話・確認を挟みながらタスクを遂行

ツールの例：

- Web検索 (最新情報の取得)
- 論文データベース検索 (CiNii、PubMed等)
- 計算・統計処理
- ファイルの読み書き
- コード実行

→ LLMが状況に応じて選択

PART 2: 研究者を支援するAI

～AIツールによる論文の執筆や査読の
支援

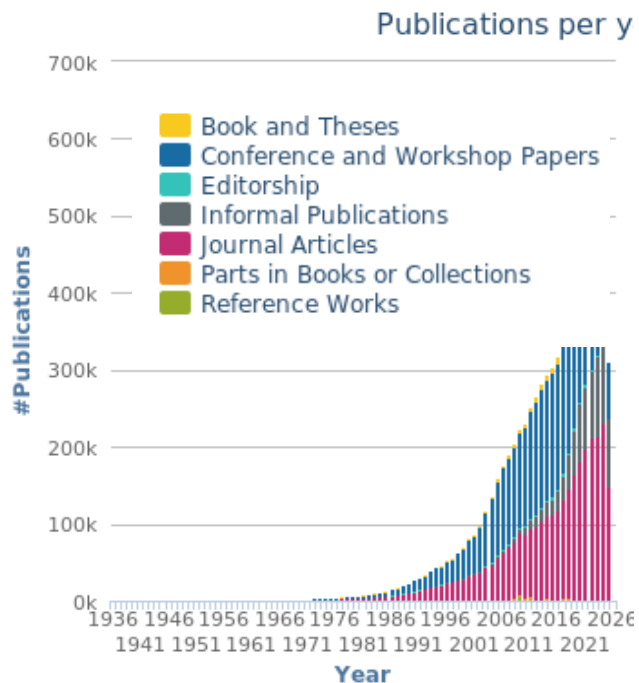
人間 with AI

vs

人間 without AI

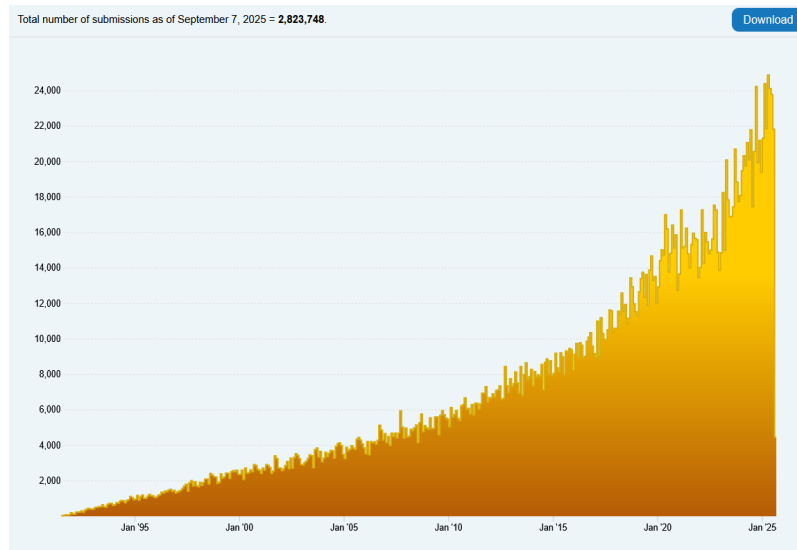
論文数の急激な増加

<https://dblp.org/statistics/img/publicationsperyear.png>



DBLP

https://arxiv.org/stats/monthly_submissions



ArXiv

<https://www.ouvrirlascience.fr/excessive-growth-in-the-number-of-scientific-publications/>

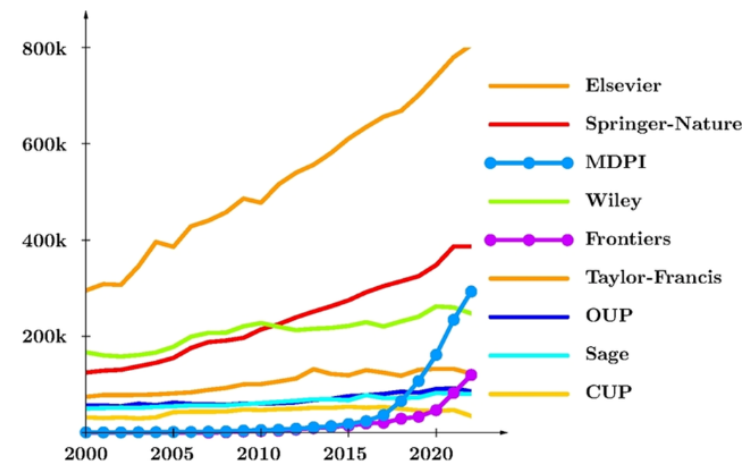


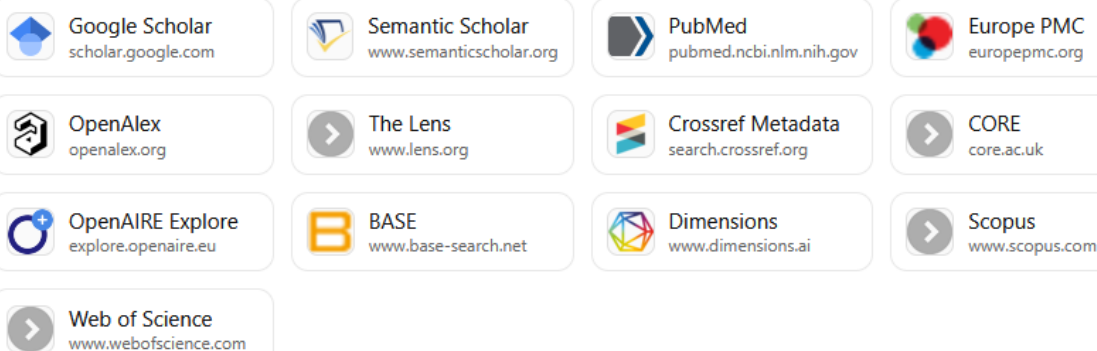
Figure 1: Evolution of the total number of articles published per year and per publisher, data from OpenAlex.

OpenAlex

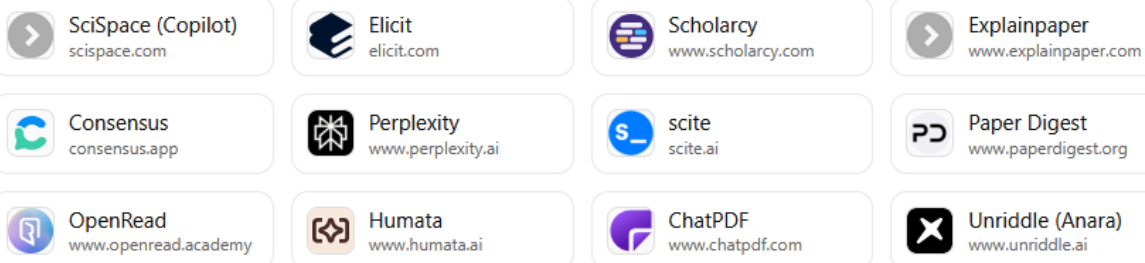
論文執筆を支援する様々なAIツール

ChatGPT 5 pro 出力 (2025.9.7)

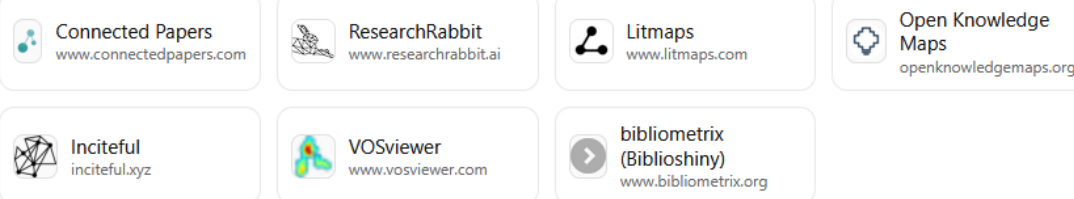
① 総合・学術検索エンジン



② AI 要約・QA・読解 (PDF含む)



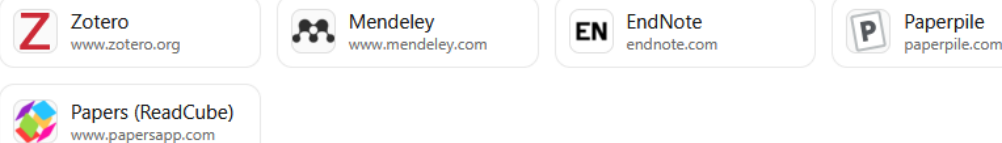
③ レコメンド/可視化



④ システマティックレビュー支援



⑤ 参考文献管理 (リファレンス管理)



論文執筆を支援する様々なAIツール

ChatGPT 5 pro 出力（2026.2.12）

ワークフロー（クリックで絞り込み）

ステップを選ぶと、下のツール一覧がその工程に絞り込まれます。

ワークフローの各ステップとツール:

- ステップ 1:** テーマ探索・研究質問 (8 ツール)
- ステップ 2:** 計画・設計・プロトコル (7 ツール)
- ステップ 3:** 文献探索 (一次探索→深掘り) (8 ツール)
- ステップ 4:** 全文入手・文献ライブラリ化 (8 ツール)
- ステップ 5:** 選別・抽出 (特にSR) (7 ツール)
- ステップ 6:** 読解・注釈・知識化 (8 ツール)
- ステップ 7:** 解析・図表・再現性 (8 ツール)
- ステップ 10:** 開・共有 (17 ツール)

中央のテキスト:

多様化
高機能化
ワークフロー化

おすすめ構成（例）

おすすめ構成（例）:

- 無料中心・最小構成 (6 tools):** Google Scholar, Zotero, Obsidian, Quarto, GitHub, Zenodo
- LaTeX共同執筆 (理工系定番) (4 tools):** Overleaf, Zotero, GitHub, iThenticate
- Word/Docs共同執筆 (共同著者多め) (5 tools):** Microsoft Word, Google Docs, Zotero, Writefull, Grammarly
- システムアイトツクリビュー/メタ分析 (6 tools):** PubMed, Rayyan, ASReview, Elicit, PRISMA, EPPI-Reviewer
- 図表を一段上に (Graphical Abstract含む) (4 tools):** GraphPad Prism, BioRender, Inkscape, Mind the Graph

AI生成論文に対するポリシー

2022～

▶ 多くの出版社や学会が、AI生成論文に対するポリシーを策定

2024. “Publishers’ and Journals’ Instructions to Authors on Use of Generative Artificial Intelligence in Academic and Scientific Publishing: Bibliometric Analysis.” BMJ (Clinical Research Ed.) 384 (2024, January).

2024.1に出版された報告（上記）によると、著名学術誌トップ100の87%が生成AIの利用に関するrecommendationを作成



「生成AIを使った場合に明記」、「生成AIが著者になることを禁止」
は、多くの出版社や学会で一致

AI生成論文の影響

▶ 2023年以降の論文中で使われる“Science English”の急速な変化

- ▶ 例 : aligns, emphasizing, underscores, boasts, showcasing, surpasses, delve, ...
 - ▶ Tom S Juzek and Zina B. Ward. Why Does ChatGPT “Delve” So Much? Exploring the Sources of Lexical Overrepresentation in Large Language Models. COLING-2025.

▶ 参考文献のハルシネーションは難関会議でも

- ▶ 自然言語処理分野における国際会議の採択論文について、参考文献の誤りが年を追うごとに増加との報告
 - ▶ “HalluCitation Matters: Revealing the Impact of Hallucinated References with 300 Hallucinated Papers in ACL Conferences” <https://arxiv.org/abs/2601.18724v1>
 - ▶ Citation Check Tool である GPTZero が著名国際会議NeurIPSの採録済み51論文に100のハルシネーションを発見と報告
 - ▶ <https://gptzero.me/news/neurips/>

▶ ハルシネーションは参考文献だけではない

- ▶ 論文の数だけではなく、1論文当たりの誤りの数も増えていることを指摘
 - ▶ To Err Is Human: Systematic Quantification of Errors in Published AI Papers via LLM Analysis <https://doi.org/10.48550/arXiv.2512.05925>

論文査読システムは危機的状况に

Our most recent check indicated that you **have not committed to review / AC** a sufficient number of papers for this cycle. We regret to inform you that if you do not satisfy this requirement, **we will have to desk reject your submission.**

◀ 査読をノルマ数こなさないと
デスクリジェクト

査読者が足りなくて採択通知が出
せないで、投稿論文あたり1名
査読者を出して下さい

We regret to inform you that **the review process is currently delayed due to a shortage of reviewers** with sufficient research experience. If we are unable to recruit additional qualified reviewers promptly, further delays in the review process—and consequently in the acceptance notifications—are expected. **Nominate one reviewer from among the authors of your submission**

Reviewers or chairs deemed highly irresponsible such as **missing deadlines without warning**, violating guidelines on LLM use and professional tone, extremely terse reviews, **may be barred from committing their work to <conf_name>** and (re-)submitting to the next <submission> cycle

◀ 予告なしに査読期限を遅れたら、
投稿とりけし、次回も投稿不可

査読者のAI利用

査読者のAI利用に対する研究者の不安

公平性 → AIの査読品質に関する懸念

責任所在 → 人間が責任を持ってチェックしないことに関する懸念

機密性 → 未公表の論文をAIにアップロードすることに関する懸念

たとえば、Natureによる研究者へのアンケートでは、**研究者はAI論文執筆よりAI自動査読に、より強い懸念**を持っている (Diana

Kwon: Is it OK for AI to write science papers? Nature survey shows researchers are split. Nature, Springer Nature. May 14, 2025. <https://www.nature.com/articles/d41586-025-01463-8.pdf>)

AI査読に関するトピック

ICLR = 機械学習分野の
トップ会議、査読を公開
している

▶ 「AIが支援する査読」の現状分析

- ▶ AIの痕跡がみられる査読が増加（2023と比べて2024年ではICLRにおける推定で15.8%）
- ▶ （分析ではAI支援査読のスコアは甘めと報告されている。ただしAI-Assisted=AIが査読したとは限らないことに注意）
 - ▶ Latona, Giuseppe Russo, Manoel Horta Ribeiro, Tim R. Davidson, Veniamin Veselovsky, and Robert West. "The AI Review Lottery: Widespread AI-Assisted Peer Reviews Boost Paper Scores and Acceptance Rates." <http://arxiv.org/abs/2405.02150>.

▶ AI査読の試行導入

- ▶ AAI 2026 reviewにおける予備的試行
 - ▶ 第一段査読で人間の査読者に参考情報を提供（あくまで支援、決定は人間の査読者であることを強調）
 - ▶ <https://aaai.org/aaai-launches-ai-powered-peer-review-assessment-system/>
- ▶ ICML 2026における試行
 - ▶ 著者は、「査読にAIを用いるのは不可」「AIを補助的に用いてもよい」いずれかのポリシーを選択できる。ただし、査読者のAI利用を望まない著者は、自分が査読するときもAIを使ってはならない。
 - ▶ <https://icml.cc/Conferences/2026/LLM-Policy>

人間の査読に関するトピック

▶ 「査読工場」 (Review mill)

- ▶ 内容とは無関係な定型的な文言で効率的に査読を生成

- ▶ Oviedo-García, M. Ángeles. 2024. "The Review Mills, Not Just (Self-)Plagiarism in Review Reports, but a Step Further." *Scientometrics* 129 (9): 5805–13.

▶ 組織的な科学不正の可能性

- ▶ 論文工場 (paper mill)、ブローカー、ハゲタカジャーナルなどによる組織的な科学不正が急増しているとの指摘

- ▶ Richardson, Reese A. K., Spencer S. Hong, Jennifer A. Byrne, Thomas Stoeger, and Luís A. Nunes Amaral. 2025. "The Entities Enabling Scientific Fraud at Scale Are Large, Resilient, and Growing Rapidly." *Proceedings of the National Academy of Sciences of the United States of America* 122 (32): e2420092122.

▶ 査読負担の不公平性

- ▶ 投稿者と査読者の比率が地域等によって不均衡

- ▶ Adam, David. 2025. "The Peer-Review Crisis: How to Fix an Overloaded System." *Nature* 644 (8075): 24–27.

近年の「AI査読者」

- ▶ **よいAI査読者は論文の質を高めるための強力な武器**
- ▶ **AI査読者は研究を支える基盤（研究・分析も盛んに行われている）**

“Stanford Agentic Reviewer”（2025.11） <https://paperreview.ai/>

- 論文を改善するためのコメント出力を重視
- 複数観点から査読

査読コメント例

実験評価に関する評価（7つの観点のうちの1つ）

- ✓ 2クラスのみでのSCITAB評価の問題点： 2クラス分類のみを報告しており、現実的な検証において重要な NEI（Not Enough Information）クラスが除外されている。この選択については正当化が必要であり、3クラス評価の結果、または NEI事例をどのように扱う／除外しているのかの分析を併せて示すべきである。
- ✓ 報告数値の内部不整合： ベースラインを上回ったという主張と表1の数値との不一致、また4.2節の主要指標と4.4節／アブレーション結果との乖離など、内部的に整合しない数値が見られる。これにより、各コンポーネントの貢献を正確に評価することが.....

PART 3: 研究者と協力するAI

～仮説-検証のサイクルを実行する
AIサイエンティスト

人間 with AI

vs

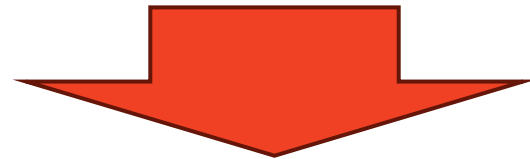
AI with 人間

リサーチタスクの自動化

▶ リサーチタスク

- ▶ アイディア生成、アイディア評価、実験計画、実験実行、結果分析、図表作成、関連研究俯瞰、考察、論文執筆、校正、...

▶ あらゆる場面でAIの活用が可能に



AIサイエンティスト

研究者向けのダッシュボード例

研究基盤としての取組み例

AI4EOSC

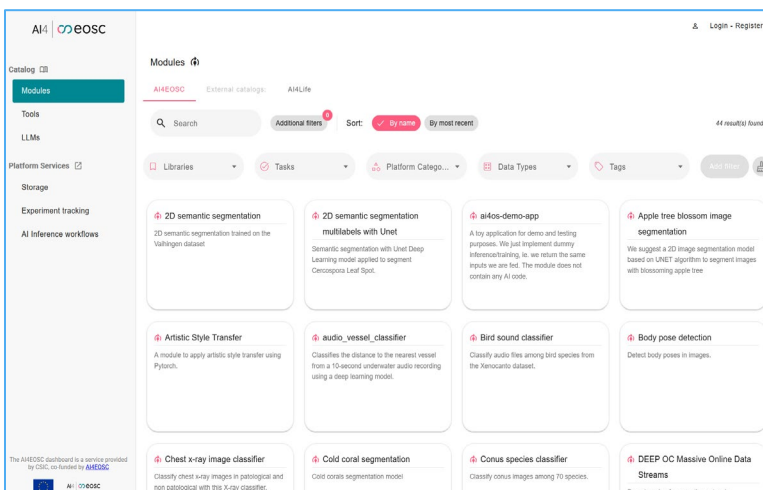
Artificial Intelligence for the European Open Science Cloud

<https://eosc.eu/horizon-europe-projects/ai4eosc/>

Countries:

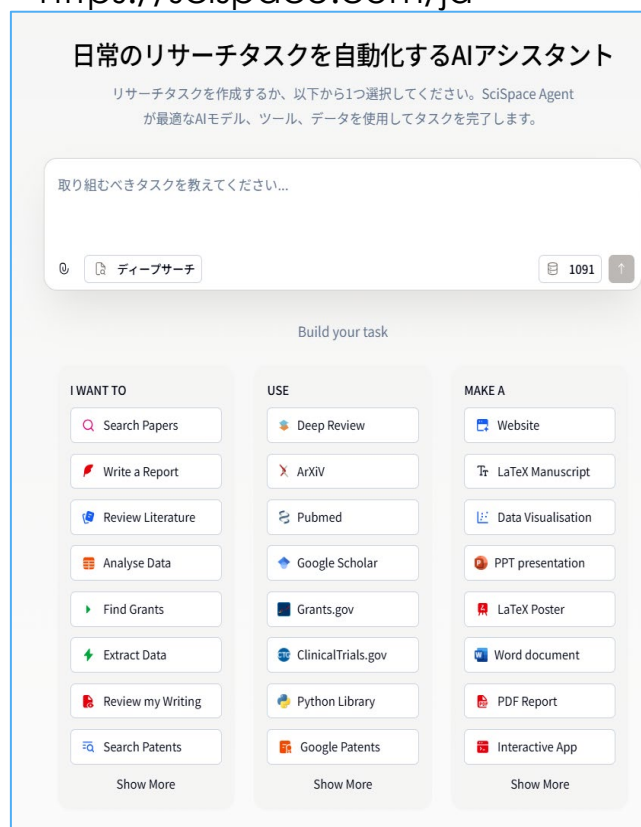
 Spain  Germany  Poland  Slovakia  Italy  Portugal

<https://dashboard.cloud.ai4eosc.eu/>



研究に特化した商用サービスの例

<https://scispace.com/ja>



出版社によるサービスの例

<https://natureresearchassistant.com/>
研究特化の "responsible AI tool"

オープンなツールの例

Maestro : <https://murtazanadir.github.io/maestro/>
自分でホストして複雑な研究タスクもカスタマイズ可能

商用AIサービスのProモード

ChatGPT, Gemini, ...
例：論文PDFをアップロード、追加実験の提案からコード生成や簡単な結果報告まで

研究者向けの支援ツールのポイント

▶ コンテンツの信頼性

- ▶ EOSCならEUのデータクラウド基盤
- ▶ 出版社系のツールは査読済の高品質科学論文を利用

▶ AIの高い能力

- ▶ 豊富な科学的知識
- ▶ コーディング、数学能力、推論機能
- ▶ 外部ツールを呼び出すエージェント機能

▶ 使いやすさ

- ▶ ユーザインタフェース
- ▶ 安心して使える

AIサイエンティストの登場

- ▶ **(2024.8) SakanaAI: AI Scientist**
 - ▶ 研究のアイデア生成、実験、論文執筆、評価のサイクルを自動化
 - ▶ コスト圧縮（高並列化）が可能であることを示した
- ▶ **(2025.4) SakanaAI: AI Scientist 2**
 - ▶ より洗練された自動化へ
 - ▶ ICLR 2025ワークショップで初のAI生成論文採択を達成
- ▶ **(2025.4) Autoscience Institute: Carl**
 - ▶ アイディア生成、実験、論文執筆を人間の研究者との協働で実現
 - ▶ ICLR 2025ワークショップで4本中3本採択（採択率75%）
- ▶ **(2025.5) IntologyAI: Zochi**
 - ▶ ACL 2025本会議で初のAI単独論文採択（採択率約20%）

最近のAI事情①

その後desk rejectされ、
会議での正式の採録には
なっていない

▶ AIツール”Zochi” によるAI生成論文 (2025.5)

- ▶ 自然言語処理分野の最高峰の会議であるACLにおいて、査読段階では採録レベルと判断

"TEMPEST: Multi-Turn Jailbreaking of Large Language Models with Tree Search" on arxiv

◀ 論文本文 8 ページの画像イメージ

<https://www.intology.ai/blog/zochi-acl>

人間が与えたのは **“novel jailbreaking methods”** という指示だけ。

From this starting point, Zochi **independently identified the research direction of multi-turn attacks⁴, formalized the Tempest method, implemented and tested it, conducted all experiments, and wrote/presented the paper** (excluding figures and minor formatting fixes made by humans).

最近のAI事情②

- <https://agents4science.stanford.edu/>
- <https://youtu.be/7pXqAeedqOo>

▶ Agents4Science 2025 (2025.10)

- ▶ AIエージェントが第一著者および査読者となるはじめてのオープンな会議形式のイベント
- ▶ ⇒ AI時代の科学研究のあり方を探る

“ほとんどすべての学術誌や学会は、AIを著者として認めることを禁止している。このような慣行によって、**研究者はAIの貢献を隠したり、最小限にとどめたりするよう動機づけられている**。この禁止は、将来の科学研究においてAIがどのように参加していくのかを理解し、形づくる私たちの能力（の醸成）を妨げている。”

“Exploring the Use of AI Authors and Reviewers at Agents4Science.” arXiv [Cs.AI]. arXiv.
<https://doi.org/10.48550/arXiv.2511.15534>.

Agents4Science 2025参加者の立場から感想

①科学的発見におけるAIの可能性と限界

- ▶ 「人間の貢献をなるべく減らす」という制約以外は、やっていることは、日ごろの研究活動におけるAI利用と大きな違いはなかった（AIスキルアップには役立つ）
- ▶ 「自律的なAI研究者」が価値を発揮するためには潤沢な計算リソースが必要（いかにコストダウンして研究を並列化できるか？）

②規範の確立

- ▶ AIを使ったかどうかより、「間違えない」「十分な情報を提供する」などの品質保証が重要（ハルシネーションは沢山ある）
- ▶ 強力なAI査読について、自前のものがない限り、アイデアをサーバにアップロードするか、優れたAI抜きで頑張るかの選択）

③AIの関与の開示

- ▶ 誠実にAIの関与を示そうとすると、記述が煩雑になり、結局それを理解するのにAIの助けが必要となる（たとえば膨大なログの解析など）

現状で問題は山積み...

適用できるトピックが限定されている (サイバーのみ)

コード生成の失敗やコーディングコストの高さ、危険なコードの生成

虚偽の報告を含む
さまざまなハルシネーション

学術的な価値の低いドラフトの
大量生産

それでも「扉の向こう」が覗ける

AI for Scienceの動きの中で

https://www.mext.go.jp/content/20251006-mxt_jyohoka01-000045188_01.pdf

AI4Sの考慮点：分野によるバーチャル操作/フィジカル操作度合い

JST CRDSの資料から引用

主分類	分野	サブ分野	バーチャル操作	フィジカル操作	■V/P■レシオ
自然科学	物理学	古典物理、量子力学、素粒子、宇宙物理	理論モデリング、シミュレーション、データ解析	機器操作、実験計測、試料調整	
	化学	物理化学、有機化学、無機化学、材料化学	分子モデリング、反応予測、データ解析	合成、計測解析、物性評価	
	生物学	分子生物学、細胞生物学、遺伝学、生態学	システムモデリング、データプロセッシング	培養実験、顕微鏡計測、フィールドワーク	
	地球科学	地質学、計測学、海洋科学	環境モデリング、システムシミュレーション	フィールド調査、試料解析、モニタリング	
形式科学	数学	純粋数学、応用数学、統計学、数理科学	理論導出、数値解析、モデリング	データ収集、検証、デモ	
	計算機科学	コンピューター科学、ソフトウェア工学、サイバーセキュリティ	アルゴリズム開発、システム設計、ソフトウェア・プログラミング	ハードウェアテスト、システム開発、メンテナンス	
応用科学	工学	機械工学、電子工学、化学工学	設計モデリング、シミュレーション、最適化	製造、試験、システムインテグレーション	
	医学	臨床医学、生体医工学、薬学	イメージング、データ解析、トリートメントプランニング	臨床試験、ラボ試験、患者ケア	
	農学	作物栽培学、園芸学、畜産学	成長モデリング、システムシミュレーション、データ解析	圃場実験、ブリーディング、耕作	
人文・社会科学	社会科学	経済学、社会学、政治学	データ解析、行動モデリング、シミュレーション	フィールド調査、行動調査	
	人文科学	哲学、史学、文学、芸術	デジタル解析、アーカイバル・プロセッシング	フィールド調査、工芸物解析、造形	
学際科学	バイオインフォマティクス	システム生物学、計算生物学	コンピューター解析、モデリング、予測	実験バリデーション、データ収集	
	認知科学	神経科学、心理学	認知モデリング、データ解析、シミュレーション	脳イメージング、行動実験	
	環境学	環境科学、持続可能性、気象学	環境モデリング、環境影響評価	フィールドモニタリング、サンプリング	
	ナノテク	ナノエレクトロニクス、ナノバイオ工学	ナノデバイスシミュレーション、プロセスモデリング	デバイス製造、材料合成・加工、センシング	

サイエンスの
多様性

V:バーチャル
P:フィジカル

Googleによる3つの論文発表

【査読機能】 ScholarPeer: A Context-Aware Multi-Agent Framework for Automated Peer Review (2026/1/30)

- ▶ <https://arxiv.org/pdf/2601.22638> 複数のAIエージェントが協調して論文査読を実行
- ▶ QA生成、抜けている比較対象手法の指摘、関連文献の調査、新規性や健全性チェック、など

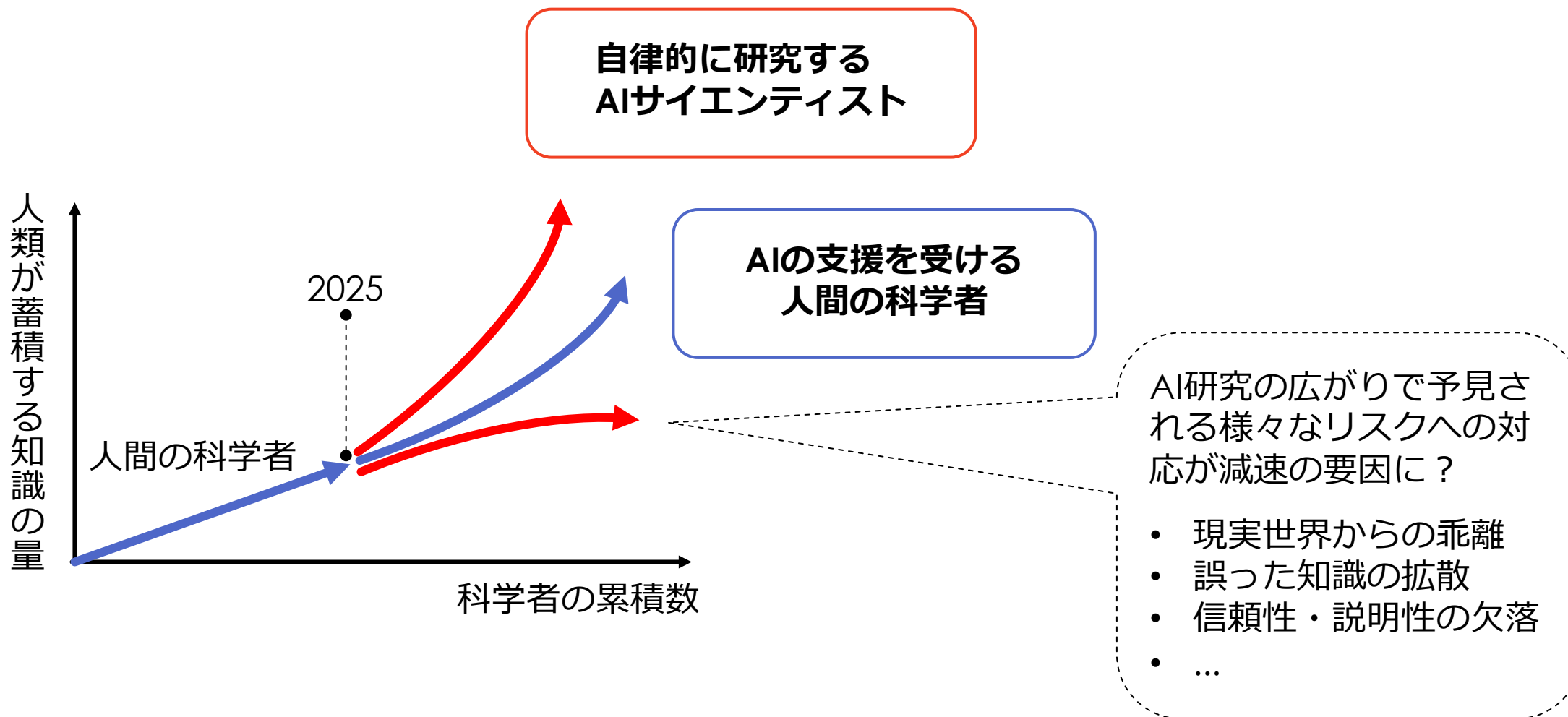
【描画機能】 PaperBanana: Automating Academic Illustration for AI Scientists (2026/1/30)

- ▶ <https://arxiv.org/abs/2601.23265> <https://dwzhu-pku.github.io/PaperBanana/> (概要：作図例あり)
- ▶ <https://github.com/dwzhu-pku/PaperBanana> (データセットとコード公開予定)
- ▶ 正確性や視認性、デザインの美しさにおいて優れた学術的な図版を生成する視覚言語モデル

【研究AIエージェント】 MARS: Modular Agent with Reflective Search for Automated AI Research (2026/2/2)

- ▶ <https://arxiv.org/abs/2602.02660>
- ▶ <https://github.com/jfc43/MARS>
- ▶ 研究を「最適化されたソフトウェアレポジトリの探索問題」として捉えて、自律的なAI研究のために最適化されたフレームワークを設計

AIサイエンティストが示唆する方向性



ご清聴ありがとうございました