

2025 年度第 1 回 J-STAGE セミナー
報告書

令和 8 年 2 月 6 日 発行

国立研究開発法人科学技術振興機構 情報基盤事業部

1 開催概要

1.1 趣旨

2025 年度の J-STAGE セミナーは、年間テーマを「学術情報流通のワークフローにおける生成 AI 活用の展望」と題し、学術情報流通の分野における生成 AI 活用の現状および展望についての情報を提供した。

第 1 回となる本セミナーは、国際 STM 出版社協会（International Association of STM Publishers: STM）と共催で実施し、サブテーマを「Utilizing Generative AI for the Advancement of Academic Journal／学術ジャーナルの改善に資する生成系 AI の活用」として、生成 AI に関する海外出版社のガイダンス、研究ワークフローへの生成 AI 導入、日本における基盤モデル開発の最前線の状況等について紹介が行われた。

1.2 開催日時、場所

日時：2025 年 11 月 4 日（火）15:00～17:00

開催形態：Zoom によるオンラインセミナー

1.3 出席者の数および所属

セミナー当日は途中出退席含め 219 人が出席した。出席者の所属が把握できたのは事前予約の上出席した 59 人であり、うち最も多かったのは大学 22 人、次いで公共機関・団体 12 人、印刷会社・出版社が 7 人、研究機関の 6 人であった。他の主な所属は、図書館、企業、その他であった。

1.4 プログラム（時刻は計画時のもの）

15:00～15:05 開会挨拶（JST）

15:05～15:15 「Overview of STM Guidance on the Use of AI in Publishing Research Outputs」

－Caroline Sutton 氏（STM）

15:15～15:35 講演 1 「論文執筆における生成 AI の実践的活用とその限界」

－笠井 大 氏（千葉大学大学院）

15:35～15:55 講演 2 「The Use of AI in Publisher Workflows: Benefits and Risks Involved in Using AI Solutions」

－Tony Alves 氏（Highwire Press）、Joshua Routh 氏（MPS）

15:55～16:00 休憩

16:00～16:20 講演 3 「Harnessing Generative AI for Actionable Insights and Societal Impact」

－柿田 佳子氏（Elsevier）

- 16:20～16:40 講演 4「信頼される日本語医療 LLM を目指して：透明性のある構築プロセスの重要性」
ー小林 和馬 氏（国立がん研究センター研究所）
- 16:40～16:55 ディスカッションと Q&A
- 16:55～17:00 閉会挨拶（JST）

2 講演概要

2.1



【Overview of STM Guidance on the Use of AI in Publishing Research Outputs】

Caroline Sutton 氏
(STM)

ありがとうございます。再び日本に戻り、今年も JST の皆さまと一緒に、このように私たちにとって非常に重要でタイムリーなテーマを扱うウェビナーに参加できることを、大変うれしく思っております。まずは、例年と同様に、本日まで登壇いただき講演者の皆さまに改めて感謝を申し上げます。また、金子理事、久保田調査役をはじめ、本日のイベント準備にご尽力いただいた JST の皆さまにも、深く御礼申し上げます。そしてもちろん、本日のためにお時間を割いてご参加くださった皆さまにも、心より感謝申し上げます。

スライドを共有しながらお話ししてまいります。本日は、私たちが AI に関して進めている取り組みと、特に出版における AI 利用の指針について、いくつか情報をご紹介します。

STM では毎年、12 月にロンドンで、将来を見据えた先進的な視点を持つ方々にお集まりいただき、年次の検討会を開催しています。今年も開催も間近に迫っています。この取り組みでは、「5 年後に学術出版に最も大きな影響を与えるトレンドは何か」というテーマに焦点を当てています。2017 年には、私たちがフューチャー・ラボと呼んでいるグループが集まり、2022 年にどのような大きな潮流が訪れるのかを議論しました (図 1)。その際に予測されたのが、AI でした。ご存じのとおり、この予測は的中しました。2022 年には ChatGPT が広く知られる存在となり、以来、私たちは AI について語り続けています。

2023 年、STM は AI に関連する 4 つのワークストリームを立ち上げました。ご覧のとおり、私たちは多くの国や法域において、政策面での取り組みを進めています。また、AI コミュニティそのものを対象とした活動も行っています。さらに、出版社では研究公正を支援するツールの開発も進められており、AI の影響で増加している偽論文や偽データ、画像操作といった問題に対処できるよう、出版社がワークフローの中で適切なベストプラクティスを導入し、実践できるよう支援しています。

本日は、この 4 つ目のワークストリームについてお話しいたします (図 2)。このワークストリームは、研究コミュニティやその他の関係者を支援することに重点を置いたものです。その取り組みの一環として、論文原稿の作成における AI の活用や、研究プロセスの中

で AI を使用した際にどのように申告すべきかといった点について、ベストプラクティスを定義する作業も含まれています。

これらのベストプラクティスについて詳しくご説明する前に、まず AI に関する私たちの一般的な見解について、少し触れておきます (図 3)。STM として、また STM 出版社として AI を考えるにあたり、私たちは新しい技術と共に進化しながらも、活動の根幹にあるもの、守るべき原則を常に意識する必要があると考えています。第一に、研究とは累積的であり、自己修正的なシステムであるということです。新しい知識は既存の研究成果のうえに積み重なっていきますし、ときには新たな情報によって、それまで真実とされていたものが覆されることもあります。

第二に、帰属は極めて重要です。それは、引用のプラクティスが研究者に正当な評価を与えるための手段であるだけでなく、引用や帰属を明示することで、知識がどのように受け継がれ、どのように発展してきたのかという系譜をたどることができるからです。こうした知識の発展の節目を把握することは、新しい研究成果が既存の知識に基づいて構築されていく過程において、その妥当性を検証するために非常に重要です。

また、検証も極めて重要な要素です。これは、皆さまのような出版社が取り組んでおられる活動そのものであり、研究コミュニティが信頼できる確かな知識の基盤の上に新たな研究を築くことができるようにするため、そして社会が研究という制度そのものに信頼を寄せるために、不可欠な役割を果たしています。

そして最後に、グローバルな対話の重要性です。私たちは、知識が、学術文献や学術的知見との対話や関与、そして理論やモデルをめぐる議論や検証を通じて発展していくことを理解しています。

また、本題に入る前に、もう一点お伝えしておきたいことがあります。STM のメンバーは、AI について考えるうえで、非常に特有の立場にあります。というのも、STM のメンバーは、大規模言語モデルや AI が学習に用いるコンテンツを出版しているだけでなく、自ら AI ツールを開発している場合もあるからです。そのため、私たちは AI について、バランスの取れたあり方や最適な発展の方向を、共に真剣に考えていく必要があると強く認識しています。STM の使命は「信頼される研究を発展させること」です (図 4)。私たちは、この使命というレンズを通してあらゆる物事を捉えており、AI についても例外ではありません。

私たちのビジョンは、真実性・責任性・信頼性を備えた AI システムを、大規模に実現していくことです。その実現に向けて、私たち自身および STM のメンバーは、新規および既

存の関係者と協力しながら、この取り組みを世界規模で進めていきたいと考えています。そして私たちの考えでは、協働は不可欠です。AI システムが科学の理念を尊重する形で進化していくためには、協働が不可欠であり、また著作権やライセンスは、そのような敬意あるプラクティスを担保するための重要な手段であると考えています。

それでは、先ほどお話ししたとおり、私たちの原則のいくつかについてご説明いたします。STM は 2021 年に、AI に関する最初のホワイトペーパーを発表しました (図 5)。これは「学術コミュニケーションにおける AI 倫理」に焦点を当てたもので、後ほど改めて触れます。また、2023 年には、ChatGPT の登場を受けて、著者による生成 AI の利用に関する初期的な指針を示す第二のレポートも公表しました。そして、この 2023 年のレポートは、現在さらに次の段階へと進んでおり、その点についても、この後ご紹介いたします。

さて、2022 年に共有した「学術コミュニケーションにおける AI 倫理」についてご説明します (図 6)。スライド右上の白い部分にご注目ください。これらの基本原則は、私たちが AI と科学について考えるうえで、今もなお中核を成すものだと考えています。私たちは透明性と説明責任を重視しています。さらに、品質と研究の健全性についても考えなければなりません。さらに、プライバシーとセキュリティを尊重すること、そして公平性についても配慮する必要があります。加えて、これらのツールを活用しながら前進していくにあたり、持続可能な発展についても考えていく必要があります。

先ほど、研究者や著者による生成 AI の利用に関する指針を示した 2023 年のレポートについて触れましたが、私たちはその取り組みをさらに一歩進めております。その成果として、今年、ほんの数か月前のことですが、学術論文の作成における AI 利用の分類に関する提言をまとめたレポートを公表しました (図 7)。これは、出版社を支援することを目的として作成されたものです。レポートそのものは非常にコンパクトですが、背景説明が含まれており、その一部が現在のスライドに示されています。

ただし、内容としては基本的な表になっています (図 8、図 9)。ご覧いただいているとおり、4 つの列には全部で 9 つの項目があり、まず 1 列目では、原稿作成における AI の利用、つまり AI をどのように使用するかというユースケースを示しています。2 列目には、その活動内容の説明があります。続く列では、その活動の具体的な例を示すとともに、含まれない内容、すなわち対象外となるものについても明記しています。項目数は 9 つだけです。現在、この資料は英語版のみとなっていますが、今後、日本語に翻訳される可能性もあります。こちらのリンクから STM のウェブサイトでご覧いただけます (図 10)。また、ご希望の方のために、このリンクをチャット欄で共有いたします。

それでは、モデレーターにお返しします。改めて、本日ご参加くださった皆さま、そしてすべての登壇者の方々に心より感謝申し上げます。画面共有を終了します。

ありがとうございました。

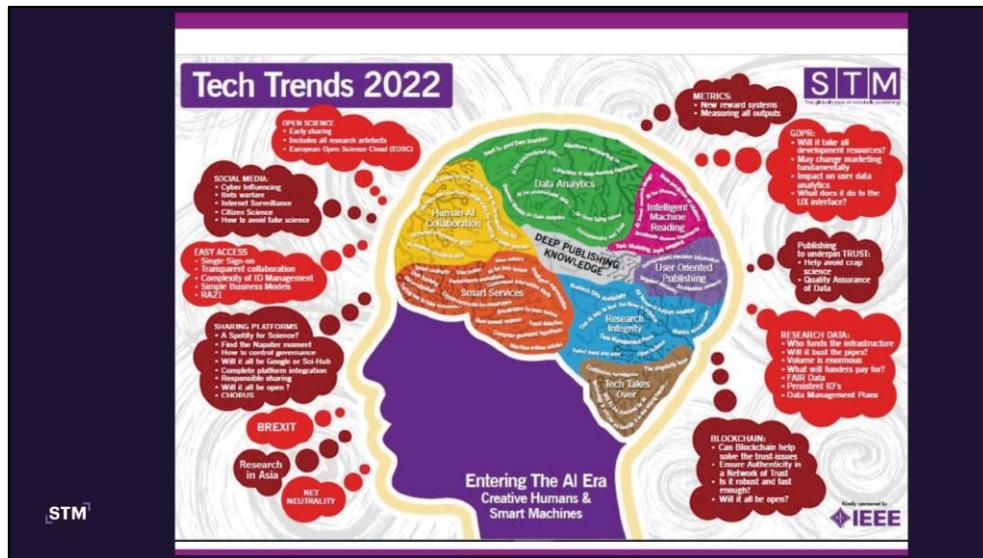


図 1

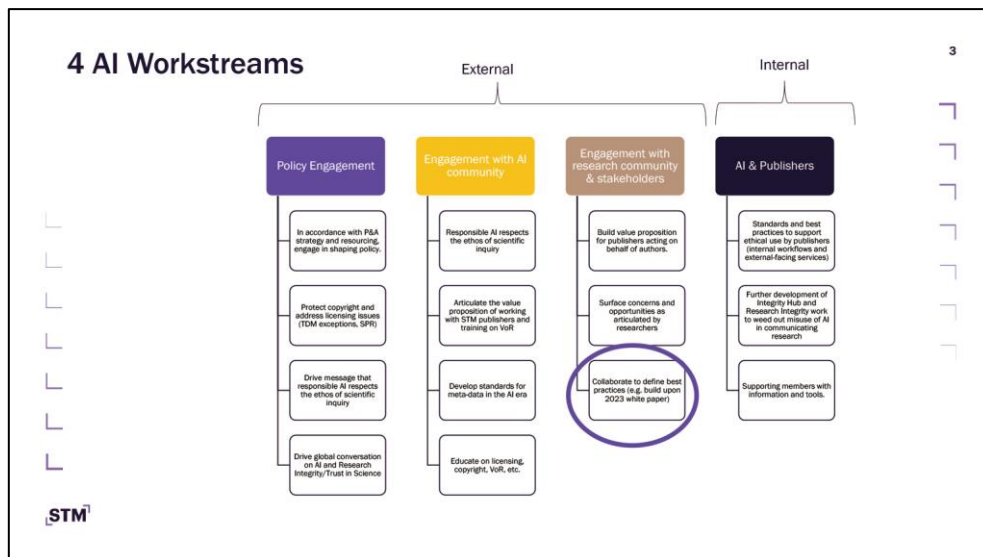


図 2

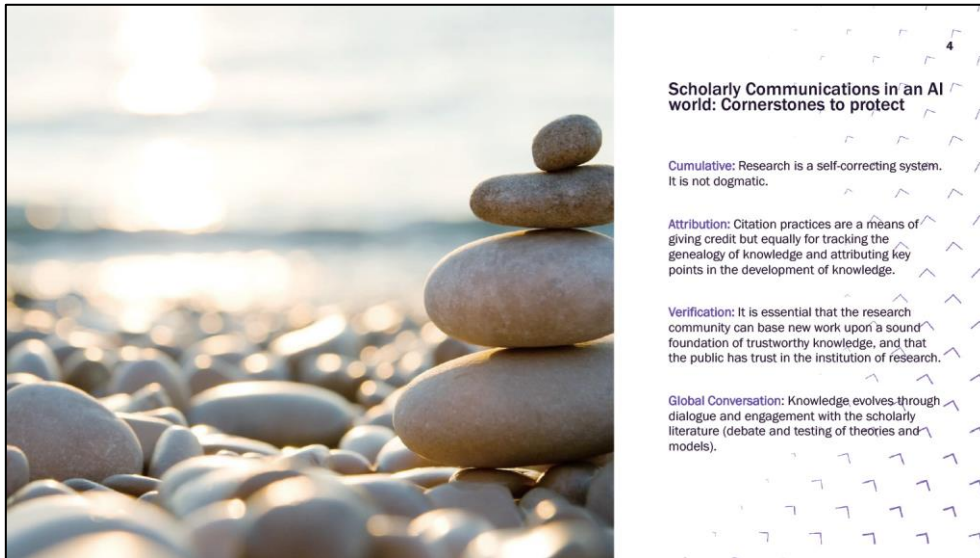


図 3

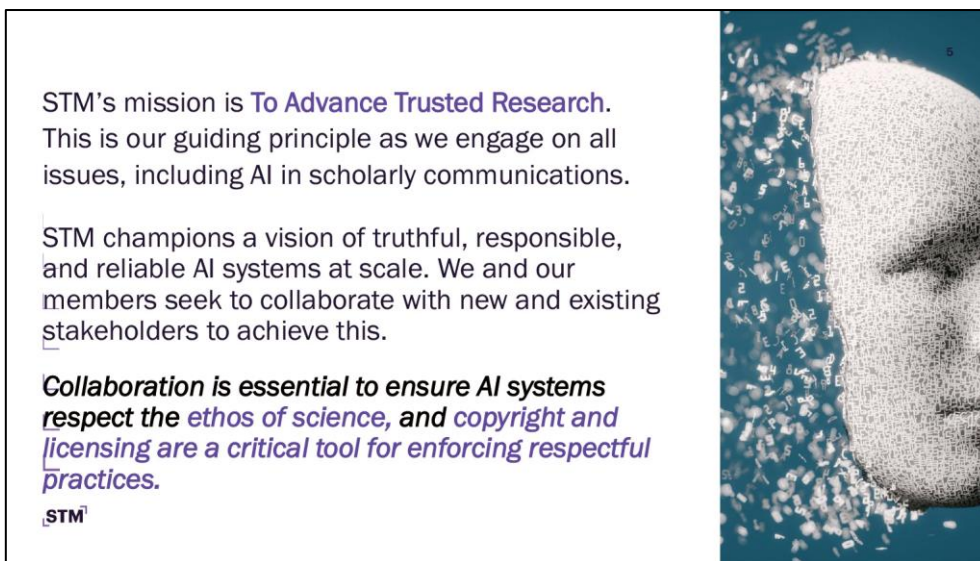
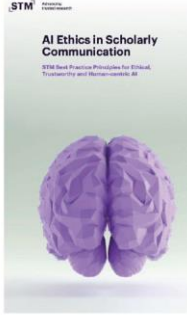


図 4

STM White papers – supporting the publishing and researcher communities



AI Ethics in Scholarly Communication
STM Best Practice Principles for Ethical, Transparent and Responsible AI

Uses of GenAI by Authors

Key Indicator	Permitted—disclosure not necessary	Disclosure necessary—permission by editorial teams	Not permitted
Basic author support tool (proof, correct, edit, and format text and documents)	✔		
Uses transcending basic author support tool		✔	
Create, alter, or manipulate original research data and results			✘
Credit GenAI as an author of a published work			✘

6

Key principles: transparency and accountability, quality and integrity, privacy and security, fairness, sustainable development.




図 5

AI Ethics in Scholarly Communication



Publishers are involved with AI in many ways, including:



1/ Content, context & data providers



2/ Supporting internal workflows and services



3/ External-facing tools and services

Full report:
https://www.stm-assoc.org/2021_05_11_STM_AI_White_Paper_April2021.pdf

STM Advancing trusted research

STM best practice principles for ethical and trustworthy AI

Sustainable development
AI systems are ideally positioned to address areas of global concern like climate change. Funding and other incentives for suppliers of high-quality input data can help maintain the vital supply of actionable knowledge.

Transparency and Accountability
Community driven standards of transparency and accountability should be applied to the data used in AI input and training, as well as in the use of AI technology in publisher's tools, processes and services. Publishers encourage working with other actors to adapt standards where appropriate.

Fairness
To avoid bias, discrimination and the suppression of novel ideas, data selection and the application of AI must be carefully analyzed, planned, reviewed, and continuously monitored. Feedback mechanisms should be developed to report and address bias.

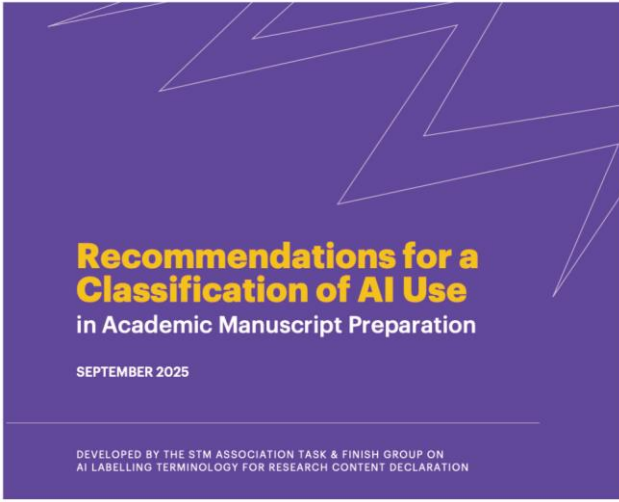
Quality and Integrity
Quality and integrity should be at the heart of the AI lifecycle, from the design and building of algorithms, to inputs used to train AI tools and services, to the application of AI. An appropriate IP framework is essential.

Privacy and Security
Principles that focus on data protection, data privacy and security can and should be used to respect and uphold privacy rights, data protection and ensure the security of datasets used in training or operating AI systems.

Legal and policy framework
The further development of AI should be guided by and grounded in clear legal standards and sound ethical principles. AI policy should respect intellectual property and continue to incentivize investment in high-quality content, datasets, and curated databases that can be used in AI applications. Any AI-enabling policy framework should foster the development of community-based standards and, where possible, build on existing initiatives. Due to the fast changing nature of the AI landscape, inflexible legislative tools should be avoided.

図 6

- Determine whether each activity is permissible for authors to use when preparing manuscripts
- Determine which permitted AI activities must be transparently declared during the submission process (e.g. to editors and peer reviewers).
- Determine which permitted AI activities must be declared in the content of manuscripts, to be included in the final publication and visible to readers.
- Provide clear policies and guidance to authors regarding the use and declaration of AI assistance in preparing manuscripts.



STM

☒ 7

STM RECOMMENDATIONS FOR A CLASSIFICATION OF AI USE IN ACADEMIC MANUSCRIPT PREPARATION 9

9 Recommended Classifications of AI Activities

Use of AI in the preparation of academic manuscripts (*)	Description of the activity	Examples of the activity	Activity does NOT include
1 Refinement, correction, editing or formatting the manuscript to improve clarity of language (**)	Machine tools were used to suggest language improvements within the manuscript	Using spell checkers, grammar checkers, and similar tools (such as Microsoft 365's inbuilt editing tools) to refine text written primarily by humans	Using AI tools to generate text from prompts or generate summaries of text; using AI tools to analyse or summarise textual documents as part of the research process
2 Writing or drafting manuscript content	AI tools were used to generate part or all of the manuscript text	Using AI tools to generate text from prompts; using AI tools to significantly expand on or rewrite text; using AI tools to generate machine summaries of text (e.g. to summarise arguments made in another publication)	Use of simple spelling and grammar checkers; analysing or summarising textual documents as part of the research process
3 Translation of manuscript text for the purpose of publishing	AI tools were used to assist translation of an author's original work into a secondary language for inclusion in the manuscript	Using AI tools such as Google Translate or ChatGPT to assist with translating a manuscript draft	Translation of materials (such as source documents) as part of the research process, unrelated to manuscript preparation
4 Refining or formatting of data reported in the manuscript	AI tools were used to assist with refinement of the presentation of data reported in the manuscript	Using AI tools to improve clarity or readability (e.g. on the level of language) or format research data submitted as part of the manuscript or its associated materials (e.g. in supplementary materials or appropriate repositories)	Using AI tools to visualise data sets (activity 6) or data manipulation; any generation, correcting, or editing of data used as part of the research process
5 Generation, refinement, correction, editing or formatting of images, diagrams or other figures for illustrative purposes only	AI tools were used to generate images, diagrams, or other figures in the manuscript for illustrative or aesthetic purposes only	Using AI tools to create an image of e.g. people farming in an ancient society – the sort of image that might otherwise be created by a human artist as an illustration	Visualisation of actual data or research outputs; editing of visualisations of actual data or results; presentation of generated images as research outputs in themselves or as representing research outputs

STM

☒ 8

10

STM RECOMMENDATIONS FOR A CLASSIFICATION OF AI USE IN ACADEMIC MANUSCRIPT PREPARATION

9 Recommended Classifications of AI Activities

Use of AI in the preparation of academic manuscripts (*)	Description of the activity	Examples of the activity	Activity does NOT include
6 Generation, refinement, correction, editing or formatting of visualisations of research data or results	AI tools were used to visualise or refine visualisations of research data/results in the manuscript	Using AI tools to generate graphs, tables, or other visualisations of research datasets	Use of traditional statistical software (such as STATA, SAS, and R) to visualise data; generation of images or other content without any basis in real research data/outputs
7 Refinement or formatting of code reported in the submitted manuscript	AI tools were used to assist with refinement of the presentation of code used in the research process and reported in the manuscript	Using AI tools to improve the readability and clarity of code submitted as part of the manuscript or any associated materials (e.g. in supplementary materials or appropriate repositories), without altering its functionality	Using AI tools to generate new code for use in research processes, or alter the functionality of code used in research processes before manuscript submission
8 Assisting with gathering references	AI tools were used to suggest references to include in the manuscript's reference list	Using AI tools to find and identify articles that are likely to be related to the research reported in the manuscript	Using automated tools (such as BibTeX) to generate formatted citations for references identified by human authors; using AI tools to generate plausible-sounding references that do not exist; using AI tools to identify publications that may be useful to the research process
9 Presentation of any kind of content generated by AI tools as though it were original research data/results from non-machine sources (***)	AI tools were used to create data, text, images, graphs, spectra, or other content that is presented as though it were original research data/results collected or analysed from other non-machine sources	Any content generated wholesale by AI tools without any basis in original research data/outputs, presented as though it is based on original data/results	Using AI tools to visualise original research data/results (activity 6); carrying out research on the outputs of generative AI (in which case outputs generated by the AI being studied can be considered research data)

(*) We reiterate that use of AI in other stages of research processes is outside the scope of this classification. This classification considers only activities that use AI tools to assist with the preparation of manuscripts.

(**) 2023 STM Guidelines recommend that disclosure is not necessary for this use of automated tools.

(***) 2023 STM Guidelines recommend prohibiting this use of AI tools.

7

図 9

11

<https://stm-assoc.org/document/recommendations-for-a-classification-of-ai-use-in-academic-manuscript-preparation/>

STM

図 10



【The Use of AI in Publisher Workflows: Benefits and Risks Involved in Using AI Solutions】

Tony Alves 氏 (Highwire Press)

Joshua Routh 氏 (MPS)

出版社のワークフローにおける AI の活用について発表する機会を与えてくださった JST と STM に感謝申し上げます。トニー・アルヴェスと申します。HighWire Press にて製品部門の上級副社長を務めております。HighWire は MPS Limited のプラットフォーム部門です。本日は、人工知能技術を用いたツールやサービスを用いて学術出版のワークフロー上のプロセスを実行する際に生じる利点とリスクについてお話しいたします。

まず、全体的な利点とリスクの概要をお話しし、その後、具体的な内容をさらに詳しく掘り下げたいと思います (図 11)。私たちは一般的に、AI 活用には 4 つの大きな利点があると認識しています。第一に、効率性とスピードの向上です。AI は、コピー編集、翻訳、XML 構造化といった反復的な作業を自動化し、人間の編集者が判断や品質管理といったより高度な業務に集中できるようにします。第二に、品質と一貫性の向上です。自然言語処理や機械学習モデルにより、大量のコンテンツに対しても、コピー編集の均質化、正確なメタデータ付与、編集基準の一貫した適用が可能になります。第三に、アクセシビリティとインクルージョンの強化です。AI による代替テキスト (Alt-Text) 生成や、多様性・公平性・包括性 (DEI) チェックといったツールは、研究コンテンツをあらゆる読者にとってより利用しやすいものにするとともに、公平性とバイアスへの配慮を伴ったコミュニケーションを可能にします。そして第四に、研究公正の強化です。画像フォレンジック、AI コンテンツ検出、科学的妥当性の検証ツールなどを用いることで、改ざんや剽窃、AI によって生成された文章を出版前に特定し、研究の真正性を守ることが可能になります。

一方で、リスクも存在します。私が特に重要と考える主なリスクは、次の 4 点です。第一に、データのプライバシーとセキュリティです。サードパーティの大規模言語モデルや外部 API を統合することは、厳格なプロトコルの下で運用されなければ、原稿の漏えいやデータの不正利用といったリスクを招く可能性があります。第二に、バイアスとハルシネーションです。先ほど述べたように、AI モデルは学習データに含まれるバイアスを反映、あるいは増幅してしまふことがあります。また、もっともらしいが不正確な結果 (ハルシネーション) を生成することがあるため、多様なデータセットを用いた継続的な検証が必要です。第三に、透明性と説明責任の喪失です。AI が下す判断は、編集上の根拠を見えにくくする可能性があります。監査証跡や説明可能性が確保されていなければ、編集プロセスに対する信頼が損なわれてしまいます。第四に、自動化への過度の依存です。AI が人間の判断を補助するのではなく置き換えてしまうと、出版社は編集上の厳密さの低下、倫理的監督の弱体化、さら

には学術記録全体の信頼性の低下というリスクに直面することになります。まとめると、AI にはスピード、品質、包摂性、そして研究公正を向上させる可能性があります。しかし、それが実現するのは、プライバシー、バイアス、透明性、そして人間による監督とのバランスが取られている場合に限られます。

ここからは、AI が出版社のワークフローをどのように変革しているのか、AI がもたらす測定可能な利点、そして適切なガバナンスや監督なしに使用された場合に生じる現実的なリスクについて掘り下げていきます。MPS Limited および HighWire Press では、過去数年間にわたり、原稿作成や査読から、制作、ホスティング、出版後の分析に至るまで、出版ライフサイクル全体に人工知能を組み込んできました。当社の研究開発拠点である MPS Labs では、現在、十数件を超える AI を活用したプロジェクトを運用または試験導入しています(図 12)。これらのツールは抽象的な概念ではありません。当社のワークフロー・プラットフォームである DigiCore Pro や、ホスティング・プラットフォームである HighWire Hosting の中で稼働する統合型マイクロサービスとして実装されており、当社の各種サービスに組み込まれているほか、他のサードパーティや外部システム向けに API 経由でも提供されています。では、これらの技術がどのような形で具体的な価値を提供しているのか、そしてどの点において注意が必要なのかを見ていきましょう。

AI の最も目に見える利点は、大規模な効率化を実現し、反復的な手作業を削減しながら、成果物の一貫性と品質を向上させる点にあります(図 13)。当社の自動言語評価およびガイド付き編集システムでは、自然言語処理(NLP)と自然言語理解(NLU)技術を活用し、文法や表現、包摂性、内容の不一致といったエラーを検出します。これらのシステムはすでに本番環境で完全に稼働しており、人手によるコピー編集作業を 25~30%削減しています。また、当社の機械学習を用いたコンテンツ構造化ツールは、非構造化された原稿を自動的に JATS XML 形式に変換し、1 論文あたり数時間の作業時間を削減するとともに、書式の不整合を解消します。このツールを試験導入したプロジェクトでは、ワークフロー全体を 20~30%短縮する効果が確認されています。

編集ワークフロー以外の領域においても、AI は効率性を向上させています。当社の AI 搭載チャットボットは、オンラインツールを利用する編集者や著者を支援しており、現在ではサポート問い合わせの 70~85%を処理しています。顧客満足度は 92%に達し、平均初回応答時間は 18 分まで短縮されました。これにより、人的サポートはより複雑な案件に集中できるようになり、さらにタイムゾーンを超えた 24 時間体制のサポートが可能になっています。その利点は、AI によって人が機械的な修正ではなく、編集上の判断に専念できるようになる点です。一方で、適切な管理がなされないと、AI による提案が微妙な言語上の誤り

やバイアスを拡散してしまうリスクもあります。そのため、すべてのモデルには人によるレビューと分野専門家による検証プロセスが組み込まれています。

出版までのスピード向上は、すべての出版社に共通する目標です。AI ツールは、査読や制作プロセスの中でも特に時間を要する工程、すなわち適切な専門家を見つける作業を効率化します。これらのツールは、PubMed、CrossRef、OpenAlex といったデータベースを活用し、原稿の内容に適した編集者や査読者とマッチングします。また、専門性を自動的に検証し、利益相反の有無を確認するとともに、査読可能な作業負荷も考慮します。その結果、従来は数日を要していた割り当て作業が、現在では数時間で完了するようになっています。

当社の AI 翻訳パイプラインは、中国語やスペイン語を含む 6 つの主要言語に対応しています。このシステムは、英語を母語としない著者、特に中国や中南米の若手研究者が公平に研究活動へ参加できるよう支援するものです。その利点は、査読プロセスの効率化や、多言語出版の推進につながる点が挙げられます。一方で、AI 翻訳に過度に依存すると、専門的な意味内容が歪められるリスクもあります。そのため、すべての翻訳は出版前に必ず人間の編集者によるレビューを受けています (図 14)。

当社の Alt-Text 生成ツールは、複数の大規模言語モデルを組み合わせたマルチモーダル手法を用いて、画像をその文脈に沿って記述します。キャプションや周辺テキストを抽出したうえで、アクセシビリティ基準に準拠した説明文を生成します。大手出版社との試験運用では、精度は 85~90% に達しており、非常に高い費用対効果を示しています。

DEI コンプライアンスツールは、大規模言語モデルとコンピュータビジョンを活用し、バイアスを含む表現、不均衡な描写、そして固定観念的な画像表現を検出します。その後、分野専門家によるレビューに続いて品質保証を行う 2 段階の人によるレビュー工程により、編集上の責任が確実に担保されます。このように、利点としては、これらのツールが、拡大しつつある DEIA およびアクセシビリティに関する要件への対応を出版社にとって容易にする点が挙げられます。一方で、リスクとしては、モデルの学習データを通じてバイアスが入り込む可能性があることです。そのため当社では、厳選された DEI 用語辞書を使用するとともに、専門家による継続的な検証を行っています (図 15)。

当社の AI 画像操作検出ツールは、ペーパーミルによる投稿でしばしば見られる操作である、図表内のスプライシング、クローニング、レタッチを特定します。このツールは、大規模な環境においてリアルタイムでの検証を提供し、編集者がフラグ付けされた画像に集中できるようにします。内部検証の結果、手動によるチェックと比較して、誤検出が 40% 減少したことが示されています。さらに、当社では人間が書いたテキストと AI によって生成

されたテキストとを区別することにも取り組んでいます。このツールは、捏造された研究や機械生成された研究を防ぐセーフガードとしても機能します。これらの研究の健全性を守る取り組みの一環として、MPS Labs は、数式が妥当であるかどうか、あるいはグラフにアクセシビリティ用の軸ラベルが欠けていないかをチェックする科学的検証エンジンの開発も進めています。これは実質的に、研究データに対する AI による Sanity check (サニティーチェック) です。利点として、こうした公正性ツールが、不正行為、盗用、誤情報から研究を保護する点にあります。一方で、いかなる AI モデルも、人間の倫理的判断に取って代わることはできません。そのため、フラグが立てられたすべての案件は、対応が取られる前に人間の専門家によってレビューされます (図 16)。

タクソノミー付与は、コンテンツの発見可能性を高めるうえで極めて重要な工程です。しかし、手作業による負担が大きいため、しばしば十分に行われていません。当社のマルチラベル分類器は、タクソノミー用語を自動的に付与し、数千本に及ぶ論文に対して一貫したタグ付けを確保します (図 17)。

現在開発中なのが、自然言語によるディスカバリー (探索) ツールです。このツールは、すでに一部のクライアントで試験運用されており、ユーザーが会話形式で質問を行い、参照情報を含む要約回答を得られる仕組みになっています。これと関連するセマンティック検索システムでは、OpenAlex および MeSH データを基に構築されたナレッジグラフを使用しています。このシステムは、キーワードが異なる場合でも、推論に基づいて関連コンテンツを発見できます。その利点は、読者にとっての発見可能性の向上と、出版社にとってのデータ再利用性の向上です。一方で、リスクとしては、継続的なコンテンツやオントロジーの更新が必要になる点が挙げられます。学習データが古くなると、関連性や精度が低下する恐れがあります。

したがって、あらゆる AI 活用の場面において、人間による監督は譲れない設計上の選択であり、人を常にプロセスに関与させることが前提となっています (図 18)。移行から DEI 分析、Alt-Text 生成に至るまで、各モデルはいずれも、分野専門家によるレビューと社内の品質保証をもって最終工程としています。これにより高い精度が保証されるだけでなく、モデルを再学習させ、性能を継続的に改善するためのフィードバックループを提供するものでもあります。AI は処理を担い、人間は判断を担います。このバランスこそが、編集チーム内においても、読者との関係においても、信頼を維持する要因となっています。利点としては、言うまでもなく、AI のスケーラビリティと人間の洞察力を組み合わせることで、両者の強みを最大限に活かすことができる点にあります。一方で、コスト圧力によって出版社が人間による監督を削減してしまえば、学術的公正性に対する信頼が損なわれるリスクが生じます。

ここで、改めて中核となるリスクについて、明確に触れておきたいと思います。データのプライバシーとセキュリティです（図 19）。生成 AI システムでは、暗号化されたデータ通信、セキュリティ確保されたストレージ、コンテンツベースのプロトコルが使用されていますが、すべての統合ポイント、とりわけサードパーティの大規模言語モデルについては、原稿の漏えいを防ぐために、厳格な審査が不可欠です。バイアスとハルシネーションです。生成モデルは、自信に満ちているものの不正確なコンテンツを生成することがあります。これを緩和するためには、プロンプトチェーン、能力スコアリング、そして最終的な公開前に人によるレビューを行う必要があります。透明性の喪失です。AI による判断は説明可能でなければなりません。私たちは、プロンプトに関する完全な監査証跡を維持しています。また、編集上の説明責任を確保するために、査読者による介入の中でモデルの出力を位置づけています。そして、自動化への過度の依存です。AI は編集の専門性を置き換えるのではなく、増強するものであるべきです。危険性はツールそのものにあるのではなく、人による検証なしにそれを使用することにあります。要するに、責任ある AI とは、単なる技術的な問題ではなく、文化の問題です。それは、透明性、品質保証、そして倫理的な利用を通じて信頼を築いていくことに他なりません。

以上をまとめますと、出版における AI はもはや実験段階のものではなく、すでに実運用の段階に入っています（図 20）。MPS および HighWire においては、迅速な制作、より高いアクセシビリティ、そして強固な研究公正を支えるワークフローの中に、AI が組み込まれています。しかし、その将来は責任ある導入にかかっており、AI が代替ではなく協働者として機能することを確保する必要があります。適切に活用されれば、AI は、学術エコシステムが常に追求してきたもの、すなわち効率性、包摂性、そして知識の伝達に対する信頼を出版社が実現する助けとなり得ます。

ご清聴ありがとうございました。

Benefits and Risks

AI enhances speed, quality, inclusivity, and integrity, but only when balanced against privacy, bias, transparency, and human oversight.

Benefits

- Efficiency and Speed
- Improved Quality and Consistency
- Enhanced Accessibility and Inclusion
- Strengthened Research Integrity

Risks

- Data Privacy and Security
- Bias and Hallucination
- Loss of Transparency and Accountability
- Overreliance on Automation

www.mpslimited.com © 2025 | MPS Limited | 2

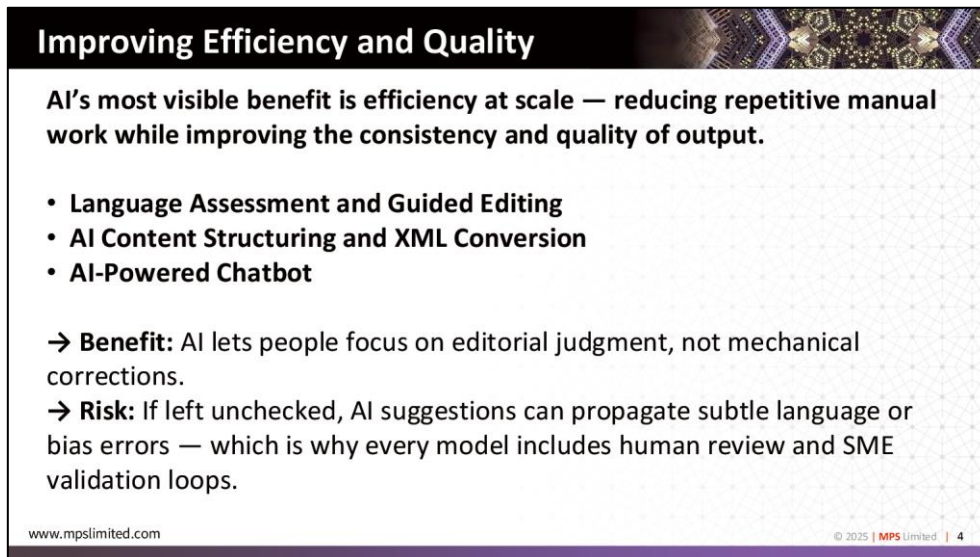
☒ 11

MPS Labs

MPS Labs is our research & development center that uses cutting-edge AI/ML/NLP technology for robust, scalable architecture & solution design that is integrated throughout our SaaS products and content services.

www.mpslimited.com © 2025 | MPS Limited | 3

☒ 12



Improving Efficiency and Quality

AI's most visible benefit is efficiency at scale — reducing repetitive manual work while improving the consistency and quality of output.

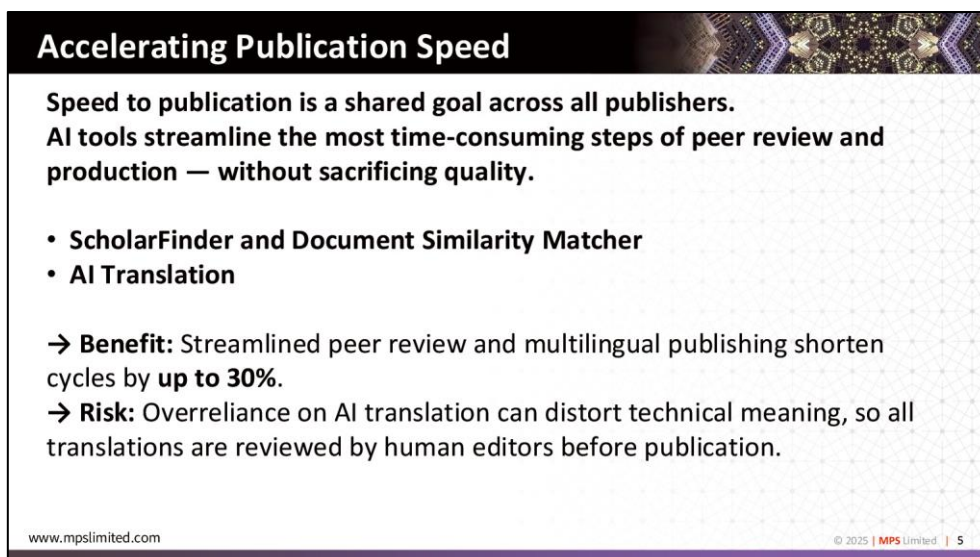
- Language Assessment and Guided Editing
- AI Content Structuring and XML Conversion
- AI-Powered Chatbot

→ **Benefit:** AI lets people focus on editorial judgment, not mechanical corrections.

→ **Risk:** If left unchecked, AI suggestions can propagate subtle language or bias errors — which is why every model includes human review and SME validation loops.

www.mpslimited.com © 2025 | MPS Limited | 4

図 13



Accelerating Publication Speed

Speed to publication is a shared goal across all publishers. AI tools streamline the most time-consuming steps of peer review and production — without sacrificing quality.

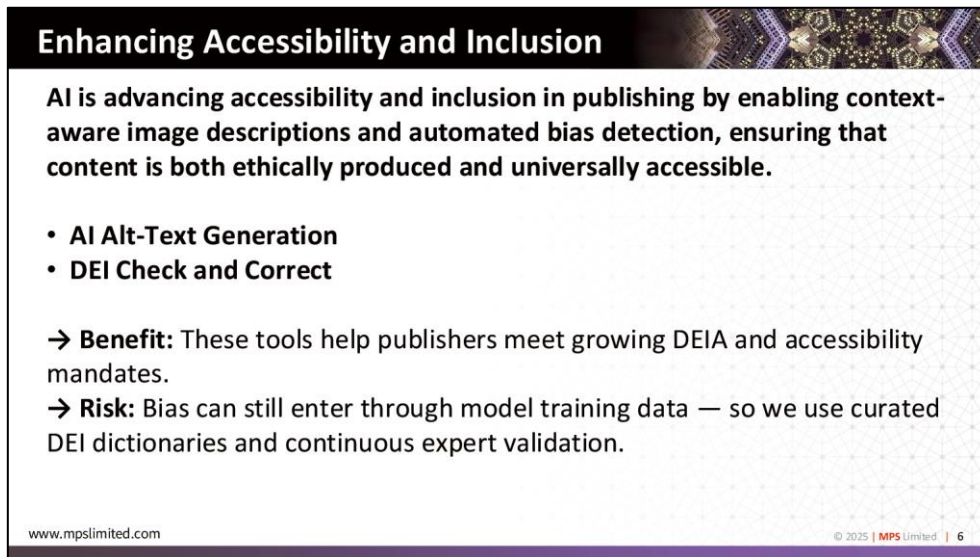
- ScholarFinder and Document Similarity Matcher
- AI Translation

→ **Benefit:** Streamlined peer review and multilingual publishing shorten cycles by **up to 30%**.

→ **Risk:** Overreliance on AI translation can distort technical meaning, so all translations are reviewed by human editors before publication.

www.mpslimited.com © 2025 | MPS Limited | 5

図 14



Enhancing Accessibility and Inclusion

AI is advancing accessibility and inclusion in publishing by enabling context-aware image descriptions and automated bias detection, ensuring that content is both ethically produced and universally accessible.

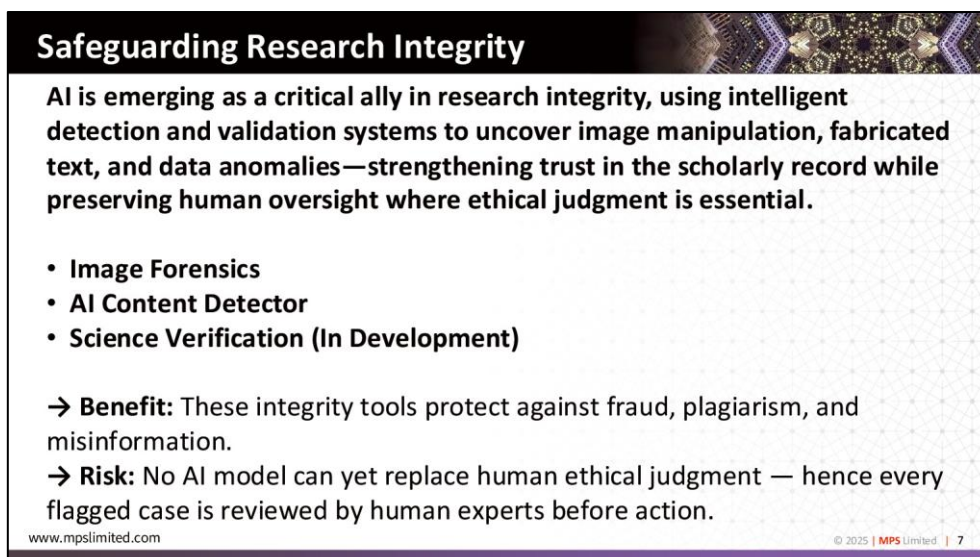
- AI Alt-Text Generation
- DEI Check and Correct

→ **Benefit:** These tools help publishers meet growing DEIA and accessibility mandates.

→ **Risk:** Bias can still enter through model training data — so we use curated DEI dictionaries and continuous expert validation.

www.mpslimited.com © 2025 | MPS Limited | 6

図 15



Safeguarding Research Integrity

AI is emerging as a critical ally in research integrity, using intelligent detection and validation systems to uncover image manipulation, fabricated text, and data anomalies—strengthening trust in the scholarly record while preserving human oversight where ethical judgment is essential.

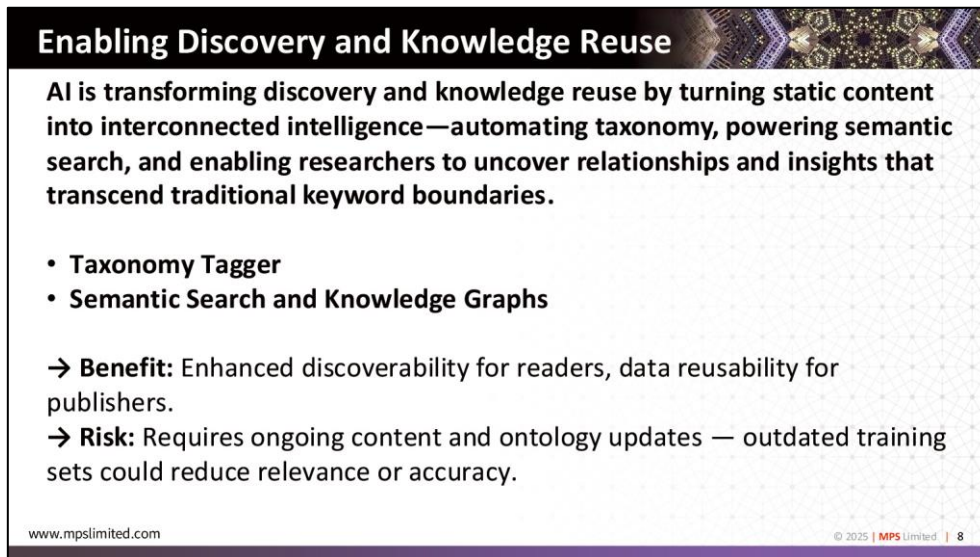
- Image Forensics
- AI Content Detector
- Science Verification (In Development)

→ **Benefit:** These integrity tools protect against fraud, plagiarism, and misinformation.

→ **Risk:** No AI model can yet replace human ethical judgment — hence every flagged case is reviewed by human experts before action.

www.mpslimited.com © 2025 | MPS Limited | 7

図 16



Enabling Discovery and Knowledge Reuse

AI is transforming discovery and knowledge reuse by turning static content into interconnected intelligence—automating taxonomy, powering semantic search, and enabling researchers to uncover relationships and insights that transcend traditional keyword boundaries.

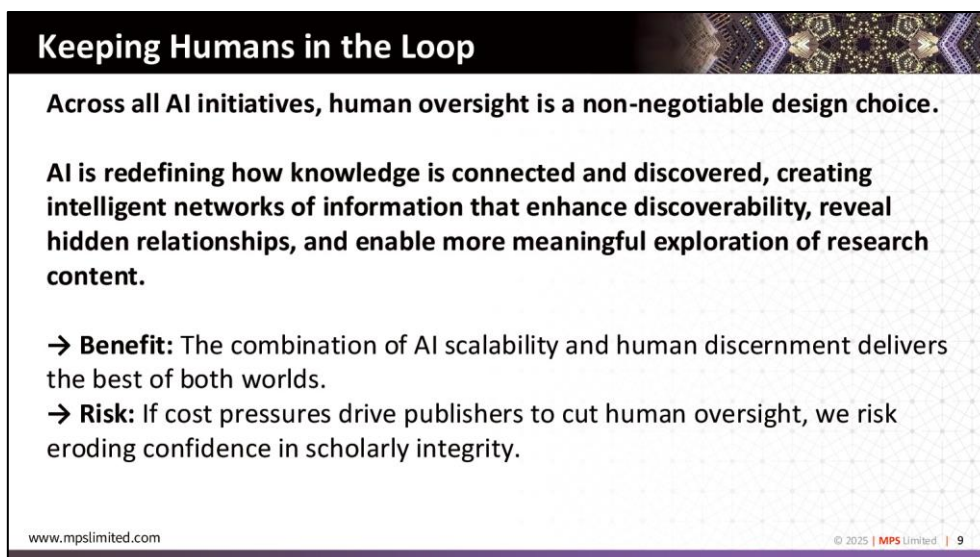
- Taxonomy Tagger
- Semantic Search and Knowledge Graphs

→ **Benefit:** Enhanced discoverability for readers, data reusability for publishers.

→ **Risk:** Requires ongoing content and ontology updates — outdated training sets could reduce relevance or accuracy.

www.mpslimited.com © 2025 | MPS Limited | 8

図 17



Keeping Humans in the Loop

Across all AI initiatives, human oversight is a non-negotiable design choice.

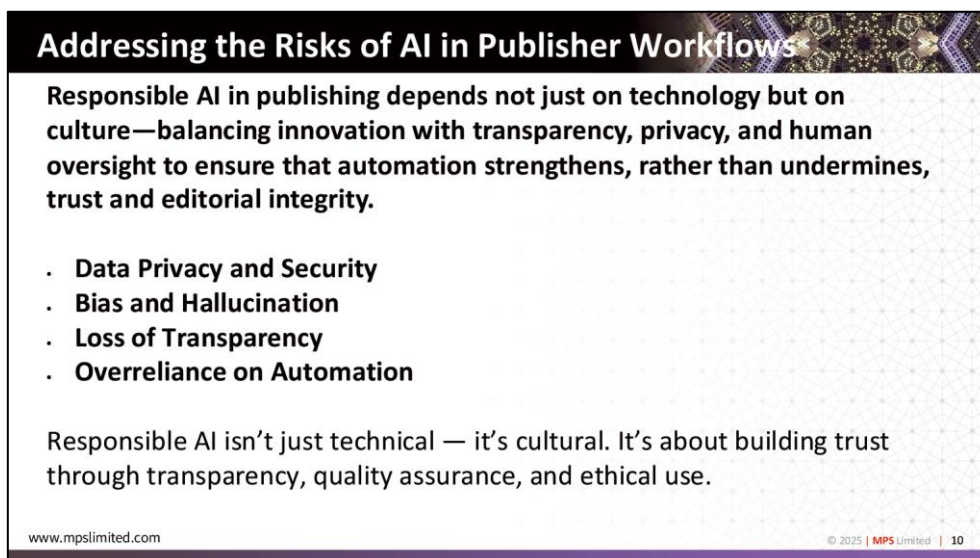
AI is redefining how knowledge is connected and discovered, creating intelligent networks of information that enhance discoverability, reveal hidden relationships, and enable more meaningful exploration of research content.

→ **Benefit:** The combination of AI scalability and human discernment delivers the best of both worlds.

→ **Risk:** If cost pressures drive publishers to cut human oversight, we risk eroding confidence in scholarly integrity.

www.mpslimited.com © 2025 | MPS Limited | 9

図 18



Addressing the Risks of AI in Publisher Workflows

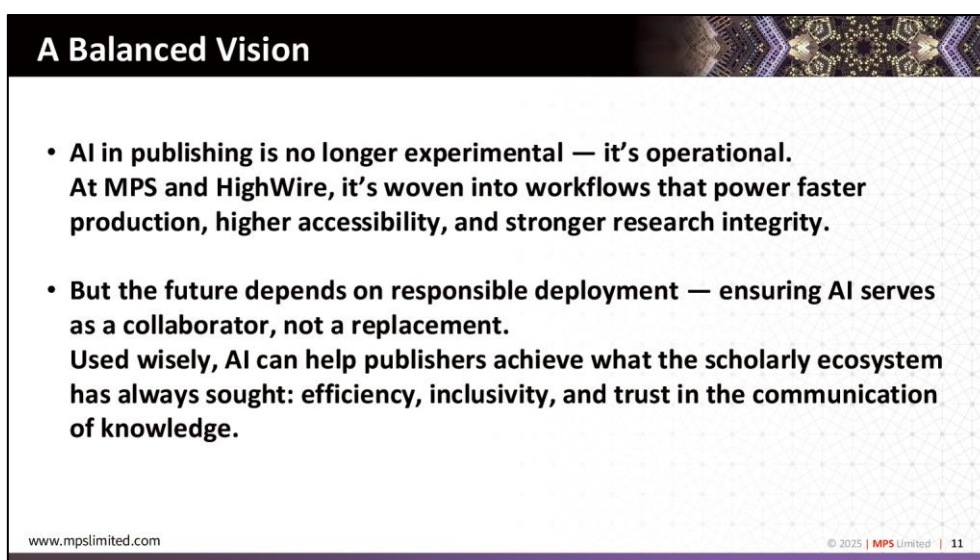
Responsible AI in publishing depends not just on technology but on culture—balancing innovation with transparency, privacy, and human oversight to ensure that automation strengthens, rather than undermines, trust and editorial integrity.

- **Data Privacy and Security**
- **Bias and Hallucination**
- **Loss of Transparency**
- **Overreliance on Automation**

Responsible AI isn't just technical — it's cultural. It's about building trust through transparency, quality assurance, and ethical use.

www.mpslimited.com © 2025 | MPS Limited | 10

図 19



A Balanced Vision

- **AI in publishing is no longer experimental — it's operational. At MPS and HighWire, it's woven into workflows that power faster production, higher accessibility, and stronger research integrity.**
- **But the future depends on responsible deployment — ensuring AI serves as a collaborator, not a replacement. Used wisely, AI can help publishers achieve what the scholarly ecosystem has always sought: efficiency, inclusivity, and trust in the communication of knowledge.**

www.mpslimited.com © 2025 | MPS Limited | 11

図 20

2.3



【Harnessing Generative AI for Actionable Insights and Societal Impact】

柿田 佳子 氏
(Elsevier)

ELSEVEIR の柿田です。本日はよろしくお願ひします。このような機会をいただき、ありがとうございます。

最初に、簡単に自己紹介をします (図 21)。私は、2006 年に ELSEVEIR 日本法人に入社したので、かなり長くなります。入社後 15 年間は営業チームに在籍し、Scopus を担当してきました。2021 年に Scopus のプロダクトチーム (製品開発のチーム) に異動し、現在は主に AI ソリューションの開発にフォーカスしています。ELSEVEIR は出版社でもあり、研究者の皆さんに使っていただく製品開発を行っています。私は製品開発に従事しているので、本日は製品開発の話をしてします。

ELSEVEIR として、なぜ AI ツールを開発をしているのかというと、プロダクトチームとして私たちは、研究者の皆さまが信頼して使えるような AI ツールをお届けしたいという思いで開発を続けています。

振り返ってみると、2022 年に ChatGPT がアナウンスされました。私も使ってみて、とても驚きました。2023 年 8 月には Scopus AI アルファ版をリリースし、2024 年 1 月に正式にリリースしました。その後も、ELSEVEIR のさまざまなプロダクトの上に AI 機能が搭載されてきました (図 22)。

2025 年 3 月には、皆さんご存じの方も多いと思いますが、電子ジャーナルの Science Direct AI を正式リリースしました。また、9 月のプレスリリースには、次世代の AI ソリューションとして開発を進めている Next gen solution ベータ版をリリースしました。私自身が関わっているプロダクトとしては、Scopus AI の開発と、現在進めている Next gen solution の開発チームを率いています。

そのような立場から、本日は、プロダクト開発で特に重視していることを話します (図 23)。たくさんある中でも、三つのポイントを挙げています。一つ目は、Trust です。本日のこれまでのセッションの中でも、スピーカーの皆さんは何度も Trust と言っていました。やはり、私たちのこのコミュニティの中では、何よりも Trust が重要ではないかと思っていますので、とにかくこの部分にフォーカスして開発を進めています。

既に、世の中には AI のツールがたくさん普及してします。その中で、なぜ ELSEVEIR が AI ツールを開発しなければいけないか、信頼できるツールはどれだけあるのかということ考えたときに、ELSEVEIR として Trust に貢献できるのではないかという思いで開発しています。

二つ目のポイントは、Actionable insight です。ELSEVEIR はコンテンツを持っているので、コンテンツと AI 機能を組み合わせることできざまな機能を生み出すことは、言ってみれば ELSEVEIR としては簡単なことです。そうではなく、信頼できるものを開発することに加えて、実際に使える情報をどのように示すかということが重要だと考えています。

三つ目のポイントは、Impact です。研究のアイデアからインパクトにつなげることは重要だと思いますが、より迅速に学術的・社会的インパクトにつなげるためにはどのようにお手伝いができるかという視点で考えています。

まず Trust です。大前提となる基本として取り組んでいることがいくつかあります (図 24)。信頼できるデータを使用しています。Scopus のコンテンツの他にも、ELSEVEIR として出しているジャーナルのデータがあります。そもそも、信頼できるデータに基づいていることが重要だと思っています。これまでも話がありましたが、リスポンシブル AI です。ELSEVEIR として責任ある AI の原則を規定しています。そこに順守する開発を進めています。

やはり、プライバシーとデータのガバナンスは非常に重要です。例えば、私たちのツールであれば、大規模言語モデル(LLM)は ELSEVEIR のプライベートの環境の中で運用しています。利用者の皆さんが入れたプロンプトなども、LLM の訓練に使われるようなことはありません。もう一つ強調したいことは、評価です。ツールを開発することは簡単ですが、やはりきちんと評価することが大事だと考えています。

この評価には三つの評価を組み合わせています。一つは、専門家による評価です。研究なので、専門家による評価が大事です。二つ目、アルゴリズムを使った評価を組み合わせています。人による評価と、機械的な評価です。三つ目は、次のポイントにもなりますが、利用者の皆さんからのフィードバックも評価の一部として捉えています。その三つを組み合わせることで、信頼性を高めるようにしています。

これらに加え、特に重視して取り組んでいることを二つ挙げます (図 25)。私たちは、毎日のように利用者の皆さまからの声を聞いています。その中で、研究者の皆さまは既にたくさんの AI ツールを利用していますが、そうは言っても信頼できない面もあると思います。

信頼ではない理由の中でも特によく聞くのは、裏で何をやっているかよく分からない『ブラックボックス』、『批判的思考(Critical Thinking)や創造性(Creativity)が損なわれる可能性がある』という懸念によって信頼できないという声をよく聞きます。

完全に解決できないとしても、解決策として取り組んでいることがあります。ブラックボックスの解決策としては、透明性の向上を重視しています。Critical Thinking に関しては、創造性が損なわれないように注意するというよりも、むしろ Critical Thinking の向上を支援するための製品開発に取り組んでいます。

透明性の向上における Scopus AI の事例です (図 26)。どのような検索が行われているか分かります。左側は回答、右側は基になった論文を簡単に確認できるようになっています。ポイントとしては、Scopus のコンテンツを使っています。論文といってもどのような論文なのか分かりません。ネット上のいわゆる論文と分類されているものであれば、その論文がどのようなジャーナルに掲載されているのか、何か規定があるのかということそうではありません。Scopus にきちんと収録されたジャーナルに掲載されている論文に基づいています。また、それぞれがどのようなサイテーションなのかを確認できるようになっています。

そういった基本的なことに加えてコパイロット機能を付けています。これは Microsoft の Copilot とは全然関係なく、自社開発したコパイロットです。プロンプトを入れた後に、裏でどのような検索が行われているのかということ、このような形で明示しています。自然言語で入れたものに加えてプロンプトをブレイクダウンし、検索式を作成しています。それらを組み合わせ、最も関連性の高い論文を引っ張ってきます。それに基づいてレスポンスを生成するプロセスです。

もう一つは、批判的に思考する力の向上を支援する Deep Research の例です。Deep Research 機能を自社開発し、今夏、Scopus AI に搭載しました。アプローチとしては、先ほどトニーさんからもありましたが、研究者の代わりにレポートを作成するアプローチでは全くありません。むしろ、研究者の方がさらに探究したくなるような多角的な視点やアイデアを提示する構成になっています。

どのような構成にして、何を提示すれば研究者の皆さんの Critical Thinking をよりサポートできるようなレポートになるのかを、研究者の皆さんの話を伺いながら開発を進めました。リリースしていますが、さらに改善しています。リプレースではありませんが、よりサポートしていく視点で全ての機能を開発しています。

次に、ポイント 2、Actionable insight です。実際に Scopus AI の画面を録画してきたものです。何か一つプロンプトを入れると、先ほどもスクリーンショットを示しましたが、Scopus AI のコパイロットが裏でどのような検索をするのか、自然言語のものと検索式でブレイクダウンしたものが出てきます。本日は話するのは、研究テーマの仮説設定の例です。

この後に一つクリックします。何をしているかという、もともと利用者の方が関心のある領域に関してさらにクラスタリングをし、小さな領域を作っています。ブラウザの機能を使い、分かりやすいように日本語にしています。小さなクラスタリングが二つのことを行います。それぞれのクラスタリングされた小さな分野が、トレンドを示します。コンスタントに論文が出ている領域なのか、あるいは論文が増えている領域なのか、今まで論文がなかった領域で、新しく論文が出始めている領域なのかを示します。また、それぞれの小さなクラスタリングの中で、仮説の提案をしています。

そのような形で、実際に研究者の皆さんがアクションにつなげられるような機能を出すことを重視し、機能開発をしてきています。すみません。少し早くなってしまいました。Actionable insight では、実際に皆さんが論文を書くとき、研究を進めていく中で、どのような情報を表示できれば研究を進めていくことができるのか、その迅速化につながるのか、プロダクティビティを上げていくことにつながるのかを考えて作っています。表示して終わるのではなく、実際に表示したものから全てクリックできるようになっているので、深掘りしていくことができるようになっています。そういった機能も、透明性を高める、リプレースするのではなく研究者を助けていく機能にするという考え方に沿っているからこそこのプロダクト開発になっています。

これまでのスピーカーの皆さんもおっしゃっていたように、信頼を得ること、リプレースするのではなく研究者をサポートすることを組み合わせて考えてきています。

ポイント三つ目、Impact です (図 27)。これまで私たちがどのような形で、どのような考え方で開発してきたのか話をしてきました。今までの製品開発は、もともと ELSEVEIR が持っているコンテンツと LLM を組み合わせ、自社が持っているさまざまな AI 技術、あるいは検索技術を組み合わせました。製品開発の方向性などについて、皆さまから非常に良い評価をいただいています。

そうは言っても、全く課題がなくなっているかというわけではなく、できることはまだまだあると考えています。特に、研究のアイデアから Impact につなげる点に関しては、やはり研究者の皆さんや、大学のような研究機関の執行部の皆さんにとっても重要なところだと思います。ファンディング機関や皆さんにとっても、どこにお金を出すかを考える上では

Impact が重要です。どのような課題にお役に立てる可能性があるかを考えたときに、二つの点に注目しています。

一つは、今申し上げたように、研究テーマの仮説設定から成果に至るまでには時間がかかります。スピーカーの皆さんもおっしゃっていましたが、研究のプロセスはとても長いです。その中でいろいろなプロセスを踏むので、全体に時間がかかる点が大きな課題です。もう一つの課題は、研究の各段階で必要なツールを選んで使う必要があることです。いろいろなツールを使っていく面白さもあると思いますが、自分でそのツールを探さなければいけません。そして、それらのツールが本当に信用できるものなのか分からないので、自分で試していかなければいけません。その上で、リスクがあることを知らずに使ってしまい、何かリスクを負ってしまったときの影響も大きい、そういった不安もあります。

それぞれの段階で必要なツールを自分で選んでいくことは個人にとっても難しいですし、研究機関にとっても難しいことです。そういった話を伺った中で、製品開発の次の段階として、次世代型エンドツーエンド AI ソリューションを9月にプレスリリースしました。Scopus AI は Scopus AI のディスカバリーの部分、あるいは Science Direct AI であれば論文を読むところを助けるといったように、一つ一つのユースケースを私たちのそれぞれのプロダクトでサポートするのではなく、もっとシームレスな AI アシスタントとして、一つのプラットフォーム上で、研究ワークフローをエンドツーエンドでサポートするようなことができれば、非常に有効なのではないかと考えています。

そのようなサポートができれば、研究アイデアのところから成果に至るまでの時間も短縮できる可能性があると思います。皆さんがなかなか大変だとおっしゃっている、それぞれの段階で必要なツール、それぞれの段階で信頼できるツールを選ぶ大変さも軽減できるのではないかと考えています。例えばそのためのアイデア創出、研究プロジェクトの計画、ファンディングの特定があります。自分の研究テーマにはどのようなオポチュニティがあるのか、アプリケーションを書くときのサポートを、最初にフォーカスするユースケースとして作り始めています。

どのユースケースにフォーカスするか、どの段階にまずフォーカスするのが最も価値が高いかを研究者の皆さんに聞いて決めているところです。一つポイントになるのは、ELSEVEIR としてはさまざまなデータを持っていますが、今まではどちらかというとサイロで使っていました。Scopus コンテンツであれば Scopus AI を使っていましたが、そうではなく、さまざまなコンテンツを意味のある形で統合して利用できるようにする予定です。

また、フルテキストに関しても、ELSEVEIR が出しているフルテキストではなく、他の出版社の皆さんが出しているフルテキストも一緒に組み込み、一緒に検索できるような形で提示できるように、現在、さまざまな出版社の皆さまと話をしています。

また、そのときに重要になってくるのは、特定の出版社、例えば ELSEVEIR だけに有利になるような論文を引っ張ってくるようなアルゴリズムではなく、中立したアルゴリズムでなければいけません。そのために独立したアドバイザリーレポートを設置し、品質、透明性、公平性を確保したいと考えています。もちろん、私たちが出すコンテンツだけではなく、利用者自身のコンテンツをアップロードして組み合わせることも可能になります。

最後に申し上げたいことは、やはり Trust です。信頼性をさらに高められるような機能も搭載する予定です。そのような形で、次のステージに向けて開発を進めています。私からは以上になります。ご清聴ありがとうございました。

自己紹介

エルゼビアでの経歴

- 2006年にエルゼビアにScopusの営業担当として入社
- 15年間営業チームに在籍
- 2021年にScopusプロダクトチームに異動
- 現在は生成AIツールのプロダクト開発に注力

ELSEVIER

図 21



図 22



図 23



図 24

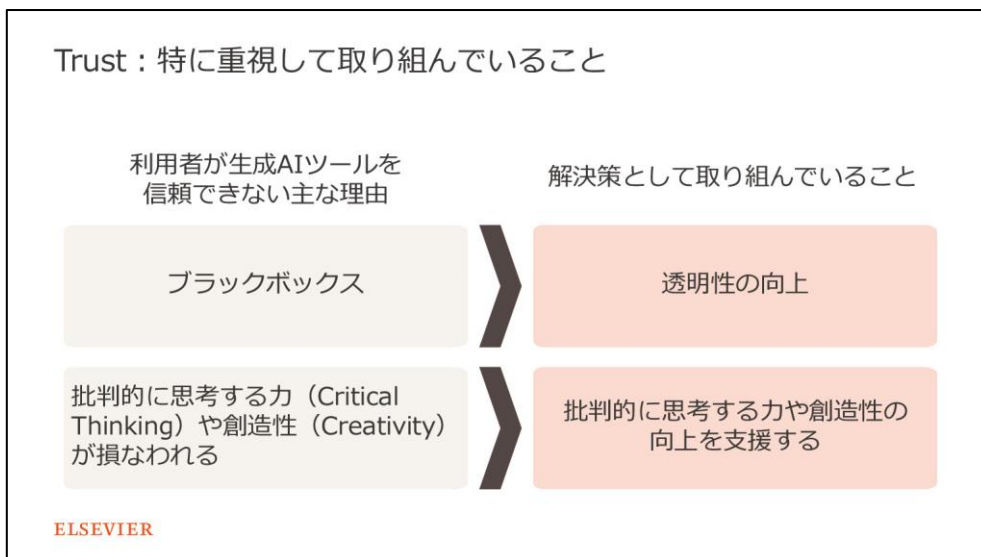


図 25



図 26

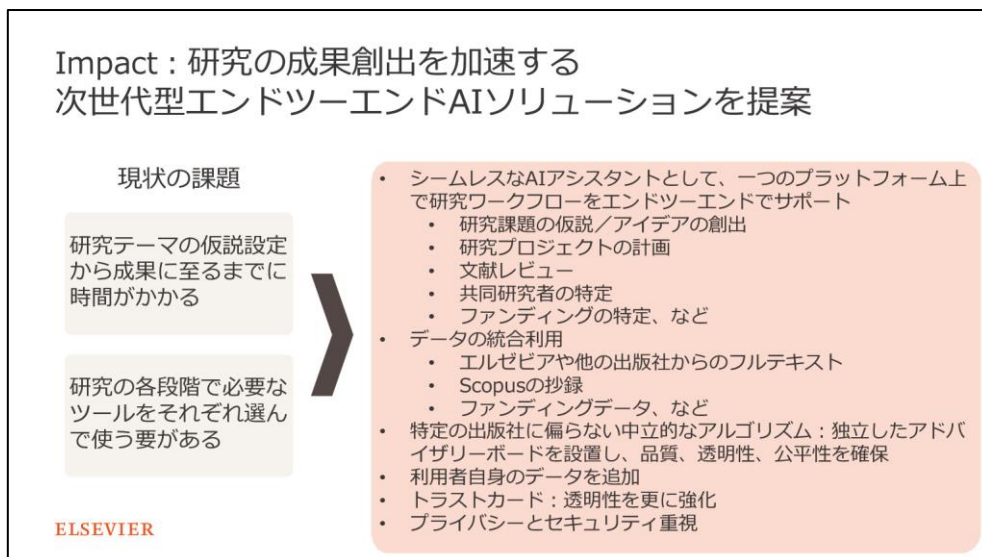


図 27

2.4



【信頼される日本語医療 LLM を目指して：透明性のある構築プロセスの重要性】

小林 和馬 氏

(国立がん研究センター研究所)

ご紹介いただき、誠にありがとうございました。国立がん研究センター研究所、および国立情報学研究所の小林和馬です。本日は、J-STAGE セミナーにお招きいただき、誠にありがとうございます。私からは、『信頼される日本語医療 LLM を目指して:透明性のある構築プロセスの重要性』と題し、私たちが進めている日本語医療従事者 LLM に関する研究開発において、なぜ透明性のある構築プロセスが重要なのかという点にフォーカスしながら話をします。

医療分野における大規模言語モデル(LLM)の可能性についてです(図 30)。近年、LLM は目覚ましい進歩を遂げています。特に、重要な点としては、LLM 登場前までは、人間の医師にとっていわばコア・コンピタンスでした。当面は AI によって置き換えられることはないだろうと思われていたようなさまざまな臨床能力、すなわち医師国家試験や診断困難例に対する診断、検査前後での診断の確率の推定、臨床技能や態度といったような評価項目において、LLM が人間と同等化しています。場合によっては、それ以上のパフォーマンスを発揮していることが報告されるようになってきました。

このような LLM と医療者との関係の在り方について、最近、示唆的な論文があったので紹介します。こちらの研究では、LLM によって支援された医師は、LLM を用いなかった医師に比べて、複雑な臨床シナリオにおける診断のスコアが優位に向上したことを報告していました(図 31)。この研究において実際に試験として用いられたのは、例えばこのような臨床シナリオになります。

この例(図 32)では、72 歳の女性が胆嚢穿孔という比較的重篤な疾患に対する緊急手術の後に、心房細動を発症しました。現在、容体は落ち着いているようですが、こうした患者さんに対して追加のモニタリング検査、あるいは治療を推奨するかどうかを LLM に問い合わせていました。実際にある LLM に対し、この症例を提示して質問を行った様子を供覧します(図 33)。

この LLM は、素早く症例を分析するとともに、ユーザーからの追加の質問に答えることにより、この症例において必要な次の診療的なステップを推奨してくれています。少し動画が速かったと思うので、先ほどの回答について詳細に分析してみたいと思います。

この LLM は、患者の容体を正確に要約した後に、ユーザーからの問い掛けにあった追加のモニタリング検査、または治療の必要性に関してさまざまな角度から分析しています(図 34)。特に、脳梗塞を予防するためのリスク評価に関し、CHADS スコアと呼ばれる、臨床スコアの計算に基づいた治療薬の投与を検討すべきと述べていました。

そこで、この患者さんにおける CHADS スコアの計算をユーザーとしてあらためて問い合わせました。正確なスコアの計算とリスク評価に基づき、この症例においては脳梗塞を予防するための治療が推奨されました(図 35)。このように LLM を用いることによって、比較的複雑な臨床シナリオにおいても、適切な治療選択へ到達できる可能性が示唆されていると思います。

既にお気付きの方も多くいると思いますが、先ほどのデモ動画で登場した LLM は、ChatGPT ではなく SIP「統合型ヘルスケアシステムの構築における生成 AI の活用」の支援をいただきました。国立情報学研究所の相澤彰子先生のチームの主導によって開発されたものです。このプロジェクトでは、1.医療の質の向上、診察・受診支援、医療業務の効率化、2.医療行政の支援、3.研究開発の支援・医学知識の発見といった医療を取り巻く社会課題の解決を掲げ、日本の医療に特化した LLM の研究開発を展開しています(図 36)。

研究開発体制です(図 37)。研究統括は、国立情報学研究所の相澤彰子先生です。大きく二つに分かれた研究体制によって実施されています。私ども国立情報学研究所や京都大学・東京科学大学・東京大学・早稲田大学の先生がたが所属する情報チームが LLM の基盤開発を担いました。自治医科大学・東京大学・九州大学等、非常に多くの医療機関の先生がたで構成される医療チームには、臨床現場のニーズに基づいたユース開発の開発や評価を担っていただきました。

私たちの研究開発のプロセスについて、もう少し詳しく紹介します。まず、私たちのアプローチの根幹にあるのは、国立情報学研究所が中心となって開発したオープン日本語 LLM である LLM-jp シリーズです(図 38)。この LLM-jp とは、学習データやモデルのパラメータを含む全ての構築過程がオープンになっている国産の LLM です。私たちはこの LLM-jp をベースモデルとし、まず大規模な日本語医療コーパスを構築しました。次にこのコーパスを使って医療ドメインに適応させるための最適なドメイン適応戦略を探索しました。そして、モデルの性能を正しく評価するための日本語医療評価ベンチマークを構築し、その評価を行いました。

こうしたサイクルを回すことによって、基盤型日本語医療 SIP-JMED-LLM シリーズの継続的な開発に取り組んでいます。この基盤的なモデルに対し、実際の臨床情報を用いたファ

インチューニング等をそれぞれの医療機関、医療機器開発企業様などの手で行っていただき、具体的な医療機器や製品といったユースケース特化型の、LLM の研究開発につなげていきたいという展望を持って取り組んでいます。

医療系 Web クローリングによるコーパス収集です(図 39)。出版社等からの利用許諾、医療系 Web クローリング、汎用 Web クローリングの三つの取り組みについて順に紹介します。まず、日本語で記載された医療テキストを可及的に収集するために、このプロジェクトにおいては出版社などから利用許諾をいただいた医学教科書に加えて、医療名のドメインに絞った Web クローリングを実施しました。

一方で、Web クローリングなどによって取得されたテキストの多くは、実際には英語で書かれたものが中心でした。これに大規模な機械翻訳を実施することによりテキストの日本語化を行い、最終的には、およそ 80 ビリオントークンからなる医療系コーパスを構築しました。1 トークンというのは、およそ 1 単語に相当する単位だと考えてください。

この 80 ビリオントークンからなるコーパスは、日本語と英語の言語の比率がおおよそ均衡したものとして構築されています。こちらの円グラフ(図 40)で示すとおり、日本語が 45.4 パーセント、英語で記載されたテキストは 45.7 パーセントでした。残りが、日本語と英語の対訳データとしてのパラレル・コーパスによって構成されたものです。

具体的なテキストの内容は、Pub Med Abstract や J-STAGE、J-GLOBAL などからなる医療分野の論文データや、ウェブから収集された診療ガイドラインです。さらには、トークン数としては比較的小規模になりますが、医学教科書なども含まれています。

評価の物差しについて説明します。構築したモデルの性能を客観的、かつ包括的に評価するために、私たちは JMedBench という、日本語医療分野に特化した評価のフレームワークを開発しました(図 41)。この JMedBench は、五つのタスクと 20 のデータセット、3 万 8000 サンプルのテストデータで構成されたものです。

具体的には、医師国家試験を含む多肢選択式の質問応答(MCQA)、医学生物学用語の固有表現認識(NER)、医学生物学文書の分類(DC)、医学生物学文書の機械翻訳(MT)、医学生物学文書の意味的類似度推定(STS)から構成されます。その上で、汎用モデルである LLM-jp の日本語医療分野へのドメイン適応を行いました。この最適な学習戦略を探索するにあたり、一つの研究的な取り組みを紹介します。

私たちの取り組みの中で、痛感したことがあります。日本語の医学テキストは、比較的希少です。実際にウェブから収集された医療関連文書の多くが英語だったため、言語比率が偏っています。つまり、継続事前学習用のコーパスの中に、日本語よりも英語で記載されたテキストの割合が圧倒的に多くなってしまうことは、今回の日本語医療分野へのドメイン適応を行いたいという目的に対し、負の影響を与えることが懸念されました。機械翻訳などを活用し、そのコーパスにおける最適な言語構成を探索しました。

具体的には、英語と日本語の比率が異なるさまざまなコーパスを準備しました。これを用いてモデルを学習した前後での、日本語医療分野のパフォーマンスを観察することで、最適な言語構成比を検討しました(図 42)。

この検討から抽出された知見を共有します。コーパスの言語的な要素、すなわち、日本語のトークン数、英語のトークン数、パラレル・コーパスのトークン数と、日本語医療タスクのスコアの変化量を観察しました。例えば、英語のトークン数は、日本語タスクを含むタスクと正の相関を示しました(図 43)。これは、LLM が英語で獲得した知識を、日本語によって問い掛けられた医学的な質問に対する答えに転用できるという、ある種の知識の転移を起こすことができることを示唆しているのではないかと考察します。

また、日本語のトークン数については、特に本邦の医師国家試験の過去問から構成されるベンチマークのような、日本固有の知識を問う、ベンチマークとの正の相関を示すことが分かりました。最終的に開発された SIP-JMED-LLM モデルは、過去 5 年間の本邦の医師国家試験において、いずれも人間の合格基準を上回りました(図 44)。また、平均して GPT-4 よりも優れた性能を示しました。現在、SIP-JMED-LLM モデル(図 45)を臨床のユースケースに特化させる方向で、医療機関ごとのデータを用いた追加的な開発を行っています。

ここまでは、主にユース的な側面と性能について話をしてきましたが、本日のもう一つの重要なテーマは、信頼性のある構築プロセスです。AI、特に生成 AI を医療のような社会の重要インフラに組み込む際には、この技術革新がもたらす大きな可能性と同時に、やはり容易には解決できない倫理的・法的・社会的課題、すなわち ELSI 課題についても考慮しなければいけません(図 46)。

例えば、学習データに含まれるバイアスと差別の問題があります。AI の判断根拠が分からない透明性の欠如、いわゆるブラックボックスの問題もあります。また、患者さんの機微な情報を取り扱う上で、プライバシーとデータ保護の観点も考慮しなければいけません。万が一、AI の判断によって危害が生じてしまった場合の責任の所在についても考慮が必要で

す。さらには、偽の情報や情報操作に悪用されるリスクや、セキュリティーリスクといったような多岐にわたる課題が、私たちに投げ掛けられているのではないかと思います。

医療 LLM における脆弱性の具体例として、データ汚染攻撃を紹介します(図 47)。データ汚染攻撃とは、LLM の学習用データセットに対して、悪意を持って偽の医療情報を注入するような攻撃です。ある研究では、学習用データセットのわずか 0.001 パーセントをこのような偽の情報に置き換えるだけで、医療 LLM が有害に医学的誤りを含む文章を生成する可能性が著しく高まると報告されています。

さらに深刻なのは、こうした汚染されたモデルであっても、標準的な医学のベンチマークでは、どうも合格してしまうようです。こうした汚染を検出することは、非常に困難であることも同時に報告されています。医療のようなハイリスクな分野における LLM を構築し、運用する際には、やはり信頼できるデータソースに基づいた学習用データセットを構築し、できれば、自ら管理し続ける必要があると示唆されます。

もう一つ重要な ELSI 課題は、機微情報、すなわちプライバシーに関する問題です。従来、機械学習モデルの学習済みパラメータそのものは、特定の個人との対応関係が排斥している限りにおいては、個人に関する情報には該当しないという見解が示されてきたと思います。こうした見解は、いわゆる識別モデルに関しては、一般的に成立するのではないかと思います。

しかしながら、LLM や画像生成 AI のような生成モデルについては、この妥当性は必ずしも自明ではありません。こちらの例では、画像生成 AI で学習用データセットに含まれていた、特定の人物の画像を、復元する意図を持ってプロンプトをモデルに投げ掛けることにより、実際に AI が元のデータセットのサンプルを生成してしまう現象が報告されています(図 48)。

こうした生成 AI が有するプライバシーリスクを踏まえた、安全なデータセットの加工や、学習法の構築、さらには、こういったものを包含するガバナンスの在り方を考えなければいけないタイミングではないかと思います。

以上、本日の話を最後にまとめます(図 49)。私ども国立情報学研究所を中心とするチームにおいては、オープンな LLM-jp シリーズをベースモデルとして、安全性・信頼性を持つ、オープンな日本語医療 LLM の研究開発に取り組んできました。現時点で開発中のモデルは、本邦の医師国家試験の合格基準に達しており、スクラッチ・モデルに始まる全ての開発過程が透明化されて国産 LLM として一定の快挙ではないかと考えます。

今後は、こうしたオープンな開発プロセスを持つ私たちのプロジェクト自体を、一つの検証環境として活用することにより、先ほど述べたプライバシーリスクをはじめとした、生成 AI を取り巻く、さまざまな課題に関する健全な知識の循環、社会受容に関する議論の場の創出を期待したいです。以上、ご清聴誠にありがとうございました。

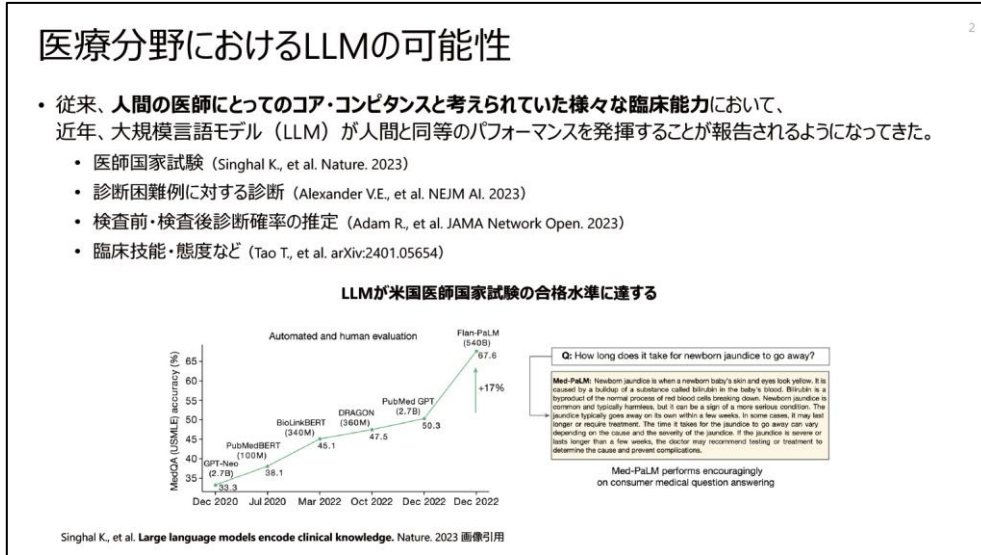


図 30

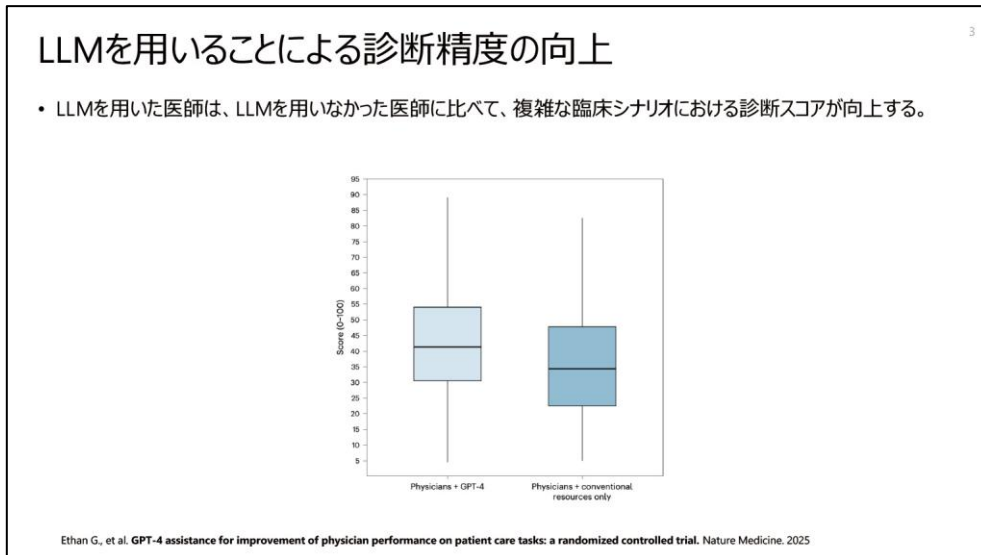


図 31

72歳女性。胆嚢穿孔の術後に心房細動を発症。

- 72歳の女性患者。GERD、2型糖尿病、コントロール不良の高血圧の既往歴があり、発熱、悪寒、腹痛を主訴として受診。腹部CT検査で胆嚢穿孔性胆嚢炎と腹膜炎が認められた。広域スペクトラム抗生物質が投与され、緊急開腹胆嚢摘出術が行われた。手術は順調に終了し、直後の合併症はなかった。術後1日目、テレメトリーで無症状の心房細動が短時間検出されたため、看護スタッフから呼び出された。テレメトリーデータの詳しい確認によると、過去24時間で心房細動が3回発生し、それぞれ15～20秒間持続していた。胆嚢炎と腹膜炎の観点からは、臨床状態は改善しており、手術後発熱や悪寒はなく、液体食を摂取可能です。血圧と心拍数は現在正常です。検査所見では白血球数の減少傾向が認められます。術前心電図と現在の心電図ともに、正常洞調律で、重大な異常は認められません。心エコー検査の既往はありません。現在、抗凝固薬や抗血小板薬は投与されていません。
- 質問: テレメトリーモニタリングを継続する以外に、現在、追加のモニタリング、検査、または治療を推奨しますか？もしそうなら、何ですか？

胆嚢穿孔にて入院・緊急手術となった72歳女性の術後に心房細動が認められた。今は落ち着いているが、追加の検査や治療は必要か？



Atrial Fibrillation

<https://www.hopkinsmedicine.org/health/conditions-and-diseases/atrial-fibrillation>

図 32



SIP-jmed-llm-2-8x13b-med-instruct

今日はどうしてお手伝いしましょうか?

Code Interpreter

- 医療関連文書からの要約情報抽出
- 医療関連文書の要約
- 医療関連文書の日本語訳

図 33

症例に対する適切な分析

6

患者の状態に関する正確な要約

追加のモニタリング、検査、または治療の必要性

<p>1. 心房細動の持続時間と頻度の評価</p> <ul style="list-style-type: none"> 現在の心房細動は短時間（15～20分）で無症状ですが、頻度が増加する場合や持続時間が長くなる場合には、より詳細な評価が必要です。 24時間ホルター心電図や長時間のテレメトリーモニタリングを検討し、心房細動のパターンをさらに評価します。 	心房細動の正確な診断に向けた追加モニタリング
<p>2. 心エコー検査の実施</p> <ul style="list-style-type: none"> 心房細動の原因として心臓や弁膜症の評価が必要です。特に、左心房の拡大や心室の有病を確認するために、経胸壁心エコー検査を実施します。 	心エコーによる器質的因子の精査
<p>3. 抗凝薬投与の検討</p> <ul style="list-style-type: none"> 脳卒中のリスクであっても、CHA₂DSスコアやCHA₂DS₋VAScスコアを用いて血栓塞栓症のリスクを評価します。 高リスク患者（例：CHA₂DS₋VAScスコア≥2）では、抗凝薬（例：DOACやワルファリン）の投与を検討します。 	脳梗塞予防のためのリスク評価
<p>4. 電解質と甲状腺機能の評価</p> <ul style="list-style-type: none"> 電解質異常（特にカリウム、マグネシウム、カルシウム）や甲状腺機能亢進症が心房細動の原因となることがあるため、血液検査を実施します。 	心房細動の誘引探索
<p>5. 術後のストレスや炎症の評価</p> <ul style="list-style-type: none"> 術後のストレスや炎症が心房細動の原因となるため、CRPや他の炎症マーカーをモニタリングします。 	周術期における影響因子の探索

結論

この患者に対しては、テレメトリーモニタリングを継続することに加え、心エコー検査を実施し、心房細動の原因を評価することが推奨されます。また、抗凝薬投与の必要性を評価するため、血栓塞栓症のリスクスコアを計算し、必要に応じて抗凝薬を開始します。さらに、電解質や甲状腺機能の検査を行い、術後のストレスや炎症の影響をモニタリングします。これらの追加評価と治療により、心房細動の管理と血栓塞栓症の予防を図ります。

推奨される検査と治療の要約

図 34

脳梗塞発症リスクの適切な推定と推奨治療

7

この患者のCHA₂DS₋VAScスコアを正確に計算してください。

スコアシステムの正しい解説

<p>1. Congestive heart failure (心不全): 1点</p> <p>2. Hypertension (高血圧): 1点</p> <p>3. Age ≥75 years (年齢≥75歳): 2点</p> <p>4. Diabetes mellitus (糖尿病): 1点</p> <p>5. Stroke/TIA (脳卒中または一過性脳虚血発作の既往): 2点</p> <p>6. Vascular disease (血管疾患): 1点</p> <p>7. Age 65–74 years (年齢65–74歳): 1点</p> <p>8. Sex category (女性): 1点</p>	スコアシステムの正しい解説
--	----------------------

この患者の背景を考慮すると、以下のようになります：

- 年齢: 72歳 (65–74歳の範囲内)
- 性別: 女性
- 既往歴:
 - 2型糖尿病 (1点)
 - 高血圧 (1点)
 - 胆嚢穿孔性胆管炎 (血管疾患や心不全の記載なし)

したがって、CHA₂DS₋VAScスコアは以下の通りです：

年齢: 1点
性別: 1点
糖尿病: 1点
高血圧: 1点

合計スコアは4点となります。このスコアに基づき、この患者は中等度の血栓塞栓症リスクを有すると判断されます。

<https://my.clevelandclinic.org/health/diseases/embolic-stroke>

**心房細動により心臓にできた血栓が
脳血管を詰まらせることで脳梗塞が発症することがある**

Embolism

Embolism (blood clot)

Carotid artery

Embolism

Embolism (blood clot)

**脳梗塞発症リスクの推定にもとづき、
血栓塞栓症予防の治療導入を推奨**

Embolism

Embolism (blood clot)

Carotid artery

Embolism

Embolism (blood clot)

**脳梗塞発症リスクの推定にもとづき、
血栓塞栓症予防の治療導入を推奨**

図 35

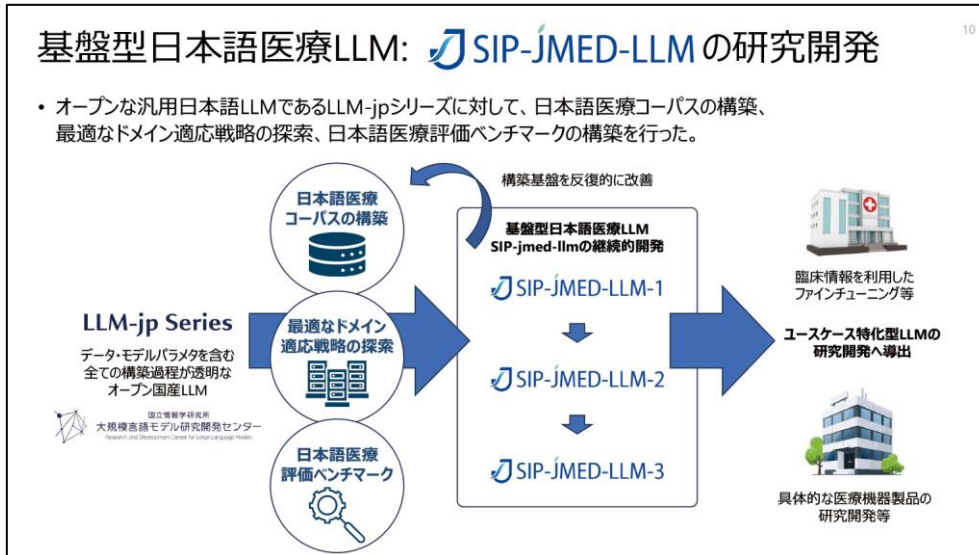


図 38

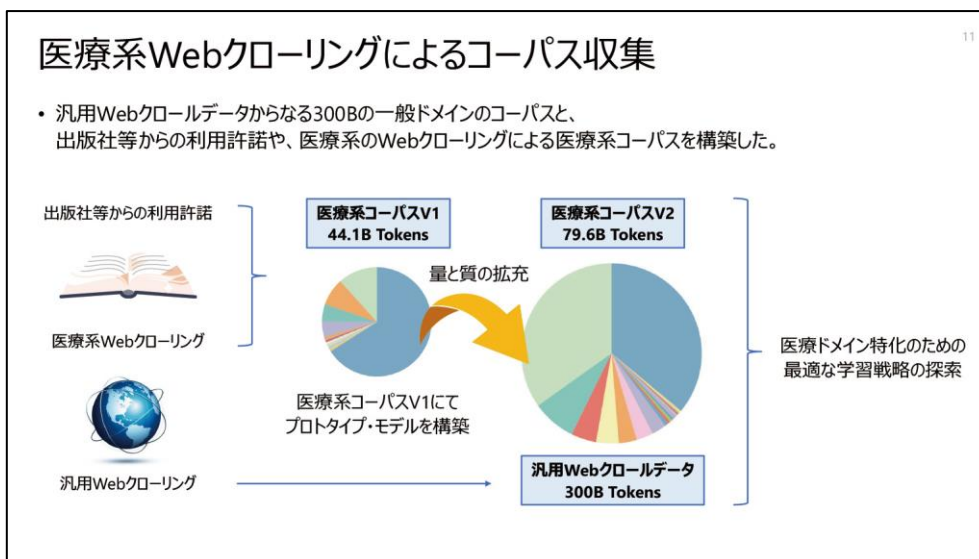


図 39

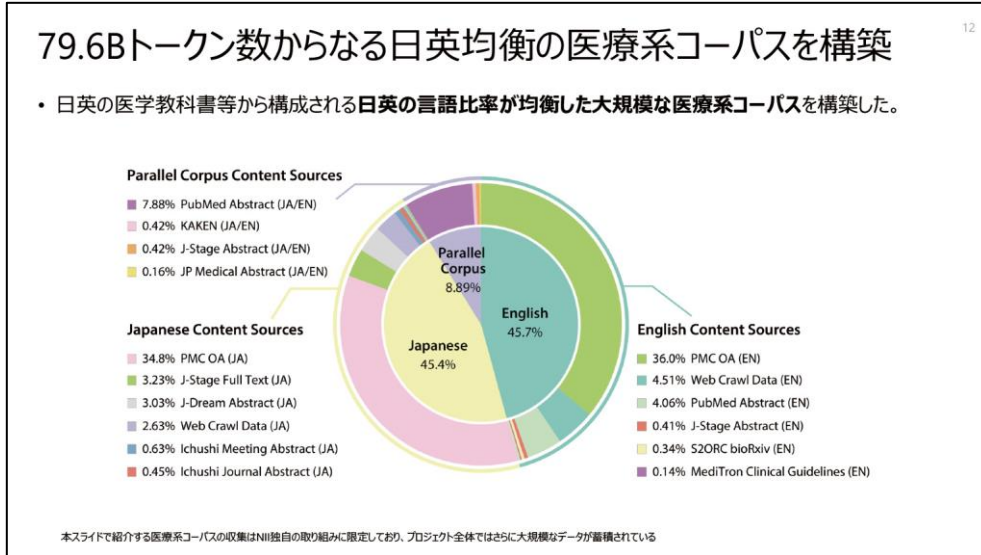


図 40

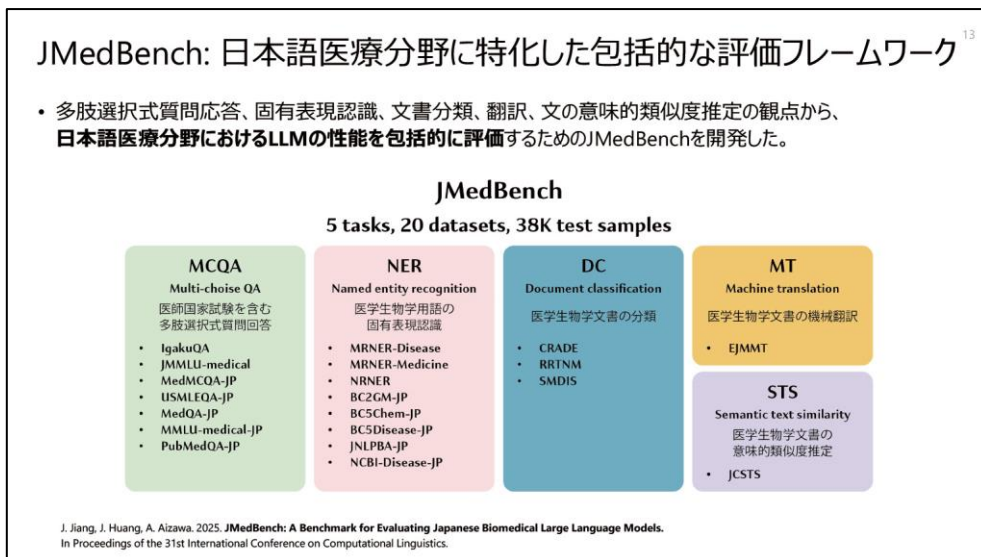


図 41

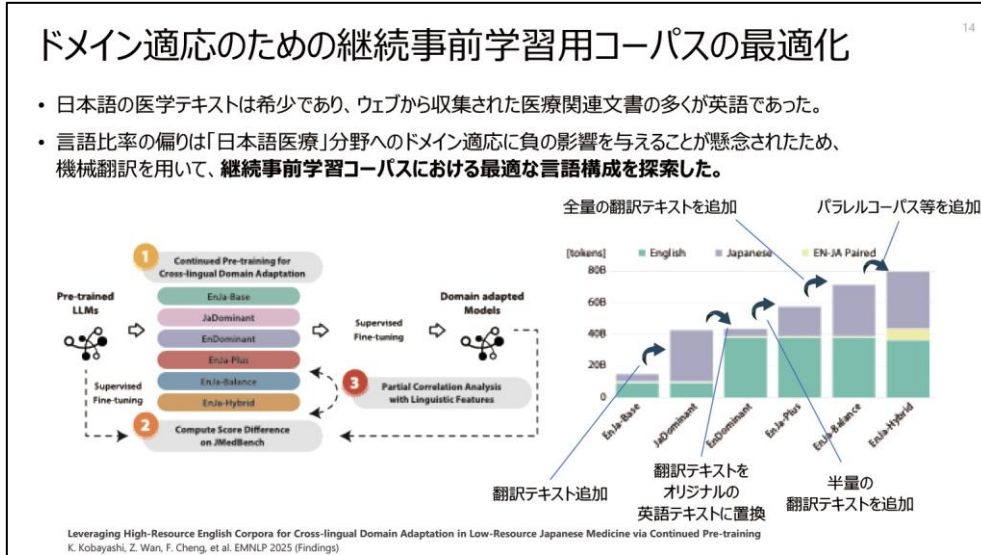


図 42

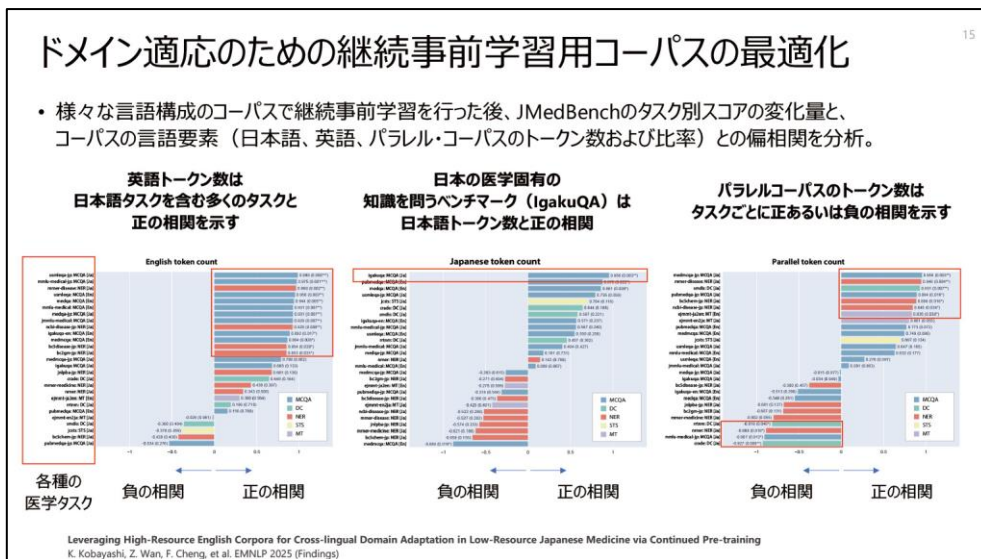


図 43

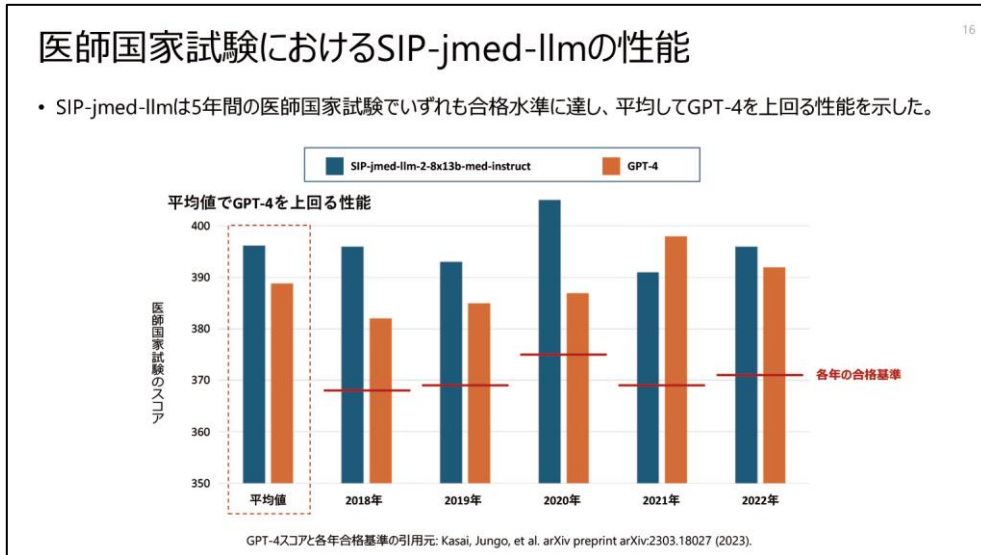


図 44

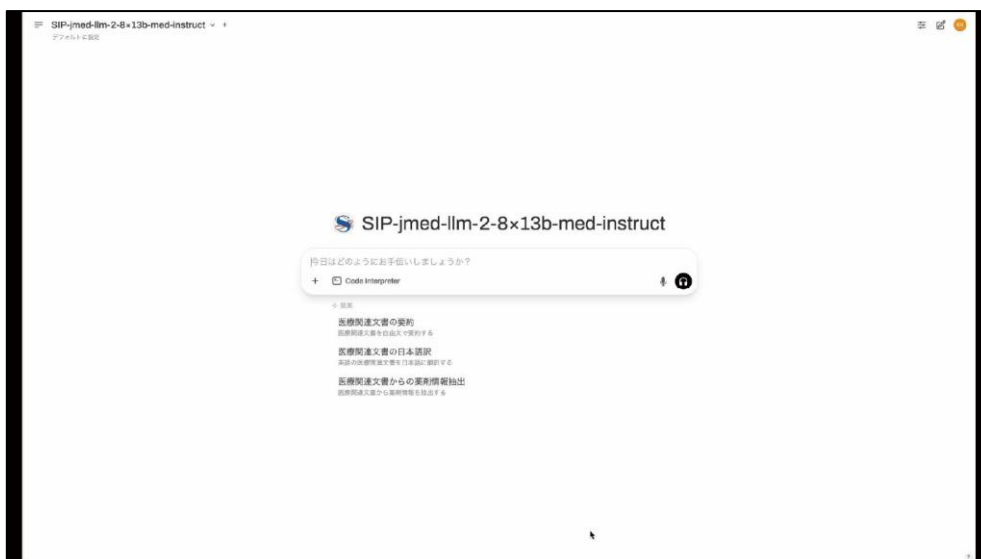


図 45

AIを社会の重要インフラに組み入れる際の倫理的課題

18

- 生成AIによる技術革新は、医療のあり方を根本から変革する力を有する一方で、容易に解決することのできない新たな倫理的・法的・社会的課題も提示している。

バイアスと差別 学習用データセットに含まれるバイアスが、AIシステムによって再生産されるリスク。

透明性の欠如 AIシステムがどのように意思決定・将来予測を行うの理解は難しく、ブラックボックスである。

プライバシーとデータ保護 プライバシーに配慮したデータ収集と、プライバシー・リスクを未然防止するためのセキュリティが必要。

雇用喪失と雇用転換 AIシステムによる自動化は雇用の喪失を含み、労働力のあり方に大きな変化をもたらす可能性がある。

アカウントビリティと法的責任 AIシステムによる自律的な判断から生じる危害に対して、責任の所在を判断するための枠組みを明確化する必要。

情報操作と誤情報 ディープフェイクの生成や情報操作、偽情報の流布などにAIシステムが利用される危険性。

セキュリティリスク 悪意者によってAIシステムがサイバー攻撃や他の望ましくない行為に悪用される危険性。

不平等とアクセス AIシステムからの恩恵が社会に不平等に分配されることで、社会的に疎外された集団に影響を与えるリスク。

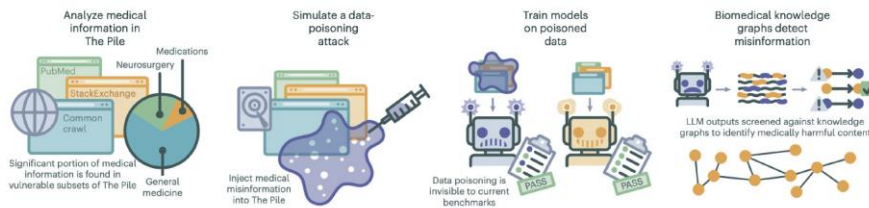
Omar Santos, Petar Randaniev. (中島 豊 訳) 2025. AIアルゴリズムからAIセーフティへ - 生成AIとLLM

図 46

医療LLMにおける脆弱性の例 - データ汚染攻撃

19

- 学習用データセットのわずか0.001%を偽情報に置き換えるだけで、医療LLMが有害な医学的誤りを含む文章を生成する可能性が著しく高まることが知られている。
- 更に、モデルの汚染を標準的な医学ベンチマークで検出することは困難であることも示唆されている。



Daniel Alexander Alber, et al. 2024. **Medical large language models are vulnerable to data-poisoning attacks.** Nature Medicine

図 47

20

機微情報を用いて学習された生成AIに係るELSI

- 従来、複数人の個人情報を機械学習の学習用データセットとして用いて構築された学習済みモデルにおいて、**内部パラメータと特定の個人との対応関係が排除されている限りにおいては、これは個人に関する情報に該当しないとされてきた。**
- この見解は識別モデルについては一般的に成立すると考えられるが、生成モデルについては必ずしも自明ではない。


個人情報保護委員会FAQ Q1-8

(個人情報)

Q1-8 複数人の個人情報を機械学習の学習用データセットとして用いて生成した学習済みパラメータは、個人情報に当たりませんか。


A1-8 複数人の個人情報を機械学習の学習用データセットとして用いて生成した学習済みパラメータ（重み係数）は、学習済みモデルにおいて、特定の出力を行うために調整された処理・計算用の係数であり、当該パラメータと特定の個人との対応関係が保持されている限りにおいては「個人に関する情報」に該当するものではないため、「個人情報」にも該当しないと考えられます。
(令和3年6月追加)

Training Set



Caption: Living in the light with Ann Graham Lotz

Generated Image



Prompt: Ann Graham Lotz

B. Balle, et al. Extracting training data from diffusion models. SEC '23: Proceedings of the 32nd USENIX Conference on Security Symposium Article No.: 294, Pages 5253 – 527 より引用。

図 48

21

まとめと今後の課題

- NIIを中心とする情報系チームでは、LLM-jpモデルをベースモデルとして、**安全性・信頼性を持つオープンな日本語医療LLMの研究開発に一気に通貫で取り組んできた。**
- 開発中のモデルは本邦の医師国家試験の合格基準に達しており、**スクラッチ・モデルに始まる全ての開発過程が透明化された国産LLMとして大きな快挙であると言える。**
- 現在、様々な実臨床のユースケースへの適合を志向した追加開発が展開されている。
- また、本プロジェクトを検証環境として活用することで、**医療LLMにおけるプライバシー・リスクの評価をはじめ、健全な知識循環と技術の社会受容に関する議論の場の創出が期待される。**

Building Effective Japanese Medical LLMs with an Open Recipe for Domain Adaptation through Continued Pre-training
A. Aizawa, Y. Arase, F. Cheng, J. Huang, Z. Huang, J. Jiang, T. Kanazawa, D. Kawahara, K. Kobayashi, T. Kodama, S. Kurohashi, Y. Oda, Y. Tsuta, Z. Wan, Z. Yang, R. Yokota. Under Submission.

図 49

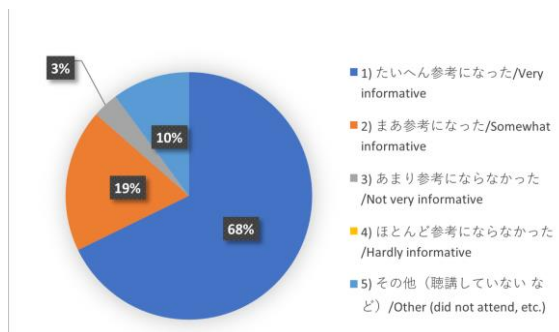
3 セミナー後の反響

3.1 アンケート

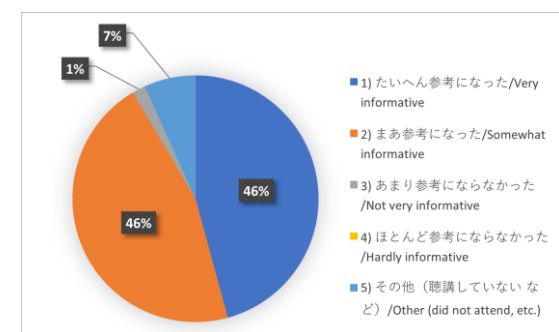
セミナー開催後、配信画面上でのアナウンスおよび参加申込者宛のメール送信によってアンケートへの回答を依頼した。開催終了後 1 週間後の回答数は 59 件であった。

◆講演内容について

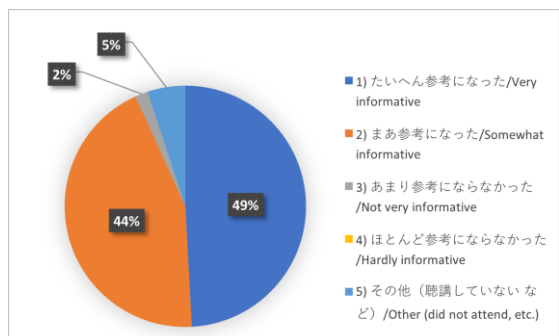
[講演 1]



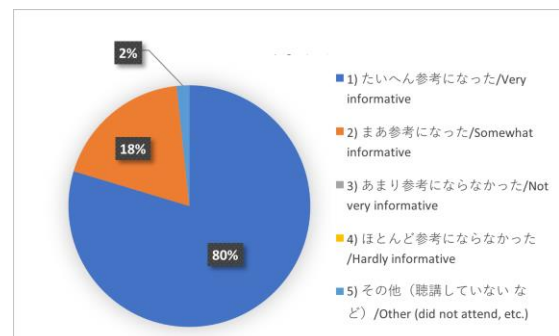
[講演 2]



[講演 3]



[講演 4]



◆セミナー全体についてのコメント

・医師であり、多くの医学論文の査読を依頼し、依頼され、かつ医学雑誌の編集長業務をしている身として、今回のセミナーは非常に有意義なものでした。とくに論文を作成支援する AI の位置、日本語の医学用語を大量に機械翻訳したものを学習内容として精度の高い医学・医療の構文を理解し生成できる AI 開発は大変参考になりました。その一方で、編集業務、とくに論文査読に AI を活用できていないのは非常に片手落ちな印象を受けました。エルゼビア社のように閉じた LLM を構築し、それが査読にも利用できるようにまもなくであることは大変心強く思います。技術開発は万人のためでもあり同時に、開発者の利益に還元されていかななくてはならないことも理解しますが、AI がなければ論文作成も論文査読もできなくなるような時代がきたときに、publication がもつ意義(「研究は自己修正だ」というのはいい定義だと思いました)をつねに考え続ける必要があると思いました。

・生成 AI 使用におけるルール設定について学びたいと思うが、このテーマについては 1 年で状況が大きく変化することが予想されるため、今後も定期的に情報提供の機会をいただけると幸いです。

・私は当会学術誌編集事務局として生成 AI 使用のガイドラインを現在最終確認中です。今日のお話にもあったように、まだまだそれぞれの学会や編集局の単位でのことが多く、非常に難しく感じています。参考になりました。

・ビジネス系の汎用 AI の話は多いですが、研究に沿った AI の情報は意外と少ないように思います。今後も、深掘りしてもらいたいですし、上手く使っている方のノウハウなどを日本全体に展開できれば良いのではと感じます。分野によって、使う AI も使い方も違うように思いますので、分野別にセミナーやワークショップがあればと思います。

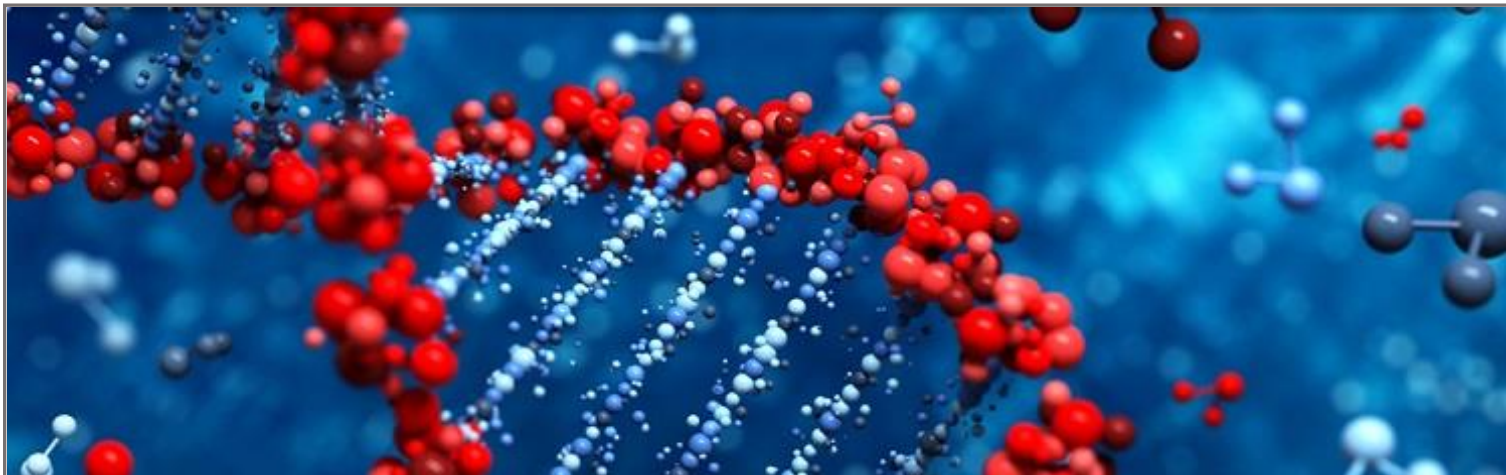
4 公開資料

本セミナーの当日配布資料を、下記 URL に掲載している。

<https://www.jstage.jst.go.jp/static/pages/News/TAB4/PastIssues/-char/ja#251126>

また、当日の講演の動画を J-STAGE YouTube チャンネルでアーカイブ公開している。

https://www.youtube.com/playlist?list=PLSXgr8_ZxZT3nAzN5B9daMAPb4NvZmzGc



2025 年度第 1 回 J-STAGE セミナー報告書

令和 8 年 2 月 6 日発行

国立研究開発法人科学技術振興機構 (JST)

情報基盤事業部 研究成果情報グループ

〒102-8666 東京都千代田区四番町 5-3 サイエンスプラザ

E-MAIL : contact@jstage.jst.go.jp

© 2026 Japan Science and Technology Agency

この文書はクリエイティブ・コモンズ [表示 4.0 国際] ライセンスの下に提供されています。

<https://creativecommons.org/licenses/by/4.0/deed.ja>
