Scholarly Communication, AI Strategies and IEEE

Paul Canning
Director International Sales
Taipei
15 November 2024

IEEE
Advancing Technology for Humanity

AI isn't new…it has been evolving since the mid 20th Century

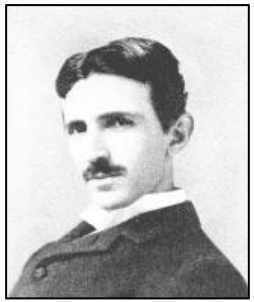
- A. Uttley, **"Discussion on Dr. A.M. Uttley's paper "Information, Machines, and Brains"**, in Transactions of the IRE Professional Group on Information Theory, vol. 1, no. 1, pp. 193-197, Feb. 1953, doi: 10.1109/TIT.1953.1188576.
- C. E. Shannon, "**Computers and Automata**," in Proceedings of the IRE, vol. 41, no. 10, pp. 1234-1241, Oct. 1953, doi: 10.1109/JRPROC.1953.274273.
- A. Newell and H. Simon, "**The logic theory machine--A complex information processing system**," in IRE Transactions on Information Theory, vol. 2, no. 3, pp. 61-79, September 1956, doi: 10.1109/TIT.1956.1056797.
- V. H. Yngve, "**The technical feasibility of translating languages by machine**," in Transactions of the American Institute of Electrical Engineers, Part I: Communication and Electronics, vol. 75, no. 6, pp. 792-797, Jan. 1957, doi: 10.1109/TCE.1957.6372595.
- R. M. Friedberg, "**A Learning Machine: Part I**," in IBM Journal of Research and Development, vol. 2, no. 1, pp. 2-13, Jan. 1958, doi: 10.1147/rd.21.0002.
- E. E. David, **"Artificial Auditory Recognition in Telephony,"** in IBM Journal of Research and Development, vol. 2, no. 4, pp. 294-309, Oct. 1958, doi: 10.1147/rd.24.0294.

Reference: IEEE Xplore.

The Beginnings of IEEE are Based on the Adoption of New Technologies

Thomas Edison

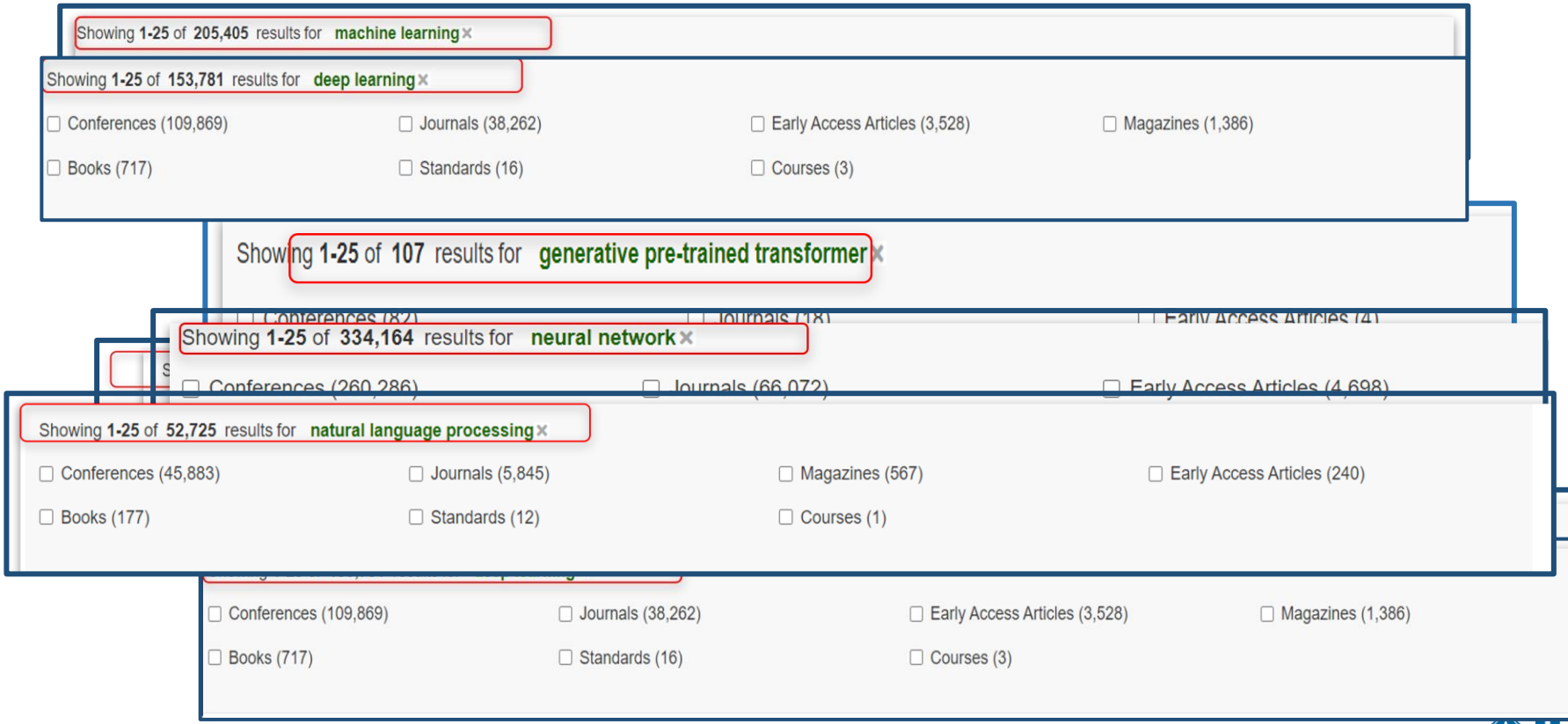
Alexander Graham Bell



Nikola Tesla

Landmark exhibition at Franklin Inst. in Philadelphia, the American Institute of Electrical Engineers held its first conference on 7–8 Oct. 1884. This was the first formal technical conference on electrical engineering held in the U.S.

AI Covers Many Key Words…



IEEE Journals covering Artificial Intelligence Topics

- IEEE Transactions on Pattern Analysis and Machine Intelligence
- IEEE Transactions on Neural Networks
- IEEE Transactions on Neural Networks and Learning Systems
- IEEE Transactions on Systems, Man, and Cybernetics
- IEEE Transactions on Knowledge and Data Engineering
- IEEE Transactions on Cybernetics
- IEEE Transactions on Fuzzy Systems
- IEEE Transactions on Image Processing
- IEEE Internet of Things Journal
- IEEE Sensors Journal
- IEEE Systems Journal
- IEEE Transactions on Intelligent Transportation Systems
- IEEE Transactions on Evolutionary Computation

IEEE

Reference: The Value of IEEE Research for Artificial Intelligence Companies

IEEE Conferences covering Artificial Intelligence Topics

- International Conference on Artificial Intelligence and Computational Intelligence
- International Conference on Artificial Intelligence, Management Science and Electronic Commerce
- World Congress on Intelligent Control and Automation
- International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing
- International Conference on Automatic Control and Artificial Intelligence
- International Conference on Machine Learning and Cybernetics
- IEEE/CVF Conference on Computer Vision and Pattern Recognition
- International Joint Conference on Neural Networks (IJCNN)
- International Conference on Artificial Intelligence and Data Processing
- IEEE International Conference on Big Data
- IEEE International Conference on Fuzzy Systems

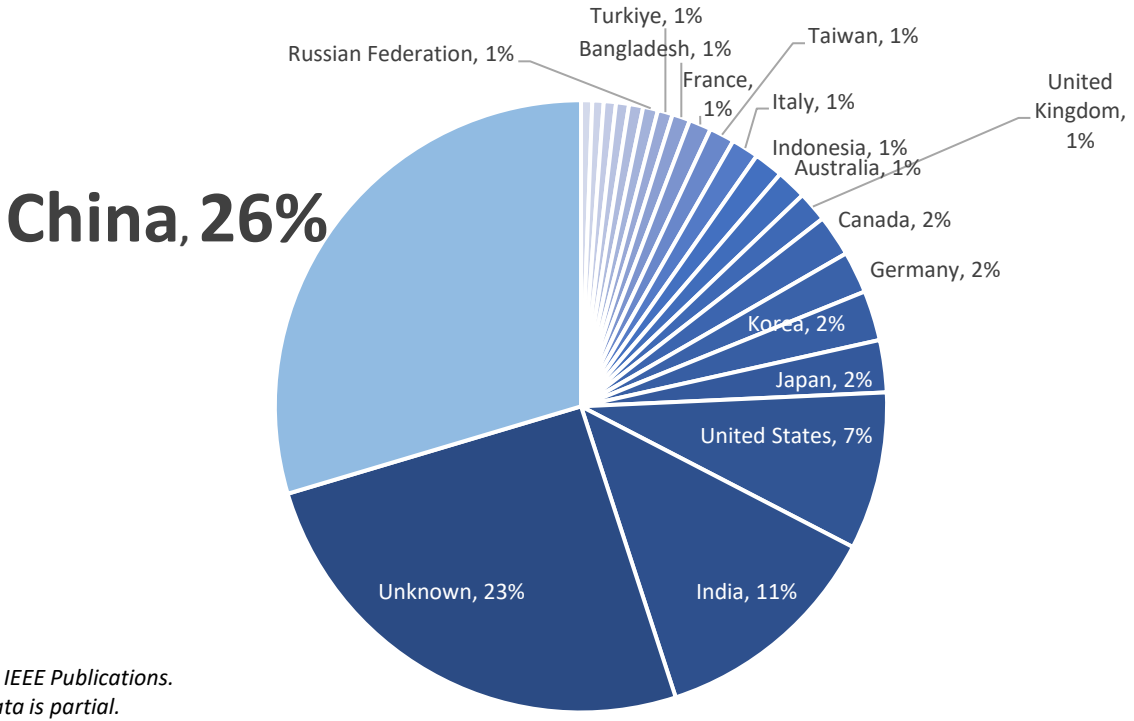
IEEE

Reference: The Value of IEEE Research for Artificial Intelligence Companies

A Sample of IEEE Standards Covering Artificial Intelligence

- **IEEE 1232.3** IEEE Guide for the Use of Artificial Intelligence Exchange and Service Tie to All Test Environments (AI-ESTATE)
- **IEEE 1855** IEEE Standard for Fuzzy Markup Language
- **IEEE 610.4** IEEE Standard Glossary of Image Processing and Pattern Recognition Terminology
- **IEEE 1872** IEEE Standard Ontologies for Robotics and Automation
- **IEEE 2755** IEEE Guide for Terms and Concepts in Intelligent Process Automation
- **IEEE P3301/D2** Adoption of Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) Technical Specification Artificial Intelligence Framework (AIF)
- **IEEE 2801-2022** IEEE Recommended Practice for the Quality Management of Datasets for Medical Artificial Intelligence

AI-Related IEEE Publishing Global Author Landscape(2022)

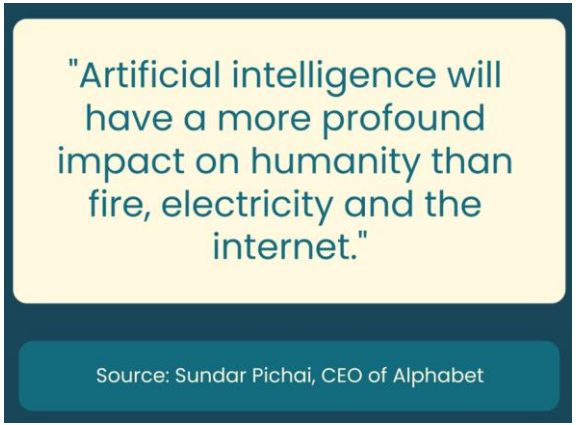


Data Source: October 2023 IEEE Publications.
**2023 publishing output data is partial.*

AI's Unprecedented Impact

Industry & Global Response to AI

- AI has outpaced existing copyright, IP & privacy laws
- Global efforts to update laws
- Many opportunities to increase revenue, streamline operations, improve efficiencies
- But businesses also need to evaluate risks and protect IP



IEEE Policy Action IEEE的政策行動

Publications Board Operations Manual:
Section 8.2.1.B Responsibilities of Authors of Articles Published by IEEE

10. The use of content generated by artificial intelligence (AI) in an article (including but not limited to text, figures, images, and code) shall be disclosed in the acknowledgments section of any article submitted to an IEEE publication. The AI system used shall be identified, and specific sections of the article that use AI-generated content shall be identified and accompanied by a brief explanation regarding the level at which the AI system was used to generate the content. The use of AI systems for editing and grammar enhancement is common practice and, as such, is generally outside the intent of the above policy. In this case, disclosure as noted above is recommended.

文章中人工智慧 (AI) 產生的內容（包括但不限於文字、圖形、圖像和程式碼）的使用應在提交給 IEEE 出版物的任何文章的致謝部分中揭露。應指明所使用的人工智慧系統，並應明列文章中使用人工智慧生成內容的特定部分，並附有有關使用人工智慧系統生成內容的等級的簡要說明。 使用人工智慧系統進行編輯和語法增強是常見的做法，因此通常超出了上述政策的意圖。在這種情況下，建議進行上述揭露。

IEEE Policy Action IEEE的政策行動

Publications Board Operations Manual:
Section 8.2.1.C Reviewers of Articles*

Information or content contained in or about a manuscript under review shall not be processed through a public platform (directly or indirectly) for AI generation of text for a review. Doing so is considered a breach of confidentiality because AI systems generally learn from any input.

不得通過公共平台（直接或間接）處理受審稿件中包含的或關於受審稿件的信息或內容，以供人工智能生成審稿文本。這樣做被認為是違反保密性的，因為人工智能系統通常會從任何輸入中學習。

**The same text was inserted in Section 8.2.2 which covers the review process in detail*

IEEE Policy Review & Next Steps IEEE的政策行動

Publications Board AdHoc Committee on AI in Publishing

下一步—出版委員會出版業人工智能特設委員會

- Beyond disclosure – What are the appropriate and inappropriate uses of large language models and other AI tools? 超越當前披露的視野—大型語言模型和其他人工智能工具的適當和不適當的使用是什麼？
- Citation of machines and artificial data – What constitutes a legitimate source? 機器和人工數據的引用—什麼構成合法來源？
- Downstream implication of disclosure – Does disclosure of the allowable use of AI for grammar improvement cause bias against non-native English speakers in the review process? 披露的下游影響—披露允許使用人工智能來改進語法是否會在審查過程中對非英語母語人士不利？
- 「Closed-loop」 language models – Is it possible to overcome confidentiality concerns by avoiding public platforms? 「閉環」語言模型——是否可以通過避開公共平台來克服保密問題？

AI Strategy from a Scholarly Publisher Perspective

- Support customer and author needs while evaluating risk
- Protect the scholarly record & ecosystem (i.e. Author Output, IP)
- Continue to support the discovery, reading, and citing of full text
- Keep learning and evolving
- Work to build our own AI tools and IEEE, perhaps, LLM

AI will improve Accessibility of Scholarly Databases

IEEE *Xplore* Accessibility Statement

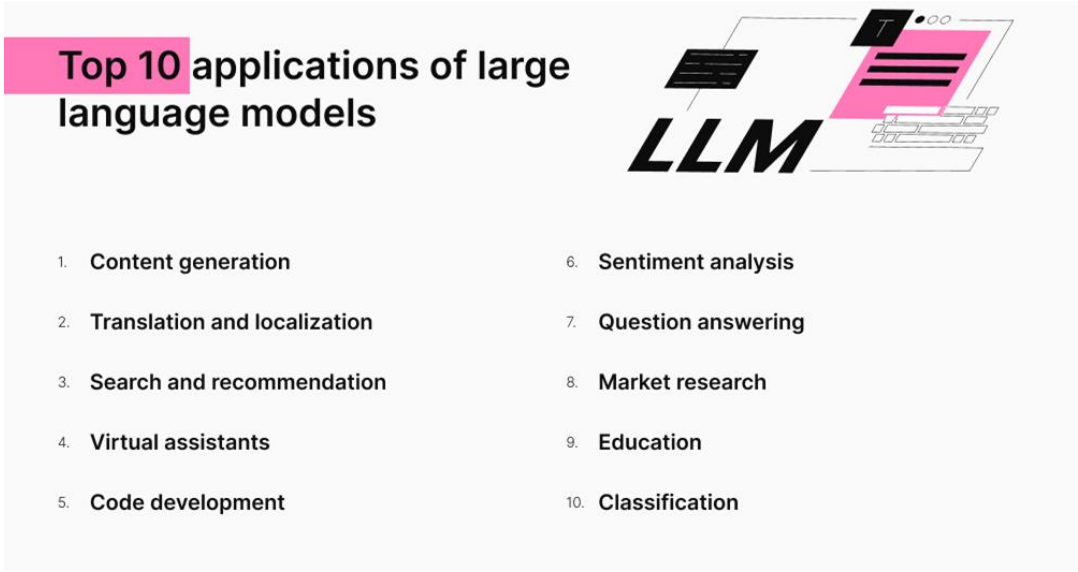
- IEEE strives to ensure that IEEE *Xplore* provides the best possible experience for all users.
- Our accessibility efforts are informed by the Web Content Accessibility Guidelines (WCAG) 2.2 from the World Wide Web Consortium (W3C) as well as the European Accessibility Act and Section 508 of the United States Rehabilitation Act.

Documentation ***(Updated: October 2024)***

- **VPAT (Voluntary Product Accessibility Template)**
  - Details out our level of conformance for each WCAG 2.2 (A+AA) accessibility guideline.
- **Accessibility Roadmap**
  - Outlines accessibility fixes recently completed and plans for fixes over the next two quarters.

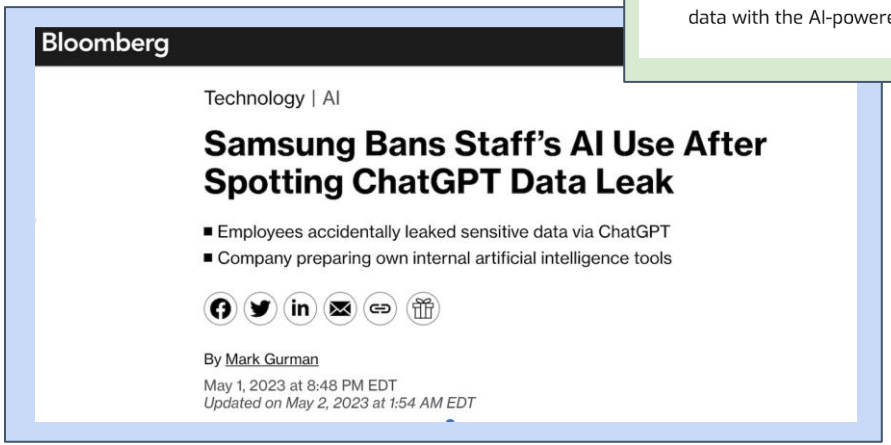
Commercial Users Want to use Scholarship for AI Use Cases

- Use tools like Microsoft Co-Pilot
- Applications analyzing trends and providing predictive analytics, data visualization, and others
- Index IEEE content for customized search/GPT applications
- Integrate IEEE content into large language models for summarization, classification, search and recommendation, question and answering, translation



https://pixelplex.io/blog/llm-applications/

Despite the excitement…
Some missteps and concerns by Big Tech



Example: Controlling Web Scraping of Content

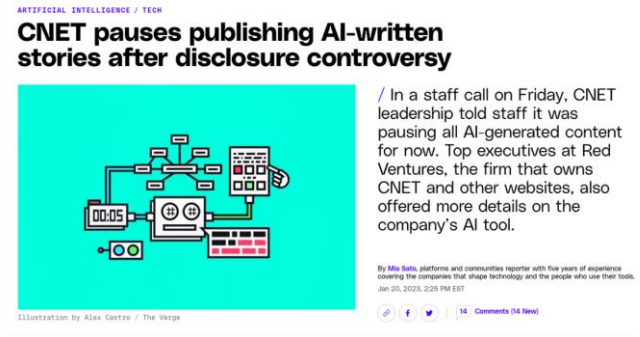
The New York Time Lawsuit

- **Background**:
  - *Case*: The New York Times (NYT) sued OpenAI over the unauthorized use of their content to train AI models.
  - *Allegation*: OpenAI used NYT's copyrighted articles without proper licensing to train its language models.
  - *Concerns*: The use of copyrighted content without permission could lead to revenue loss, copyright infringement, and harm to journalism.

Example: AI Hallucinations

CNET's AI Issues

- **Impact**:
  - CNET published articles generated by AI that contained factual errors.
  - Backlash from readers, harm to editorial reputation, financial blowback.
  - Internal review and editing costs escalated



ARTIFICIAL INTELLIGENCE / TECH

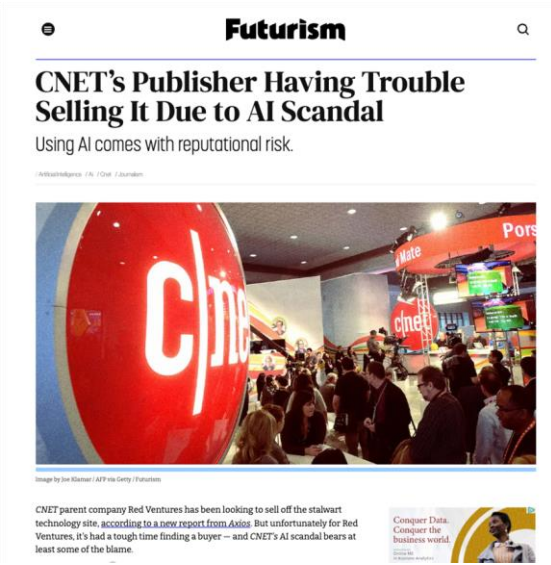
CNET pauses publishing AI-written stories after disclosure controversy

/ In a staff call on Friday, CNET leadership told staff it was pausing all AI-generated content for now. Top executives at Red Ventures, the firm that owns CNET and other websites, also offered more details on the company's AI tool.

By Mia Sato, platforms and communities reporter with five years of experience covering the companies that shape technology and the people who use their tools.
Jan 20, 2023, 2:25 PM EST

Illustration by Alex Castro / The Verge

14 Comments (14 New)



Futurism

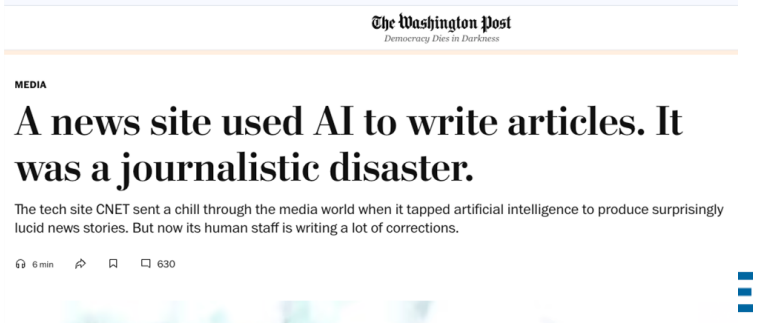
CNET's Publisher Having Trouble Selling It Due to AI Scandal

Using AI comes with reputational risk.

Image by Joe Klamar / AFP via Getty / Futurism

CNET parent company Red Ventures has been looking to sell off the stalwart technology site, according to a new report from Axios. But unfortunately for Red Ventures, it's had a tough time finding a buyer — and *CNET*'s AI scandal bears at least some of the blame.

Conquer Data. Conquer the business world.



The Washington Post
Democracy Dies in Darkness

MEDIA

A news site used AI to write articles. It was a journalistic disaster.

The tech site CNET sent a chill through the media world when it tapped artificial intelligence to produce surprisingly lucid news stories. But now its human staff is writing a lot of corrections.

6 min 630

Example: Authorship and Publishing Ecosystem Questions

Compensation and Incentives

- *Example*: Publisher licensing deals with large commercial AI tools
- *Impact*:
  - Revenue opportunity for publishers
  - Authors may be asking for new terms in the future:
    - AI royalties
    - Opt out options
  - Effect on other incentives like citations
  - Protections in the terms with AI companies can help

Academic authors 'shocked' after Taylor & Francis sells access to their research to Microsoft AI

NEWS JUL 19, 2024 BY MATILDA BATTERSBY

"We are at a crossroads in the production and dissemination of research knowledge, and in my view the biggest problem with this deal is the reduction of academic research into raw content from which data can be extracted and repackaged as knowledge," Clemens said.

Example: Fair Use

Libraries citing fair use for AI training

- *Example*: University of California Memo on Fair use
- *IEEE thoughts*:
  - Fair use allows for limited use of copyrighted content but depends on factors like the purpose, amount used, and its impact on the market. It's not a blanket exception.
  - Training AI models, especially large language models (LLMs), requires large-scale data ingestion, which goes beyond what fair use permits.
  - Using extensive copyrighted material to train AI can also hurt the market value of the original work, which is a key factor against fair use.
  - Additionally, the way AI uses data—automatically and on a massive scale—is very different from the smaller, more specific uses that fair use normally covers, like educational purposes.

How IEEE Evaluates AI Use Requests for its IP

- The purpose of the project
  - What does the end product look like?
- The end users of the AI tool and/or the Output of the AI
  - Internal to the organization? Commercial Product for multiple organizations? Publicly available?
- IEEE content is required
  - How much? Metadata? Full Text?
- Scope of full content set being ingested
  - Internal proprietary documents? Other scholarly publishers? Web scraped information?
- The security of the data, the tool, and the Output
  - What tools are being used and where are they hosted?
- Whether the IEEE content will be used for training
  - Training or non training/Retrieval Augmented Generation set up?
- Prominence of the referenced source work
  - Ensuring the referenced works with links to source are still paramount

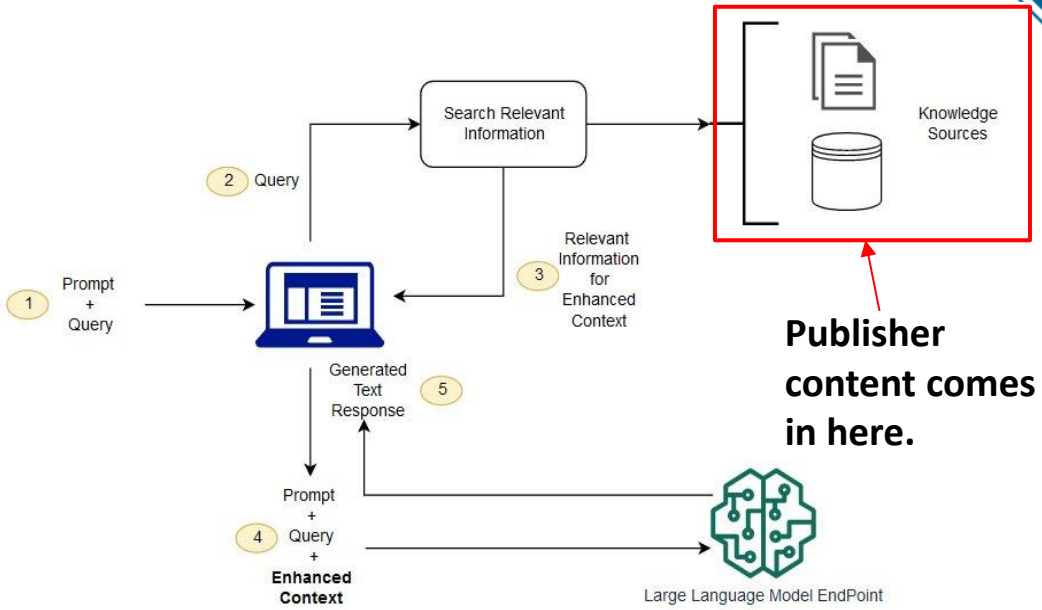
Training vs Non Training?

When the content is used for **training an AI model:**

- Once the content is ingested into the model, it cannot be removed
- We need to consider this in the risks and the price

When the content is **not used for training (non-train):**

- Most common method: Retrieval Augmented Generation (RAG)
- References authoritative knowledge base outside of its training sources before generating a response
- Content can be removed from the workflow at any time



https://aws.amazon.com/what-is/retrieval-augmented-generation/

Practical Tips for Managing the Scholarly Record

- Post AI permissions/restrictions on website terms of use
- Be clear in contracts/agreements what is permitted and what is not, including the remedies/penalties for breach
- Update robot.txt files and follow instructions for blocking web scraping from the various players (OpenAI, Google)
- Understand how AI tools work and what the output will look like
- Maintain the right to test/audit AI where your content is ingested
- Retrieval Augmented Generation Models can help with references and hallucinations
- Have open conversations with customers about what they want to do and what your needs are as a publisher

We are living and working in interesting times…

- **Clear AI Policies are Essential**: Establishing clear AI terms helps publishers keep control of the content and rights.
- **Balance Innovation and Protection**: Strategy should strike a balance, allowing for responsible AI use and increased opportunities while continuing to support sustainability of the ecosystem and trustworthy content.
- **Continue to Adapt:** The space is evolving rapidly so be prepared to change the strategy as the technology and market changes.

Questions for the Academic, Scholarly Community

- What are your AI strategies and concerns?
- Any opportunities or risks that are top of mind?
- Where do you see ways to mitigate the risks mentioned?
- What's your view of Fair Use from the author's perspective?
- Have your organizations developed AI policies?
- Do you have other practical tips to share?

Many Thanks for Another Warm Welcome at CONCERT 2024!

Paul Canning

Director International Sales

p.canning@ieee.org