Use of STM Content with Generative AI Platforms – Copyright Considerations

Question: Can I upload a journal article for analysis into generative artificial intelligence? Answer: It's complicated. Read on to find out more.

Introduction

We are all learning how to most effectively and ethically make use of generative artificial intelligence (gen AI) applications in our work, and it is clear that these tools are extremely powerful for discovering and analyzing content. This article focuses specifically on the use case of searching for, finding, and uploading Scientific, Technical and Medical (STM) journal and eBook content into gen AI applications for analysis and summarization. It highlights the main use cases, STM publisher restrictions, and the evolving solutions to ensure users can make use of AI tools with STM content in a copyright-compliant manner.

When it comes to the use of STM content for commercial purposes (by and for commercial organisations), whether that is for internal research and development at a pharmaceutical company, preparing a regulatory submission, or for a medical communications project, your ability to do this with AI applications will most likely be restricted by copyright law. STM publisher licenses and terms and conditions clearly state that their content cannot be used in major AI applications like ChatGPT, Microsoft Copilot, Claude, Gemini, DeepSeek, etc. without express permission. This also includes applications focused on scientific and medical workflows, such as Elicit or Consensus, which do not have license agreements with publishers and put the onus on users to know if they have the required rights to upload content to their solutions. The reason is that without controls on this use, these models will invariably consume the content and make use of it for training and evaluation of the AI models, often becoming permanently incorporated in the models.

There is, however, an acute need to upload content into gen AI applications. Many, if not most, corporate researchers, information professionals, and medical writers now use gen AI applications in their work and want to use them to analyze and summarize STM content. In many cases, those in a corporate environment have an "Enterprise" license to these applications, which ensures that none of the content analyzed goes back to train the model or be permanently incorporated into it. However, the infrastructure for easily providing the copyright permissions is lagging behind the technology by several years.

Copyright licensing programs

Recently, the Copyright Clearance Center (CCC) and the Copyright Licensing Agency (CLA) have both developed licensing solutions that allow corporate researchers and medical communication professionals to legally use gen AI applications with copyrighted scientific journal articles. However, note that jurisdictional limitations and exceptions can apply that require additional agreements with national Reproduction Rights Organizations, such as VG Wort in Germany, Centre Français d'exploitation du droit de Copie (CFC) in France, or the Japan Reprographic Rights Center (JRRC). In July 2024, CCC introduced AI re-use rights within its Annual Copyright Licenses (ACL), making it the first-ever collective licensing solution for the

internal use of copyrighted materials in AI systems. This license provides businesses with harmonized rights to legally integrate copyrighted content into AI applications, such as Microsoft Copilot and ChatGPT, ensuring compliance with copyright laws while compensating rightsholders. Similarly, the CLA has expanded its business and public sector licenses to include workplace AI permissions, enabling organizations to use copyrighted content as prompts within enterprise gen AI tools.

These collective licenses address a growing need for responsible AI usage in research and corporate settings. The CCC's licensing framework ensures that AI systems can analyze, summarize, and extract insights from licensed materials for internal purposes while maintaining legal and ethical standards. The CLA's AI permissions provide similar benefits within the UK jurisdiction, allowing corporate users to upload journal articles into AI platforms for internal use. While the CLA's licenses apply primarily in the UK, CCC's ACL licenses are designed for multinational businesses, covering a broad range of rightsholders across various jurisdictions. However, it is important to note that neither licensing program explicitly names any major STM publishers as participants.

Corporate licensing professionals and librarians, as well as representatives of STM publishers, report that the majority of STM content is not covered by these licenses. So, while these collective licenses offer a structured, transparent framework for AI use, their effectiveness in STM research depends on which publishers choose to participate. Therefore, it is imperative that users and organizations verify their licensing coverage with CCC and CLA before integrating AI into their research workflows.

Publishers

In the meantime, major publishers are engaging with customers to extend required rights to use their content with gen AI applications, although reportedly under very specific conditions, which can entail lengthy negotiation and licensing processes. While STM journal content is increasingly published under Open Access licenses, the pace has slowed so that the majority of new content is still not published Open Access. In addition, publishers have either already adopted or are increasingly adopting Open Access licenses that restrict commercial use. While a user can freely use journal content published under a CC-BY license with AI applications (with attribution for derivative content), users from commercial organizations, including pharmaceutical companies and communications agencies, cannot do so with content published under a CC-BY-NC (Non-Commercial) License. Therefore, users need to check with their content licensing colleagues (if they have them) or verify for themselves what permissions their organization has from publishers.

So, there is a significant gap between what copyright agencies are trying to cover and what publishers are starting to cover. Potential solutions exist and are under development to help fill that gap. When it comes to discovering, analyzing, and summarizing content, there are a number of STM-workflow-specific AI assistants and applications. However, they are either limited in coverage of journal full-text, only focused on metadata and abstracts, or may allow search of full-text but cannot surface or summarize content. When evaluating such tools, it is, therefore, critical to assess what rights the solution has organized with publishers to allow analysis and summarization of full text. An accurate list of their rights gap between what is provided

from publishers and copyright agencies. A lot of innovation and partnerships are developing in this space, and we can expect to see new solutions appearing in the months and years to come.

What to do next?

While the pace of copyright licenses, business models, and solution development to address the needs of users is picking up, it lags behind that of the technology by several years. This requires a mix of solutions to ensure copyright compliance when using STM content with gen AI applications. The checklist and decision tree included here, which provide a description and visualization, can be leveraged as part of a workflow to ensure compliance across this mix of solutions.

Checklist

- 1. Is the content Open Access (OA)?
 - \rightarrow If yes: Proceed to Step 2.
 - \rightarrow X If no: Proceed to Step 3.
- 2. Is the OA license CC-BY or CC-BY-NC?
 - → 🗹 If yes:

 \rightarrow CC-BY: You can generally use the content with attribution.

 \rightarrow CC-BY-NC: You may use the content non-commercially with attribution, but will require additional AI rights for use with gen AI and re-use rights for marketing, regulatory submissions, etc.

 \rightarrow X If no or unclear: Action – Review the specific license terms or consult legal to verify permissible uses.

3. Does your organization have a copyright license with a Reproduction Rights Organization (RRO; e.g., CCC or CLA) that includes AI rights and covers the publishers you need?

 \rightarrow If yes: You can likely use the content under that license – **confirm it includes AI** use cases.

 \rightarrow X If no: Proceed to Step 4.

4. Has your organization acquired AI rights directly from the publisher?

 \rightarrow If yes: You are covered – ensure the rights are documented and clearly scoped for your AI use.

 \rightarrow X If no: Proceed to Step 5.

- 5. Can a proprietary solution (e.g., licensed data platform, aggregator, etc.) provide either:
 - a. Copyright-compliant and attributed content? OR
 - b. The ability to acquire the necessary AI rights?
- \rightarrow If yes: Use the proprietary solution to stay compliant.
- \rightarrow X If no: Action You should not use the content with gen AI until proper rights are acquired.

Decision tree



Conflict of Interest: Chris Bendall is VP of Product Strategy at Research Solutions Inc.

This article was developed under the auspices of the International Society for Medical Publication Professionals (ISMPP) AI Task Force (May 2025).

References:

CCC Announcement of the AI Rights Updated to their Annual Copyright License – July 2024

<u>CLA Announcement of their intent to include AI Rights in their Copyright License</u> – November 2024

JISC report on the development of Open Access – March 2024

Creative Commons License Terms