

Working Group on Scholarly Content & Generative AI

RECOMMENDATIONS

Introduction

The Working Group on Scholarly Content and Generative AI was charged by the Provost, the VPR, and OGC with helping the MIT community develop a common understanding of the legal, technical and ethical issues involved in the use of scholarly content in the development and training of AI models. Specifically, the working group was asked to produce a set of resources, guidance, and principles to inform decisions made by individual MIT scholars and by MIT entities who publish, steward, or otherwise manage scholarly content.

In our recommendations, we have attempted to strike a balance between supporting the generation of human-authored scholarly content and fostering a flexible structure that allows scholarly content to be used for AI purposes in research and teaching. Importantly, we also recognize that the ongoing improvement of AI systems depends on sustaining the publication of curated, trustworthy science and scholarship.

Recommendations

- In acquiring access to scholarly content for the MIT Community, the MIT Libraries should seek terms that include broad computational access for research and teaching, as articulated in the [MIT Framework](#). Working with the Faculty Committee on the Library System (CLS), the MIT Libraries should review and update the [MIT Framework](#) to ensure it adequately addresses the needs of MIT researchers to have rights to train LLMs and AI tools on the third party content licensed for non-commercial research and teaching uses.
- The working group recognizes the potential danger that AI poses for replicating, replacing, or significantly diminishing the demand for and use of human generated scholarly content. Simultaneously, however, consistent with MIT's commitment to the value of openly sharing MIT's research and teaching with the world, the working group supports continuing to facilitate access to MIT content by all researchers, and the public at large. Members of the working group representing offices that steward MIT content and those responsible for publishing scholarly content acknowledge that there may be differences in the preferred and/or recommended terms and conditions for accessing and using content for AI purposes.
 - Therefore, the working group recommends that the Provost establish a small sub-group to develop and coordinate best practices—where feasible and to the extent aligned with the goals of all stakeholders—for seeking and granting rights

to scholarly content for AI purposes. The group should include representatives of The MIT Libraries, MIT Open Learning, The MIT Press, Office of the General Counsel, and the Faculty Committee on the Library System.

- MIT encourages a culture of seeking consultation prior to the use of any content for AI purposes. Taking time to understand the source of the content and whether its owners, authors and/or publishers have any associated rights is consistent with MIT's mission and values. MIT community members should use best efforts to comply with existing law on the use of scholarly content. Separately, but equally important, thoughtful and careful consideration of how scholarly content is used for AI purposes is consistent with academic norms (e.g., citation, attribution).
 - MIT community members should use best efforts to understand when permission is necessary, when fair use may instead be relied upon and when certain licensing terms apply. (For an extended discussion on this topic see [Machine Learning with Scholarly Content](#) and [MIT Working Group on Scholarly Content and Gen AI: White Paper.](#))
 - MIT community members should be particularly sensitive to how they use written material and third-party content licensed to MIT.
 - Training should be offered on a School or Department basis on the basics of accessing and using scholarly content, and appropriate consultations on the right to use. Schools and Departments should work with the MIT Libraries to provide such training.
- When collaborating with industry and other institutions, MIT community members should seek transparency about how scholarly content is accessed and used for research purposes and how scholarly content generated in the course of research will be used. Collaborations should establish norms at the beginning of a proposed project.
- MIT should consider leveraging its thought leadership to collaborate with those AI companies that have existing relationships with MIT in order to advocate for such companies to protect the varied interests of scholars and authors.
- Building on the Institute's rich traditions of open licensing, open publishing, and open knowledge production, MIT should seek opportunities to adopt and support the development of open source tools that embed attribution, citation and other norms of scholarly behavior and discourse in AI output, in order to address the concerns of the creators of scholarly content while simultaneously promoting a responsible framework for those researchers who use scholarly content to train AI systems.
- Given the evolving nature of this topic, the sub-group tasked by the Provost should continue to informally monitor and track concerns and needs of researchers for the next

two years and then report to the Committee on Intellectual Property (COIP) (and/or other senior leadership groups) to address any arising MIT community needs. In addition, the sub-group should:

- Establish a centralized repository for information, resources and FAQs related to this topic generally. The first iteration of this website can be found at: [Machine Learning with Scholarly Content](#)
- Seek ways to increase awareness of AI related workshops, events, and open discussions across campus; as a way to encourage dialogue and learning between those who seek to use content for AI purposes and those who generate content.
- Coordinate with relevant offices and senior leaders to communicate any changes to existing law to those who create and use scholarly content for AI purposes, and ensure resources and FAQs are updated.