PUBLIC SUBMISSION

Received: May 29, 2025
Tracking No. mba-4r96-l8i0
Comments Due: May 28, 2025 **Submission Type:** API

Docket: NSF-2025-OGC-0001
NITRD_FRDOC_0001

Comment On: NSF-2025-OGC-0001-0001
Request for Information: Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

Document: NSF-2025-OGC-0001-DRAFT-0317
Comment on FR Doc # 2025-07332

Submitter Information

Organization: STM

General Comment

Please find attached comment by Dr. Caroline Sutton, CEO STM (International Association of Scientific, Technical, and Medical Publishers).

Attachments

US Artificial Intelligence (AI) R-D Strategic Plan May 2025

Advancing trusted research

May 29, 2025

Faisal D'Souza
National Coordination Office (NCO)
Networking and Information Technology Research and Development (NITRD)
National Science Foundation
2415 Eisenhower Avenue
Alexandria, VA 22314

RE: STM Response to Request for Information on the Development of a 2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan Docket ID No. NSF-2025-OGC-0001

Dear Mr. D'Souza:

STM, the International Association of Scientific, Technical and Medical Publishers welcomes the opportunity to provide input on supporting US innovation and America's AI leadership, as NITRD and OSTP develop an AI R&D Strategic Plan. We appreciate the Trump Administration's commitment to promote human flourishing, economic competitiveness, and national security through leadership in high-quality and accurate AI. These comments build on comments STM submitted earlier this year on the AI Action Plan, but focus more squarely on the research sector, where we and our members have significant experience as providers of high-quality content for the training of AI, users of AI in internal processes, and developers of AI tools for the research enterprise.

STM represents non-profit scientific societies, university presses, and commercial publishers. Our members support hundreds of thousands of American jobs and drive innovations by American researchers that are exported throughout the world, part of a US copyright industry that adds $1.3 trillion annually to the US economy. 60 of our members are based in the US, and many more have offices in America supporting hundreds of thousands of American jobs. Our members play a vital role in disseminating high-quality, peer-reviewed reports about research that drive scientific discovery and innovation. We are deeply invested in ensuring that AI technologies are developed and deployed in a manner that promotes discovery and innovation while ensuring the integrity and reliability of the information provided by the technology, protecting intellectual property, and fostering public trust – all of which provide the foundation for future innovation and job growth.

STM and its member publishers understand that AI has the potential to radically change the way we work, learn, do research, and deal with information. STM members are already deploying AI to support accurate, transparent, and reliable research and discovery processes. From our perspective, in addition to specific research and needs and development challenges in AI, the National AI R&D Strategic Plan needs to consider the fundamental enablers and infrastructure needed to drive R&D and support the priorities: to wit, the high-quality, vetted and validated scholarly articles provided by our members, the systems and structures that enable access, integration, and use of these works, and the intellectual

property protections that incentivize and enable discovery and innovation. We expand on each of these three recommendations for the strategic plan below.

(1) The importance of high-quality, vetted and licensed content for advancing research

A pillar of AI development is the use of high-quality input for the training and validation of AI models: high-quality input is needed to train high-quality models for them to generate high-quality, trustworthy outputs to support innovation and growth. In the academic publishing field, such input can have different attributes:

i. Research data (either raw or structured into datasets or databases): which underpins research findings. STM publishers have supported data sharing best practices and created data policies to facilitate and encourage the making available of such data and linking them to the articles reporting on them.
ii. Academic articles: which explain and contextualize the experiments, data collected and findings – a unique expression of the researcher's ideas and discoveries, which is protected by copyright. There are different versions of an article: from the author's original manuscript, to the version that is submitted to a journal for publication, to the accepted manuscript after peer-review, to the Version of Record (VoR), which is the final published version.

Publishers are strategically positioned to provide both types of content, but especially high-quality, research articles, validated by peer review, that are so critical to AI training. This highly valued input is the result of publisher investments in editorial systems and processes, including research integrity checks and peer review management, to ensure that the final version of the published article, the VoR, is reliable and valuable, and updated as necessary. When used as training input, the quality of this content improves the quality and accuracy of AI tools. Publishers have also invested significant resources to develop infrastructure for services and products that build on tagging of articles for discovery, their enrichment with machine-accessible metadata and structuring content in standardized formats, all essential for machine-use.

These investments – and the value provided in the outputs of these investments – are enabled by frameworks (including those provided by the current copyright and IP frameworks) which incentivize and reward the production of high-quality, reliable reports about research. They also drive competitiveness and enable authors and users of these information-rich products to establish provenance and provide attribution.

STM encourages the AI R&D strategy to include planks that promote the quality of AI and its inputs by:

- Promoting **standards that encourage or certify** when AI tools are using **the best quality input**. For scholarly information, this would be the Version of Record which contains the latest peer-reviewed information and is preserved with integrity to include any corrections, to **promote accuracy and reduce the risk of false, misleading, or biased outputs**.
- **Requiring audits** and/or certification of the outputs of AI for accuracy and lack of bias, in coordination with the private sector and emerging tools. This could include implementation of corrective mechanisms for inaccurate or outdated information. In the scholarly space, this would

ensure that corrections and/or retractions are included in AI, so as **not to waste investments or taxpayer dollars on outdated research** reports.

- Upholding the principle of **rigorous human oversight**, which is crucial in ensuring the highest standards of quality and integrity and safeguarding against misinformation. Publishers continue to prioritize the role and expertise of human reviewers and ultimately take editorial and legal responsibility for the content they publish. Humans should always be able to review the chain of evidence across the literature.

- Working with other stakeholders to **identify and safeguard the trustworthiness of AI outputs**. For example, the STM Integrity Hub works to address challenges to the integrity of scholarly communication from human actors as well as AI tools. Through a combination of shared data and experiences, and by harnessing technological innovation, the STM Integrity Hub provides an environment for publishers to collaborate with other parties to develop and operate screening tools, including those using sophisticated AI for the benefit of the entire scholarly ecosystem.

(2) Supporting the systems and structures that enable access, integration, and use of these works, including licensing, transparency, and attribution

Publishers are excited about the potential that AI, and especially Generative AI, brings for scientific discovery and scholarship. Recognizing the importance of an objective assessment on risks and opportunities in this space, STM has supported independent analysis of the current infrastructure and gaps supporting high-quality content and its use by AI, reported by the non-profit Ithaka S+R in two reports: The Second Digital Transformation of Scholarly Publishing and A Third Transformation? Generative AI and Scholarly Publishing. Ithaka S+R also maintains a tracker of generative AI products used in research and education.

STM and its membership have been considering the potential of AI in the sector for a long time and created Best Practice Principles for Ethical, Trustworthy and Human-centric AI in 2021. After the emergence of mainstream generative AI tools and the expectation of the exponential increase in the use of these tools in research and research publishing, we released initial guidance for the use of: Generative AI in the publication process. STM also reflected on ethical principles for the use of AI in our AI White Paper.

For AI to deliver on its promise, guardrails and firm principles will be needed to ensure the responsible development of AI and its appropriate place in the ethos of science:

- AI applications operating on scholarly content should use the **final published article** (the Version of Record) because that is the most thoroughly vetted form of a research publication and gets updated over time. Corrections, errata and retractions are an important part of the updating process and serve an essential role in keeping the scholarly record clean. They can take place when serious issues are unearthed in a published article, which then gets updated with a clear revised status. In cases of a retraction, the record for the work remains publicly available as a placeholder but contains information as to the retraction along with a link to the retraction notice. Versions that are not the Version of Record often do not get this update, remain available without any cautionary

notice, and could still be ingested by and erroneously influence AI models even after the findings they report have been discredited. **These are some of the reasons why use of the Version of Record is an essential condition to ensure the scholarly community, and the public, can trust AI outputs as accurate and reliable**. STM participates in cross-sector initiatives aimed at improving the management and visibility of retractions and errata, including the CREC (Communication of Retractions, Removals, and Expressions of Concern) Working Group at NISO and the CUSAP project (Content-update Signaling and Alerting Protocol). Providers of AI models, both general-purpose and especially if intended for scientific use, should have measures in place to deal with corrections, retractions and errata to ensure the self-correcting nature of research and scholarship is included in their operation.

- AI should be **transparent**: the scholarly discourse, and the innovation and economic and public health benefits that derived from it, rest on **clear provenance** and **chain of evidence**. Where AI models and outputs are added as a component in the process of generating knowledge from knowledge, building on existing scholarly literature, AI should be placed into such a chain. It should be clear what resources AI has been trained on and learned from, and link to references (**attribution**) or other types of information should be publicly provided to ensure provenance and the ability to track assertions generated by AI back to the original source literature.

Copyright and licensing are key enablers in the practical implementation of these principles, offer legal clarity and flexibility for the parties involved and guarantee the high-quality of the input.

(3) Protecting the intellectual property regimes that incentivize the creation of these works and enable their dissemination and use

A key driver of the development of high-quality AI tools has been the use of high-quality articles reporting on research advances to train these tools. As noted above, the development and dissemination of this high-quality material relies on the current copyright and IP frameworks which incentivize and reward continued investments to structure, annotate and enrich content as required for machine use, and to turn raw information generated by experiments into high-quality, reliable knowledge. The strategy should not prioritize free circulation of overwhelming amounts of unverified, unreliable information. Instead, it should promote investment in being able to build curated, contextualized, and reproducible knowledge. A solid copyright framework that underpins a strong research sector in the US, drives competitiveness and enables researchers and publishers to establish provenance and attribution; two key enablers of a healthy research ecosystem that produces reports on scientific breakthroughs which in turn can be used by AI systems to maximize their benefit. It therefore must be made central.

Contrary to claims by some AI developers, a market for the licensing and provision of this data already exists, and some deals have been announced publicly. Ithaka S+R, an independent non-profit research and consulting organization, maintains a tracker of some of the licensing deals made.

The content licensing market for AI has developed because the copyright and contract law framework it needs already exists. The market environment these create allows willing parties to agree on mutually

acceptable terms for content licensing deals directly or via intermediaries. In order for the AI licensing market to realize its full potential, greater transparency is needed. More needs to be done to ensure that publishers and authors alike can verify that license terms on content are respected and that AI tools are providing trustworthy output based on the training materials used.

In summary, we recommend that policy in this area do the following:

- **Respect and reinforce existing intellectual property protections**, including but not limited to copyright law, to ensure the ongoing development and availability of high-quality, vetted information and content. We also refer you to an STM submission to the US Patent and Trademark Office in 2020 on the need for intellectual property protection in the context of AI innovation.
- **Leverage and build upon an already existing and well-functioning licensing market**. The market for licensing content to AI developers has developed, and continues to develop, because the copyright and contract law framework for it already exists, allowing willing parties to agree on mutually acceptable terms. In addition to its direct benefits, licensing also fosters accountability, transparency, and provenance regarding key information such as training materials used by AI systems.

Research is already being transformed by the use of artificial intelligence, and continued investments in research will in turn transform and super-charge further development of artificial intelligence itself. With strategic planning, together with appropriate oversight, standards, and regulation, the US has the opportunity to expand its leadership in AI for the benefit of society. Because advancing trusted research is at the center of what STM and its members do, STM's members are perfectly positioned to help enable trusted and trustworthy AI. The work that our members do to support digital innovation, from the community's early embrace and stewardship of digital forms of content, tagging and enriching it, creating ontologies, and even using AI to improve it, can be key enablers of a quality AI future with the continued support of intellectual property protection and licensing regimes. With our long experience with digital technologies, STM and its members stand ready to engage with the federal government in support of US leadership in AI and an associated trust and integrity in these technologies.

Respectfully submitted,

Dr. Caroline Sutton
CEO

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the 2025 National AI R&D Strategic Plan and associated documents without attribution.