

# Generative AI and Research Integrity, CNI Fall 2023

- Chris Bourg, Director of Libraries
- Heather Sardis, Associate Director for Technology and Strategic Planning
- Erin Stalberg, Associate Director for Collections and Faculty Relations Strategy

**MIT  
Libraries**



# MIT President Sally Kornbluth, Presidential Inauguration, 2023:

*We must help society come to grips with  
the tectonic forces of artificial intelligence,  
containing its risks and harnessing its  
power for good.*



## MIT CFP, 2023

- Inform public discourse and the development and application of generative AI
- Develop impact papers that articulate effective roadmaps, policy recommendations, and calls for action





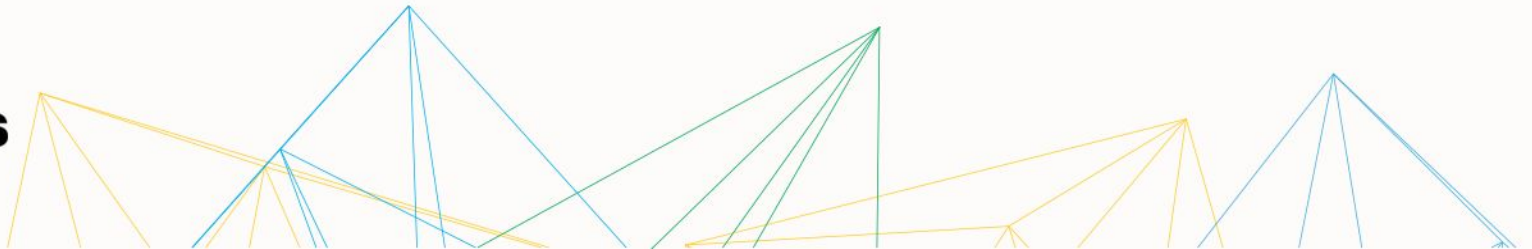
## Why research integrity?

- Public trust in academia and in research waning
- Recent high-profile instances of research misconduct
- Must leverage the power of generative AI in ways that enhance, rather than erode, public trust in science, academia and research.



# Research integrity is critical for:

- Ensuring reliable information for decision-making
- Advancing technology and innovation
- International collaboration and alliances
- Preventing exploitation and manipulation
- Nurturing public trust and support
- Counteracting disinformation







## Why MIT Libraries?

- *"Libraries and archives are institutions that help society cling to the truth."* – Richard Ovenden, Bodley Librarian Cambridge @ MIT Libraries, October 2023
- [Center for Research on Equitable and Open Scholarship](#)



# Why MIT Libraries?

- Deeply invested in supporting and strengthening the research enterprise
- Expertise already established in doing research on how scholars create and disseminate knowledge
- Leaders in open and equitable science
- MIT Libraries Gen AI Team: Micah Altman, Chris Bourg, Sue Kriegsman, Nick Lindsay, Heather Sardis, Erin Stalberg

# Goals of today's session

- We'll share our progress in:
  - Reviewing emerging and potential roles of Gen AI in science policy and as part of the scientific information infrastructure
  - Characterizing the specific applications of Gen AI that have the greatest potential for promoting impact, openness, equity, and trust in science
  - Identifying the research needed to ensure Gen AI alignment with science practice and its underlying core values





# Applying GenAI to research integrity



# Elements of research integrity

- Prohibitions against misconduct
- Conducting research with honesty, transparency, and objectivity
- Commitment to ethical principles (respect, accountability)
- Research replicability and reproducibility

# Values that underpin the research enterprise

- The *inputs* to scholarly communication
- The *content* of scientific communication
- *Participation* in science
- The *systems and processes* of scholarly communication



# Reducing peer review bias and burden in scholarly publishing

- Reduce the impact of existing reviewer biases
- Substantially reduce reviewer burden
- Make the review process more consistent and reliable across reviewers and outlets
- Accelerate the publication process
- Provide a framework for open documentation, measurement, and evaluation of the peer review system

# Enhancing the availability & accessibility of open data

- Enhancing existing data sets with automatically generated documentation and metadata
- Automating the process of checking submissions against journal data replication policies to ensure that publications are compliant with standards and transparency requirements
- Improving the interfaces to data-discovery systems and the relevance of the results that they produce



# Increasing inclusion in scholarly communication: Broaden accessibility

- Translating English language publications
- Augmenting publications with structured annotations to communicate article organization
- Describing specialized content such as figures, tables, and equations for print-disabled readers
- Generating plain-language summaries of scientific findings for non-technical audiences

# Increasing inclusion in scholarly communication: Broaden participation in publishing

- Accurate and timely translation of manuscripts into English for review
- Adapting AI authoring tools to the needs of English language learning authors
- Adapting AI authoring tools for scientific writing
- Developing AI tools to facilitate the peer-review process for English language learning writers

# Calling back to the values that underpin the research enterprise

- The *inputs* to scholarly communication
- The *content* of scientific communication
- *Participation* in science
- The *systems and processes* of scholarly communication



# Open research opportunities: scholarly communication and Gen AI outputs

- Designing foundation models so that they are reliably verifiable and transparent as to their level of uncertainties.
- Developing standards and test methods, corpuses, and auxiliary tools that researchers and the public could use to evaluate the quality of algorithmic outputs in their various contexts and use cases.
- Developing new paradigms for peer review.

# Open research opportunities: scholarly communication and Gen AI inputs

- Efficient approaches to privacy-preserving training of GenAI models
- Efficient approaches to addressing personal information in the training of Gen AI
- Mechanisms to limit memorization, track provenance, support attribution, and align machine learning outputs with the specific requirements of copyright and licenses



# Open research questions: scholarly communication and AI governance

## *Dynamics of the knowledge market and ecosystem*

- How Gen AI is, could, and ought to affect the health and operation of the scholarly knowledge ecosystem
- How does it affect the durability and sustainability of the ecosystem?
- How does it affect norms and incentives for participating in science?
- How does it affect who participates in science and how the burdens of participation are distributed?

# Open research questions: scholarly communication and AI governance

*Enclosure, durability, and sustainability of the knowledge commons.*

- Designing institutions and approaches that yield a healthy knowledge commons.
- As AI tools become increasingly integrated into the dissemination and interpretation of the scholarly record, new methods and institutions of digital preservation will need to be developed.

# Open research questions: scholarly communication and AI governance

## *Norms and incentives*

- Designing norms and practices for excellence in hybrid human-AI scholarship

## *Participation and burden in science and scholarly communication*

- Research is needed into how interventions using GenAI affect participation directly and indirectly.

# Our envisioned impact

- An actionable strategy for addressing the most promising opportunities to increase trustworthiness and integrity in the products and process of research
- Ability to mine research data and scholarly output at massive scale to surface new insights and formulate hypotheses.
- Ability to translate the academic language of research articles into a wide and accessible array of languages, formats, and reading levels.

# Our envisioned impact

- Improvements in the quality and availability of research data, both to increase research integrity and to increase the pool of viable training data for AI.
- Accelerating the open availability of scholarly literature and data to increase the quality of the data being used in Gen AI tools, and thus increase the quality of the output of the tools.





## Next steps

- Complete the paper
- Host invitational workshop, summer 2024

-----

- For more information:  
<https://libraries.mit.edu/creos/>

*(images from [Midjourney AI Bot](#))*

