

Assessing the Application of AI and Machine Learning in Cataloging and Metadata: Progress Report of the PCC Task Group on AI and Machine Learning

Charlene Chou^{1,*}, Inkyung Choi²

¹*New York University, Division of Libraries, New York, NY, United States*

²*DCMI Education Committee, United States*

Abstract

The Program for Cooperative Cataloging (PCC) Task Group on AI and Machine Learning for Cataloging and Metadata was charged to develop guiding principles, best practices, community of practices, training documentation and plans and international partnership. This report outlines the group's deliverables, methodologies, accomplishments, ongoing initiatives, and the challenges encountered during the initial phase of the project.

Keywords

metadata management, NLP (Natural Language Processing), generative AI, semantic enrichment, controlled vocabularies, LLM (large language model), information retrieval

1. Introduction

In response to the rapid advancements in AI and machine learning (ML) technologies and their potential impact on library cataloging and metadata, the PCC Policy Committee established a dedicated task group to conduct a comprehensive environmental scan. This effort is grounded in thoughtful consideration and strategic planning for the cataloging and metadata communities. The group's mandate includes developing guiding principles, fostering a community of practice, engaging with international counterparts, and creating best practices for AI/ML applications.

2. Deliverables

The PCC Task Group on AI and Machine Learning for Cataloging and Metadata is charged with eight deliverables, including the development of guiding principles, best practices, and training resources. Their work also involves fostering a community of practice, engaging with international partners, enhancing the PCC knowledge base, and assessing the broader impact of AI on the future of cataloging. [1]

DCMI-2025 International Conference on Dublin Core and Metadata Applications

*Corresponding author.

 Charlene.chou@nyu.edu (C. Chou); cik860@gmail.com (I. Choi)

 0000-0003-4736-7662 (C. Chou); 0000-0001-5048-8516 (I. Choi)

 © 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

3. Methodology

To manage the eight deliverables, diverse areas of expertise, and tight timelines, the Task Group adopted a decentralized approach by engaging members into deliverable-focused subgroups. Progress was tracked using collaborative tools, and regular meetings facilitated coordination and knowledge exchange.

4. Progress and achievements

- Publicized the “Draft PCC Guiding Principles for the Use of AI and Machine Learning in Cataloging and Metadata”[2] in March 2025, with its feedback form open until August 8, 2025.
- Developed a living document outlining AI/ML definitions, including relationship diagrams and examples like Large Language Models (LLMs) and their applications in cataloging.
- Submitted recommendations for establishing a PCC AI Metadata Community of Practice, approved with pilot meeting set for August 5.
- Engaged with relevant initiatives, including discussions on the RSC AI White Paper[3] and presentations from OCLC on AI/ML activities.

5. Ongoing initiatives

- The Best Practices Subgroup is addressing data provenance for AI-generated metadata, considering MARC field 588 and subfield \$7 for recording such information, which is still under discussion. The subgroup is drafting a comprehensive best practices document and the draft will be shared with the Standing Committee on Standards for comment.
- The International Outreach Subgroup is identifying global AI cataloging experiments and reaching out to inform best practices.
- The AI Training Subgroup, in collaboration with the Standing Committee on Training, is curating existing learning resources and developing training outlines.
- Efforts are underway to update AI resources on the PCC wiki, ensuring they reflect current developments and international activities.

6. Conclusion

Key challenges of our work include managing membership changes while ensuring continuity of expertise; keeping pace with the rapid evolution of AI by developing effective strategies to maintain up-to-date resources, communities of practice, and best practices documentation; and ensuring that the PCC Guiding Principles are effectively shared with and understood by the cataloging community, emphasizing that AI is a tool designed to support and enhance cataloging workflows, not to replace human catalogers. In conclusion, the Task Group’s work reflects a thorough review of the strategic applications of AI and ML technologies in cataloging. By developing guiding principles, fostering communities of practice, and engaging in international

collaboration, the group aims to ensure that AI serves as a tool to enhance, rather than replace, the human expertise of cataloging and metadata professionals.

References

- [1] PCC Task Group on AI and Machine Learning for Cataloging and Metadata, <https://wiki.lyra-sis.org/display/PFCCP/PCC+Task+Group+on+AI+and+Machine+Learning+for+Cataloging+and+Metadata>
- [2] Draft PCC Guiding Principles for Use of AI and Machine Learning Technologies in Cataloging and Metadata Work, https://docs.google.com/document/d/1UIDqZGU-lgVfRmWb_-UKRhZRUySyr1SdrU6qSA2f-80/edit?tab=t.0
- [3] RSC/Papers/2025/1 Considering AI in the Growth of RDA, https://www.rda-toolkit.org/sites/default/files/uploads/RSC_Papers_2025_1.pdf