Metadata Remediation through AI Collaboration

ELIZABETH ROKE

Abstract:
This presentation examines the research findings of an 8-month project at Emory University to explore the advantages and challenges of generative AI technologies for descriptive metadata of historical images. Our study combined LLM experiments with user studies to assess the potential of generative AI for the remediation of legacy metadata and the creation of accurate, inclusive, and contextually-sensitive metadata for newly-acquired collections.

Our research engaged with three central questions: (1) Can LLMs effectively identify bias in existing metadata while providing appropriate alternatives to help remediate legacy metadata? (2) How useful is AI-generated metadata to end users? (3) What might a human-AI collaborative metadata workflow look like?

Using Anthropic's Claude models and a specially designed prompt engineering framework, we tested LLMs' capabilities for generating, evaluating, and refining metadata across selected image collection samples. Our user studies, which engaged Emory students and library staff through surveys and focus groups, revealed preferences for machine-generated and hybrid-generated metadata over original human-generated content, with participants valuing context, specificity, and comprehensiveness. This work culminated in a proof-of-concept developed in collaboration with AWS's GenAI Innovation Center.

Our findings suggest that while LLMs can enhance metadata workflows by providing additional contextual information and flagging metadata quality issues such as bias, they require thoughtful human oversight to prevent overinterpretation and ensure accuracy.

About the author:
Elizabeth Russey Roke is the Archival Technology Program Manager at Emory University Libraries. Primarily focused on description, discovery, and access to special collections materials, she works on a variety of technology projects and initiatives related to metadata standards, archival descriptive practice, discovery and next generation technology such as linked data and AI.