Responsible AI: Tools for values-driven AI in libraries and archives

Summary: Montana State University (MSU), James Madison University, and Iowa State University seek $249,999 for a 3-year Implementation Grant to develop resources that support ethical use of artificial intelligence (AI) in libraries and archives. In alignment with IMLS Agency Objective 3.2., *Responsible AI* will promote ethical, values-aligned practices and strategies for minimizing harm as libraries and archives "promote access to museum and library collections" through the use of AI.

Project Justification: Over the past few years, libraries and archives have begun using AI to enhance library services, especially to support collection description and discovery. Some examples include using natural language processing, computer vision, web scraping, and geocoding to improve metadata for digitized photos [1], applying image processing and machine learning to enhance manuscript metadata [2], and others (e.g. [3][4][5][6]). Recent publications [7] [8] [9] and meetings [10] [11] [12] [13] explore the theory and practice of AI in libraries and archives. The ACRL Research Planning & Review Committee designated machine learning and AI a "top trend" [14]. Organizations like AI4LAM [15] have formed to cultivate community capacity, and Europeana is assessing AI usage in cultural heritage institutions in Europe [16]. AI in libraries and archives is part of a broader, multisector trend embracing the potential to use AI to enhance services. However, libraries and archives have tempered excitement about AI with an awareness of its potential for amplifying harms to the communities they serve. In libraries, our ethical care is our strength—as Kate Zwaard writes, "through the slow and careful adoption of tech, the library can be a leader" [25, p. 2]. Libraries are aware that because training data are biased, AI is biased as well [17] [18], and can cause at-scale discrimination based on race, gender, sexuality, class, nationality, and other factors [19] [20] [21]. The Data Justice Lab keeps a running record of the "harms that have been caused by uses of algorithmic systems" [22]—an eye-opening document that shows the pervasive challenges of AI.

This grant proposal addresses the following problem: How can we use AI in libraries and archives while minimizing harm and upholding professional values such as diversity, equity, social responsibility, and supporting the public good [23][24]?

Responsible AI proposes solutions to this problem. This project will produce a toolkit that helps practitioners consider ethical implications of AI projects in libraries and archives, thereby aligning with our professional values and avoiding harm to our communities. The project begins with the premise that AI is already in use in libraries, and that its use will continue to grow. Our project builds from that starting point. We will introduce an ethical layer to AI in libraries that will help practitioners responsibly implement AI in accordance with our professional values, thus "improv[ing] the ability of libraries and archives to provide broad access to and use of information and collections" (NLG Goal 3). By bringing together a wide range of project personnel (see below), we place an "emphasis on collaboration to avoid duplication and maximize reach" (NLG Goal 3), amplifying our shared strengths and creating pathways to help our wider practitioner community consider the potential harms of AI. The deliverables of *Responsible AI* are practical tools that support ethical use of AI, in alignment with professional values, for the betterment of society [26] [27].

Project Work Plan: **Phase 1. Environmental scan, Aug 2022-Jan 2023.** We will conduct a review of AI projects mentioned on library websites (expanding on [28]) and widely distribute a survey (taking into account [29]) to gain understanding of the landscape of AI usage in US libraries and archives of various sizes and missions. The scan will be complementary to efforts in Europe [16][30]. The goal of the scan is to gather information about current applications of AI tools and processes, so as to learn what risks and potential harms occur when using AI in libraries and archives. Results will be published as a journal article in 2023/4, and shared as conference presentations in 2023. **Phase 2. Case studies and practices exchange, Feb-Aug 2023.** We will solicit case studies to seed an online *practices exchange* that highlights successes, challenges, and missteps of AI projects in libraries and archives—through the lens of ethical, values-driven implementation. (See [8, p. 10] [31, p. 469] [32] for more information on the idea of *practices exchanges*). The exchange will model "transparency as a means to work against repeated community mistakes" [8, p. 10], and will be organized according to different ethical concerns and potential harms. We will use Open Science Framework or a similar system for discovery, access, shared use, and preservation. The case studies and practice exchange will be shared via conference presentations and a webinar. **Phase 3. Build ethically-relevant harms analysis tool, Sep 2023-Mar 2024.** The tool will weigh the harms of AI against potential community benefits**.** The tool will include methods for auditing AI approaches, and will be based on models that support responsible practice in other contexts—i.e. Data Ethics Decision Aid [33], Social Inclusion Audit [34], and Envisioning Cards [35]. The Envisioning Cards are a close cousin to the tool we hope to produce; they are colorful and easy to use, and they help designers of new technologies consider the long-term and indirect effects of their tech products. Our proposed tool will support librarians and archivists engaging with AI, exploring

potential benefits and harms by posing such questions as: How will the project affect libraries, archives, and society in the short and long term? What consequences could result if this particular AI implementation became common practice? How might this AI project affect the lives of different library stakeholders (e.g. workers, users, administrators)? How does this AI project align or misalign with our values as librarians and archivists? To prototype an initial version of the tool, the grant team will host a workshop in Bozeman, MT, in Fall 2023. Workshop participants will include practitioners with diverse identities and a variety of experiences. The workshops will use survey results from Phase 1 and case studies from Phase 2 to support development of the tool. **Phase 4. Tool assessment and validation, Apr 2024-Jan 2025.** The tool will be distributed to community experts, who will use it during the planning process for AI projects; they will then provide structured feedback that will be used to revise and improve the tool, taking into account continuing results from Europeana's AI in Relation to GLAMS Task Force. Outreach in 2024-2025 will focus on encouraging use of the practices exchange and tool. We plan to host workshops in venues such as Learn@DLF and OCLC Works in Progress webinar series. **Phase 5. Final tool and handbook, Feb 2025-Aug 2025.** The handbook will introduce fundamental issues, summarize the state of the field and the results of the environmental scan, and provide documentation to facilitate the use of the grant deliverables. Final deliverables will be presented at conferences in 2025. The team will also host workshops about putting the tool into practice. **Sustainability and dissemination**. The DLF Privacy, Ethics, & Technology (PET) Working Group is home to an active community of scholars and practitioners. Co-PDs Shorish and Young are past co-conveners and current active members of the group, and PD Mannheimer will also join the group to steer the sustainability of grant deliverables. Mannheimer will leverage the DLF PET community to support use and dissemination for the tool, and will invite the community to participate in an annual review and revision of the tool for at least 5 years after the grant ends.

Diversity Plan: AI can generate harms that disproportionately impact people from minoritized communities. It is therefore imperative that these communities be core to our work. To this end, we will convene an Advisory Board that will help center and amplify the voices of BIPOC, LGBTQIA+, and other minoritized communities. The Advisory Board will meet twice yearly for the duration of the project to ensure that a diversity of perspectives are considered through review and active participation in the development of project outcomes. Mark Mattienzo - Stanford (software engineering, UX, human rights archives); Dorothy Berry - Harvard (African American special collections, overlooked & erased histories); Thomas Padilla - Center for Research Libraries (collections as data, responsible AI); Bohyun Kim - U Rhode Island (Chief Technology Officer, AI researcher); and Stephanie Russo Carroll - U Arizona (co-founder of Indigenous Data Sovereignty Network) have agreed to serve on the Board. Board members will be compensated for their time and expertise.

Project Results: This project contributes to development of consistent guidelines and practices for the ethical use of AI in libraries. *Responsible AI* provides strategies for methodical consideration of potential harms of AI projects, with a goal of supporting decision-making. *Responsible AI* deliverables will help practitioners consider ethical implications as they embark on AI projects that support increased impact and new uses of library resources. We expect that this project will reach hundreds of librarians and archivists in the first year through the survey and practices exchange. The harms analysis tool may ultimately reach an even broader audience—practitioners in a range of contexts (including public libraries and museums) can use the tool in practice; teachers can use the tool in the classroom to consider ethical implications of technology; administrators and leaders can use the tool to weigh harms and benefits when deciding whether to adopt new technology in libraries, archives, universities, and the public sector. *Responsible AI* deliverables are designed in an academic library/archives context, but they can ultimately act as models beyond the profession.

Project Team: Project Director Mannheimer is a data librarian and PhD candidate in Information Science at Humboldt University in Berlin; she was previously PD for the IMLS-funded *Dataset Search* project [36]. The project team includes practitioners and researchers with experience and knowledge in the following areas: social science research methods (Mannheimer, Young, Rossmann); participatory workshop facilitation (Young, Scates Kettler); ethics and privacy for library technologies (Mannheimer, Shorish, Young, Rossmann); ethics and philosophy of technology (Sheehey); supporting diverse representation in cultural heritage collections (Scates Kettler); systems administration (Rossmann); software support and application development (Rossmann, Clark [37]); technology maintenance and preservation (Clark , Rossmann, Scates Kettler [38]); and IMLS grant administration (Mannheimer [36], Clark [39], Young [40]).

Budget Summary: The estimated budget is $249,999 to support salaries & benefits (26%); conference registration/travel (29%); on-site participatory workshops (6%) (budget will include flexible options due to COVID-19); social science research support services [41] (2%); Advisory Board & community expert honoraria (5%); tool graphic design & production costs (5%); and indirect costs (27%).