Dr Yaming Fu
School of Cultural Heritage and Information Management, Shanghai University
Prof. Simon Mahony
Department of Information Studies, University College London

Title: Beyond Boundaries: Integrating Humanities and Generative AI in Archival Resource Development

Keywords: Human-in-the-Loop (HITL), Multimodal Archival Resources, Credible Knowledge Services, Human-Machine Collaboration

This proposal presents a human-machine collaborative framework for archival knowledge services, built around the core concept of "human-in-the-loop" (HITL) to ensure credible control and trustworthy outcomes. By embedding human expertise into the key stages of data processing and knowledge generation, the proposed model emphasizes the accuracy, evidentiary value, and historical contextuality of archival content. Professional judgment and manual review by archival staff play a critical role, ensuring that the generated knowledge aligns with archival standards and retains its cultural and historical authenticity.

Taking the AIGC Scientist Knowledge Service Platform as a case study, our research demonstrates how HITL can enhance multimodal archival resource development. The model integrates advanced machine learning techniques with human intervention at pivotal decision making points (Wu, et al., 2022), including the large language model generated metadata information, full text results via OCR tool and chat function by RAG (Retrieval-Augmented Generation) (see Figure 1), balancing computational efficiency with the nuanced understanding that only human professionals can provide. This collaborative approach extends to multimodal data processing, knowledge content transformation, and the development of generative AI applications, creating a system where technology and human intelligence coexist in a dynamic interplay.

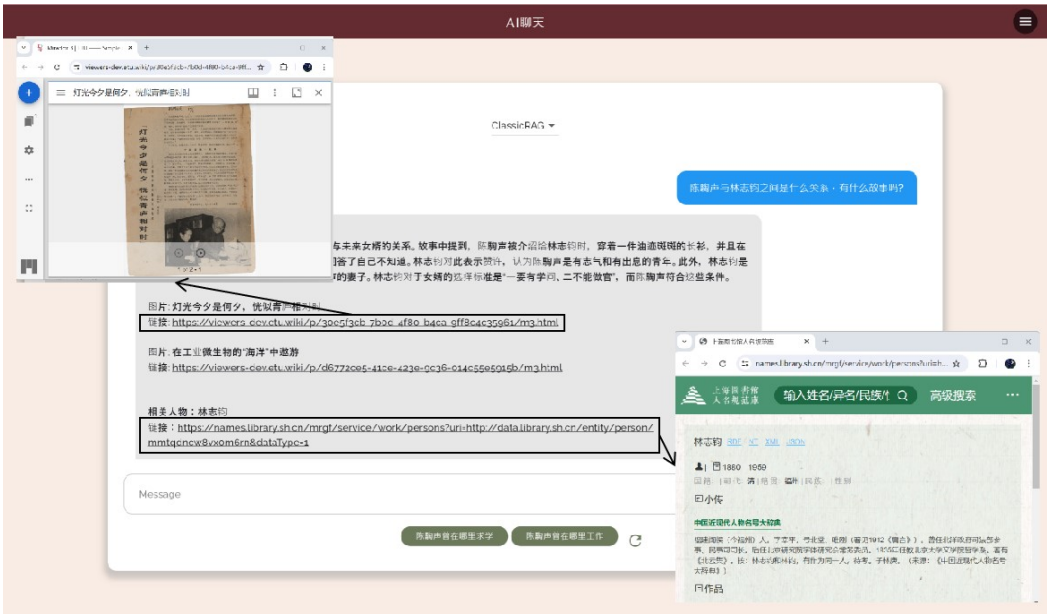

Figure 1 The Chat interface of the knowledge service based on RAG

The framework not only ensures credible knowledge services but also introduces innovative methodologies for leveraging "credible" technologies in the humanities (Thiebes, Sebastian, Ali, 2021). By focusing on interdisciplinary collaboration, it addresses challenges in aligning machine-driven processes with human-centric values, offering a sustainable and ethical approach to archival management.

This presentation will delve into the design and implementation of the HITL framework, highlighting its application to multimodal archival resources and its potential to foster cross-disciplinary dialogues in DH projects. It aligns with the conference theme, "Collaboration beyond Boundaries," by illustrating how human-machine collaboration can redefine archival practices, expand the boundaries of knowledge services, and inspire new possibilities for working with both human and non-human agents.

(Words in total: 293)

References

Thiebes, Scott, Sebastian Lins, and Ali Sunyaev. "Trustworthy artificial intelligence." Electronic Markets 31 (2021): 447-464.

Wu, Xingjiao, et al. "A survey of human-in-the-loop for machine learning." Future Generation Computer Systems 135 (2022): 364-381.