

# Enhancing Access to Legacy of Slavery Records Using Generative AI

**RAJESH KUMAR GNANASEKARAN, RICHARD MARCIANO, LORI PERINE, AND MARK CONRAD**

## **Abstract**

The presentation discusses how to harness the potential of generative AI to support exploration and discovery in archival records. The talk leverages an IMLS-funded project called GenAI-4-Arch where we partner with the Maryland State Archives (MSA), and focus on a subset of the Legacy of Slavery “Domestic Traffic Ads” (DTA) collection. The DTA collection encompasses newspaper advertisements from 1824 to 1864, which shed light on the abhorrent practices of chattel slavery—where buyers and sellers engaged in transactions involving human beings.

We use Large Language Model (LLM) interfaces to interact with records conversationally. The use of LLMs may carry significant disadvantages: (1) the potential to transmit sensitive and potentially offensive content back to the models with the risk of it being incorporated into their training data, and (2) the potential for falsehoods and distortions (aka hallucinations) to be generated.

To mitigate these shortcomings, our project actively engages with collection content experts at the MSA in Annapolis, Maryland, and a focus group of local cultural and historical experts and community members at the Kennard African American Cultural Heritage Center in Centreville, Queen Anne’s County, Maryland (located on the Eastern Shore, where many of the DTA Collection ads originate).

In this talk, we demonstrate: (1) how to make archival records AI-ready, (2) how to tune LLMs to incorporate community feedback, and (3) how to evaluate the resulting outputs based on the propensity of LLMs to hallucinate.

## **About the authors:**

**Richard Marciano** is a Professor at the University of Maryland iSchool. He is also the Director of the Advanced Information Collaboratory (AIC). His research interests focus on opportunities and challenges of disruptive technologies for archives and records management (including CAS – Computational Archival Science, AI, ML, Digital Curation, etc.), while promoting ethical information access and use.

**Mark Conrad** is an archivist with over 30 years of experience working with electronic records, including 28 years at the US National Archives. He is a member of the working group that developed and maintains ISO 14721 - OAIS Reference Model and ISO 16363 - Audit and Certification of Trustworthy Digital Repositories. He is a Co-Founder of the Advanced Information Collaboratory (AIC).

**Lori Perine** is an Assistant Professor of Math, Stats, and Data Science at Montgomery College (Maryland) and a Research Fellow at the AIC at the University of Maryland. She explores innovative uses of data science in social science and humanities applications, with an emphasis on sociotechnical dimensions of technology development and applications. A dedicated advocate

for bringing technology and innovation expertise into government to improve policy and services, she also is a special advisor to the Maryland Insurance Administration for implementing AI management and governance.

***Rajesh Kumar Gnanasekaran*** is the Assistant Director of AI Solutions at the Division of Information Technology (DIT) at the University of Maryland and a Research Fellow at the AIC. He helped launch the HERITAGE-AI Initiative: “Harnessing Enhanced Research and Instructional Technologies for Archival Generative Exploration using AI”. He holds a Ph.D. from the U. Maryland iSchool. His focus is to explore, analyze, and apply computational treatments on cultural datasets using advanced data science-based approaches such as LLMs, ML, AI, natural language processing, and graph databases.