

Workshop Proposal, “Archiving America/American Archives” Siegen University, June 12-14, 2025

Can Archives Be More Than Datasets?

The rapid implementation of Generative AI since the early 2020s has affected several assumptions at the core of academic practices. Standards and expectations related to evaluation, production, labor, innovation, and iteration were fundamentally unsettled by generative tools now available to the broader public. Just as importantly, GenAI has trawled petabytes of academic writing archived on both private and public servers to develop the affordances it is most commonly used for: categorizing, summarizing, contextualizing, and generating text, as well as making its generative production epistemically plausible. The affordances of GenAI are hence indebted to academic archives both as *organizational systems* and as *datasets*.

The “panoptic capture” (Michele Elam) of GenAI in turn alters the affordances of the academic archive itself. The essays, reviews, and monographs produced by institutions of higher education now provide GenAI with datasets that have a “prospective” orientation: GenAI trawls the archive to produce work oriented towards new uses and new contexts. Additionally, this new prospective orientation is marked by the properties that have long marked academic work: the archive (or dataset) is overwhelmingly Anglophone, white, and male.

The moment of AI also reframes the traditional physical archive as potentially auratic and resistant to the deprivations of AI trawling. Physical archives of writing, sound, and visual and audiovisual media remain a field inaccessible to AI, asking us to reconsider their function and importance.

This workshop addresses the consequences of AI as an important shift in the infrastructural makeup of academic practices in and with archives. We welcome papers that engage with the intersection of AI and the archive in a number of ways, including but by no means limited to questions such as:

- What is the “archive” of a LLM or an algorithm?
- What is the relationship between AI’s datasets and other archives, both empirically and theoretically?
- How might we deploy archives to resist algorithmic bias?
- What is the relationship between AI’s global scraping of data and America specifically?
- What is the utility of AI tools and applications for making (better?) use of physical archives?
- How are physical archives reframed as (less accessible) stores of knowledge in an age of AI?
- What does it mean to think of all previous textual production as “datasets” of an “archive” of AI?
- How does the AI archive—and GenAI textual production—relate to theoretical ideas about textuality and literariness?
- What is the role of American Studies in navigating this new situation?

Confirmed Speakers:

1) Sarah Ciston, University of Southern California/AI & Society Lab, Alexander von Humboldt Institute for Internet and Society, Berlin

“ELIZA’s Archives”

2) Rebecca Roach, University of Birmingham

“Encoding/Decoding, Stuart Hall and the Black Archival Imaginary”

Organizers:

PD. Dr. Tim Lanzendörfer, Goethe-Universität Frankfurt

Prof. Dr. Pierre-Héli Monot, Ludwig-Maximilians-Universität München