

Bias Is Not a Bug: Towards Methodological Awareness in AI-Assisted Humanities

Sven Burkhardt

sven.burkhardt@unibas.ch

Digital Humanities Lab, University of Basel, Switzerland

0009-0001-4954-4426

Abstract

This talk examines the intersecting layers of bias that emerge when applying large language models (LLMs) to historical sources—specifically, a corpus of letters from Nazi-era Germany (1939–1945). Drawing on a Digital Humanities project that uses neural networks and LLMs to extract named entities (persons, roles, places, organizations), I argue that machine bias does not act in isolation, but rather interacts with and amplifies existing archival and source-based distortions.

The first layer is archival bias: not all individuals, voices, or materials are preserved equally. The second is source bias: even within surviving documents, representation is uneven—particularly for women and marginalized groups. For example, female actors often appear only in relational forms, making them nearly untraceable. The third layer is model bias: LLMs trained on modern or biased datasets tend to misclassify, erase, or flatten these already fragile traces.

Rather than viewing these biases as mere technical errors, I propose treating them as structural epistemic problems. The talk discusses how we address these issues within the annotation pipeline by implementing review flags (needs_review), context-sensitive deduplication, and uncertainty tracking—turning machine limitations into analytical tools.

This presentation aims to offer perspectives on how we engage with AI in the humanities. Existing biases in our research data can be amplified by the tools we use—especially by powerful systems such as LLMs. What we call “AI” becomes, through our application of it, a co-producer of historical interpretation and meaning. Recognizing and documenting bias at every level—within archives, within source texts, and within computational models—is not a limitation, but a methodological necessity.

This case study offers a practical and critical reflection on the ethical responsibilities of DH projects that rely on AI, and calls for bias-aware annotation frameworks that foreground what remains invisible, uncertain, or contested in historical data.

Biography

Sven Burkhardt is a master’s student in Digital Humanities and History at the University of Basel, where he also completed his bachelor’s degree in Sociology and History. His academic work explores the intersection of historical research and computational methods, with a particular interest in social networks and historical data modeling.

He is currently writing his master’s thesis, which investigates personal and institutional networks in Germany between 1939–1945. The project focuses on extracting messy data from a corpus of historical letters and analyzing it using a mixed-methods approach that combines neural network-based OCR, language processing with LLMs, and algorithmic data extraction for network analysis.

Alongside his studies, Sven works as a research assistant at the Digital Humanities Lab Basel (DHLab) and completed a research internship with the University of Basel's RISE program (Research Infrastructure and Support in the Humanities). There, he contributed to projects dealing with the conceptual design of computer-assisted research, the processing and analysis of historical data, and the development of sustainable strategies for the open and reusable publication of digital research data.

Beyond academia, he is actively involved in university governance and student representation. He currently serves as President of the Student Council (Studierendenrat) and is boardmember for the Studentassociation Digital Humanities (Fachgruppe Digital Humanities)