# AI and the Archive: Change and Cultural Impact in an Evolving Digital Landscape

Jacob Muller

Luddy School of Informatics, Computing, and Engineering

LIS 581 – Archives and Records Management

Professor Jenifer Monger

December 17, 2024

**AI and the Archive: Change and Cultural Impact in an Evolving Digital Landscape**

The significant technological strides made over the course of the past few decades have led to drastic shifts in the ways which users interact with and search for information, even more so as artificial intelligence (AI) systems and algorithms become increasingly prominent in their everyday use. Such changes necessitate the reevaluation of "Archives" in the traditional sense and the role of information professionals and archivists in this vastly changed digital sphere. If we, as information professionals, want to stay relevant in this digital era we must come to understand the changes in how users seek out information, the potential cultural impact of AI in archival spaces, and how to effectively harness AI as a tool to stay relevant while uplifting marginalized voices.

The rapid expansion of the digital landscape has itself upended traditional notions of information seeking and retention. Billions of gigabytes of information are created every day, and it is functionally impossible to retain and sift through all of it for what might be considered "valuable." Yet, as Jane Winters and Andrew Prescott (2019) point out, "Each age has felt overwhelmed by the quantity of information and has sought to develop new tools and methods to assimilate the mass of new data" (p. 393). Thus, it is our duty to rise to that challenge. Just as it is impossible for archivists to store all of this information in its present state, the user has to find a way to navigate this information overload. This need has brought the rise of Google and other major search engines whose algorithms enable the use of simple, structureless searches: Even when such methods produce poor or misleading response[s]… we nevertheless trust in the ability of the free text search to retrieve the information we want" (Winters & Prescott p. 395). Using Google or a similar

search engine enables users to "[sidestep] the gatekeepers," eliminating the knowledge barrier required to navigate digital databases in the eyes of the user (Theimer, 2019, p. 5). Thus, the question is begged: How can we ensure that archives (both physical and digital) remain relevant in the digital landscape and accessible to patrons when we ourselves struggle to navigate the constantly changing algorithms being used across it?

Numerous different approaches have been advocated for in response to this need, but all call for a significant shift in focus within the archival field. Michael Moss and Tim Gollins (2017), for instance, suggests a shift in focus from the challenges of digital preservation "toward the other core principles of archival science: namely appraisal (what to keep), sensitivity review (identifying material that cannot be disclosed for ethical or legal reasons), and access" (p. 1). Kate Theimer argues for a slightly different approach, the harnessing of algorithms and AI systems to further support the "bypassing the gatekeepers" in order to make archival materials more accessible for users (Theimer p. 5). Conversely, Clifford Lynch (2017) suggests the abandonment of traditional archival ideals altogether in favor of reimaging the field in its entirety (para. 40). While each of these approaches contains their own merit, reaching an effective solution will prove to be tedious no matter what direction the field moves toward.

While the degree to which AI will permeate archival spaces or even everyday life is itself uncertain, the potential cultural impact of AI in archival spaces is incredibly clear. In this "Age of Algorithms," as Lynch dubs it, algorithms that depend on consumer interaction and engagement to operate will only continue to grow in their reach (2017). Where AI algorithms fundamentally differ from humans is in their inability to critically think and

contextualize information: “the process by which these systems improve themselves … [is] a process of trial and error and based on probabilistic assumptions. It lacks the ideological, ethical and cultural awareness which play an important part in human decision making” (Winters & Prescott p. 399). These algorithms are still very much limited by the bias of their creators and datasets they have been trained on. Digital archives also fundamentally differ from search engines like Google in that they don’t contain the contextual information that would drive a search conducted utilizing many of these algorithms. “Web archives are also subject to change over time: they are not static archives, but transform in front of our eyes… An archived website can ‘appear’ or ‘disappear’ … and take-down notices can result in the immediate removal of material from access” (Winters & Prescott 398). It is crucial that we examine the role of critical thought and intentionality in the archival field because it will continue to distinguish the human from the algorithmic moving forward.

Many scholars have also highlighted the benefits of examining the impacts of AI in archival spaces from an interdisciplinary approach. For instance, Shakir Mohamed, Marie-Therese Png, and William Isaac (2020) advocate for utilizing various decolonial theories, arguing for the “[reappraisal] of what is considered the foundation of an intellectual discipline by emphasizing the legitimacy of marginalized knowledge” (p. 664). In reassessing the foundations of archival studies in accordance with other marginalized voices, we open the way for a reimaging of the Archive as a space in the “Age of Algorithms.” Abhishek Gupta and NIkitasha Kapoor (2020) suggest that “AI-enabled mechanisms can enhance interactivity and can become a tool for gaining insights from both dominant and lesser dominant sources of content and knowledge” (Gupta & Kapoor

p. 6) This approach sees potential in leveraging AI algorithms to uplift minority voices that have historically been underrepresented and silenced in archival spaces. Mohamed et al. (2020) further elaborate that:

> By connecting instances of algorithmic oppression across geographies, new approaches that consider alternative possibilities of using technology in socially complex settings in more critical and considered ways will emerge, and so too will designs that incorporate inclusive and well-adapted mechanisms of oversight and redress from the start" (p. 667).

Making efforts to leverage AI algorithms for the purpose of uplifting underrepresented voices offers a great deal of opportunity in connecting with users and promoting access and engagement within digital archival spaces.

Some within the field have also argued for the implementation of critical technical practices (CTP) and engaging with AI in a more critical way throughout the search process. CTP "practices take a middle ground between the technical work of developing new AI algorithms and the reflexive work of criticism that uncovers hidden assumptions and alternative ways of working" (Mohamed et al. p. 672). As previously discussed, the inability of AI to engage in critical thought is one of the fundamental differences in how humans and AI algorithms interact with information. Jo Guldi (2018) outlines a process they dub a "Critical Search," which advocates for affording "the digitally-aided scholar a set of advantageous techniques for the recovery and analysis of social experience through the mass-digitized archives so widely available today" (p. 8). Guldi's Critical Search seeks to

move past standard models for searches and engaging with algorithms to promote a transparent model for users to conduct research in the digital age. “Critical Search does not depend on a particular algorithm or set of algorithms, but rather suggests how questions of interpretation and scholarly selection permeate the entire process of applying digital tools and using their results” (Guldi p. 7). Guldi’s Critical Search has the shows great potential to alter the way users conduct research both within the archive and without.

If we are to move forward with harnessing AI algorithms as tools in digital archival spaces and reevaluating archival ideals, we must also consider the implications of such. Namely, we must examine the role of transparency and accountability. With the rapid technological developments of the past few decades and increasing complexity of the algorithms being implemented in systems all across digital spaces, the growing opacity of AI algorithm must be acknowledged. This issue has caused the rise of what has been dubbed “explainable artificial intelligence,” or XAI, which is described as the “challenge of shedding light on opaque machine learning (ML) models in contexts for which transparency is important” (Bunn, 2020, p. 144). Jenny Bunn argues that recordkeeping and XAI share the fundamental goals of transparency and accountability, but that XAI “sits in a context in which this vision is framed much more strongly as fairness and the avoidance of bias” (Bunn p. 148). Archiving and recordkeeping, conversely, are often very open about bias and its potential impacts of it on collections and materials. There is legitimate cause for concern with the narrative that AI algorithms are free from the flaws of bias and humanity, as neglecting to acknowledge that limitation creates further opportunity for misuse and malpractice within these digital spaces.

The growing presence of artificial intelligence and fundamental changes in how users interact with information in the present digital age necessitates the re-assessment of the archival field at its core. Whether we choose to merely adjust our focus from preservation or abandon our traditional notions of archival ideals in their entirety, archives must change to meet the ever-evolving needs of users in the digital age. We may come to utilize AI for finding and storing digital information, creating metadata for, checking the sensitivity of, or even promoting access to materials. For many aspects of our field, only time will tell, however: “Narrative, storytelling, meaning-making, context providing – most of the archivists I know are already great at this and we know that it is what resonates most about archives for many people” (Theimer p. 13). Regardless of the drastic change taking place across our field and the fears it may spark, we can be certain that digital archives will find ways to adapt to user needs and endure, just as archives always have.

**References**

Bunn, J. (2020). Working in contexts for which transparency is important: A recordkeeping view of explainable artificial intelligence (XAI). *Records Management Journal, 30*(2), 143-153. https://doi.org/10.1108/RMJ-08-2019-0038

Guldi, J. (2018). Critical search: A procedure for guided reading in large-scale textual corpora. Journal of Cultural Analytics. https://doi.org/10.22148/16.030

Gupta, A., & Kapoor, N. (2020). Comprehensiveness of Archives: A Modern AI-enabled Approach to Build Comprehensive Shared Cultural Heritage. *Workshop Datafication and Cultural Heritage – Provocations, Threats, and Design Opportunities*, 1–10. https://arxiv.org/pdf/2008.04541.pdf.

Lynch, C. (2017). Stewardship in the "Age of Algorithms". First Monday, 22(12). https://doi.org/10.5210/fm.v22i12.8097

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial ai: Decolonial theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, 33(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Moss M., & Gollins T. (2017). Our Digital Legacy: An Archival Perspective. *Journal of Contemporary Archival Studies*, 4(2), 1-31. https://elischolar.library.yale.edu/jcas/vol4/iss2/3.

Theimer, K. (2019). Its the End of the Archival Profession as We Know it, and I Feel Fine. In *Archival Futures* (pp. 1–18). essay, Facet Publishing.

Winters J., & Prescott A. (2019). Negotiating the born-digital: a problem of search. *Archives & Manuscripts*, *47*(3), 391–403. https://doi.org/10.1080/01576895.2019.1640753