LETTING AI LOOSE IN AN ARCHIVE

TECHNOLOGY TO MANAGE OR TO MANAGE WITH

by Isto Huvila

INTRODUCTION

Artifical intelligence has been called both a misnomer,[1] and a revolution coming very soon but not quite being here already for decades.[2] During the past couple years there has been a growing consensus that it actually might be here – for good and bad.[3] A question remains, however, what is it that we have and what does it mean. The genuinely spectacular capabilities of rapidly advancing, especially machine learning based natural language and image processing techniques have raised optimism on their capacity to carry out tasks beyond the capabilities of traditional computational methods. In archives and records management, they provide opportunities to process large quantities of records beyond what has been possible before computationally or relying on human labour. However, at the same time, the rapid development of diverse AI techniques makes it difficult to grasp what consequences, implications, benefits and downsides they have.
A large number of projects and initiatives around the world have started to work on questions pertaining to AI and archives. The author is affiliated with two such initiatives. The international InterPARES TRUST AI project[4] has worked from 2021 to inquire into how AI can be used to provide support in archives and records management tasks. An example of a smaller project is the "Artificial Intelligence as a risk and opportunity for the authenticity of archives"[5] funded as a part of the large Swedish Wallenberg AI, Autonomous Systems and Software Program – Humanity and Society (WASP-HS) research programme.
The work in these two projects and numerous initiatives beyond them, much of which is done at archival institutions, provide an increasing number of pivotal insights into how AI can assist in a broad range of critical records management tasks. Automatic character and hand-written text recognition makes transcribing texts faster in an entirely different order of magnitude compared to manual work.[6] Automatic classifiers and indexing can help to organise and document archives on the level of individual records making them directly searchable and accessible to users much faster, and with at least slightly thorough understanding of their administrative and organisational provenance.[7] Large language models have also already now proven their usefulness in summarising contents of large quantities of text. They hold even a lot of promise in supporting archival appraisal.[8] NLP can help to classify records according to a set criteria both for keeping and weeding.[9] Further, similar techniques can be used to identify sensitive information helping archivists to make decisions on whether they can be released and to whom.[10] These are merely examples of emerging opportunities.

AI AS AN IMPERFECT SOLUTION

All this is naturally more complicated in practice. Applying AI techniques requires that all processed information is digitised or born-digital. Especially with older material, digitisation still takes a lot of time meaning that a lot of material is and will remain non-digital for the foreseeable future. Considering the required resources, it is also relevant to ask how much is it reasonable to digitise. Further, even if AI is capable of producing remarkable results, they are far from perfect and it is difficult to know when they are trustworthy and on what premises. Automatic classification can be largely accurate but may well miss sensitive information or suggest that important records should be disposed. Similarly, while AI-based indexing would be mostly accurate, it can miss a lot of relevant details and perspectives and misrepresent records and their contents. Accuracy can be measured but the measurements reveal inaccuracies only in relation to what is measured, risking to lead to comparable problems information retrieval research has been struggling with for decades with how to measure and ensure the relevance of search results in information systems.
The imperfections of AI techniques in records management tasks are not, however, what makes AI truly a conundrum. Also records creators, managers and archivists are imperfect, missing details and making debatable decisions. The enigma of AI lies rather in how it is accurate and how it makes mistakes. The underpinning premise of AI development is what Birhane describes as turning challenges to 'problem/solutions formulations' where a challenge is formalised as a mathematical problem and solved by developing an ap-

propriate solution to the problem.[11] There is an obvious risk that the challenge is only partly captured by the 'problem'. The premise of how machine learning approaches 'learning' is closely similar. At best quantitative learnability – how an AI can 'learn' to produce statistical solutions that have a low likelihood of being erroneous – can be a close approximation of real-life learnability – how a human-being can learn to make informed decisions – but seldom exactly the same.[12] However, instead of the approximations themselves being an issue, the more serious problem is how they are often portrayed as solutions to the challenge itself. The deliberate black-boxing of AI systems as automatic question-answering machines coupled with the inherent opacity of many popular AI techniques put them at odds with the conventional logic of archives as process-bound information. In contrast to simpler algorithms with all decisions and threshold values explicitly encoded, probabilistic techniques do not follow set procedures producing outputs based on pre-programmed sets of instructions and consequently, their outputs are not fully predictable. This means also that making AI to work requires more than digitisation and data cleaning. It is at least equally critical to understand and communicate the opportunities and limits of AI to diverse stakeholders of archives from records creating organisations to researchers and the public, and manage their expectations. Instead of insisting on giving a (false) impression of the completeness of search results generated by AI-powered search engine, their incompleteness and possible complementary search strategies can be highlighted through the design of search systems and how the results are presented. Automatically generated summaries can similarly be flagged and perhaps not being shown by default at all in order to indicate their unverified status. Similarly, even if it would be tempting to flag for how the use of automatic classification of records facilitates and speeds up the public release of records, perhaps the best strategy is to avoid overestimating its impact and downplaying the importance of human labour in the process.

RECORDING AI

A parallel question to how to manage AI in archives and records management tasks is how to get grips with it in the broader context of how records and recording intersect with diverse areas of social life. In parallel to how AI is increasingly used to manage records at archives, they are to an even larger extent used to generate and process information and records across the society. Currently perhaps the most debated examples relate to automated and AI-supported decision-making that are becoming increasingly popular in administrative processes in areas like processing welfare benefit requests and credit applications. AI is also gaining popularity in detecting potential cases of tax and benefits fraud and other criminal activities.[13] In all of such cases, AI is used in decision-making that need to be appropriately documented for adequate evidence much similarly to how AI needs to be kept accountable when used in archival tasks. In some cases, decisions can be completely automated, which raises a question of who *de facto* created a record: developer of an AI model or system, its operating organisation, or perhaps the AI itself. Even when AI is used to augment rather than automatise human decision-making, it is relevant to ask how and to what extent AI use should be documented. The borderline between using AI as a piece of office automation with at the most minor consequences to records creation processes and as their key constituent is far from being sharp and straight.

The increasingly apparent risks relating to hallucinating generative AI systems, biassed and incorrect outputs of AI classifiers and general difficulty to understand how AI techniques work and what problems they solve have unsurprisingly led to mounting demands of transparency, explanations and information on how different AI techniques function,[14] including calls for recordkeeping of AI development itself.[15] One of the measures proposed to provide neces-

1 Jaron Lanier: There Is No AI. In: The New Yorker 20 (2023), pp. 04–23.
2 Thomas H. Davenport: The AI Advantage: How to Put the Artificial Intelligence Revolution to Work. Cambridge, MA 2018; Michael I. Jordan: Artificial Intelligence – The Revolution Hasn't Happened Yet". In: Harvard Data Science Review 1.1. (2019); Spyros Makridakis: The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms. In: Futures 90 (2017), pp. 46–60.
3 Sarah Griffiths: Move over, Indy! AI Is Here. In: Engineering & Technology 18 (2023) H6, pp. 28–31; Stuart A. Thompson: The Dark Side of A.I. Is Here. In: The New York Times (2024), B1–B1; Niels van Hove: Commentary: AI Is Here to Automate the Knowledge Worker. In: Foresight 61 (2021), pp. 36–38.
4 On the Internet www.interparestrustai.org (accessed 04.11.2024). See also the article by Luciana Duranti in this issue of Archiv theorie & praxis.
5 Artificial Intelligence as a risk and opportunity for the authenticity of archives. On the internet at https://www.uu.se/en/department/alm/research/research-projects/ongoing-projects/artificial-intelligence-as-a-risk-and-opportunity-for-the-authenticity-of-archives (accessed on 04.11.2024).
6 E. g., Dick Kasperowski, Karl-Magnus Johansson and Olof Karsvall: "Temporalities and Values in an Epistemic Culture: Citizen Humanities, Local Knowledge, and AI-supported Transcription of Archives". In: Archives & Manuscripts 51 (2024) H2, pp. 3–22.
7 E. g., Ciaran B. Trace: Maintaining Records in Context? Disrupting the Theory and Practice of Archival Classification and Arrangement. In: The American Archivist 83 (2020) H2, pp. 322–372; Jason Franks: Text Classification for Records Management. In: Journal on Computing and Cultural Heritage 15 (2022). H3, 42:1–42:19.
8 E. g., Shitou Zhang et al.: Archives Meet GPT: A Pilot Study on Enhancing Archival Workflows with Large Language Models". In: iConference 2024 Proceedings 2024.
9 E. g., Naya Sucha-xaya: Documenting Events in Times of Crisis: Investigating Traditional and Contemporary Thai Archival Practices. In: Archivaria 93 (2023), pp. 42–71; Basma Makhlouf Shabou et al.: Algorithmic Methods to Explore the Automation of the Appraisal of Structured and Unstructured Digital Data. In: Records Management Journal 30 (2020) H2, pp. 175–200.
10 E. g., Camille Moret: Police Casework Recordkeeping in England and Wales: An Archival Denial of Evidence. In: Archives and Records 45 (2024) H2, pp. 176–195; Lise Jaillant, ed. Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitised Archival Collections. Bielefeld 2022.
11 Abeba Birhane: Algorithmic Injustice: A Relational Ethics Approach. In: Patterns 2 (2021) H2.
12 Edward B Kang: Ground Truth Tracings (GTT): On the Epistemic Limits of Machine Learning". In: Big Data & Society 10 (2023) H1, p. 20539517221146122.
13 E. g., Hilde Weerts et al.: Can Fairness Be Automated? Guidelines and Opportunities for Fairness-aware AutoML". In: Journal of Artificial Intelligence Research 79 (2023), pp. 639–677 cf. Alexandra James and Andrew Whelan: "'Ethical' Artificial Intelligence in the Welfare State: Discourse and Discrepancy in Australian Social Services". In: Critical Social Policy 42 (2022) H1, pp. 22–42.
14 Michael Ridley: Explainable Artificial Intelligence (XAI). In: Information Technology and Libraries 41 (20922) H2; Christian Borch and Bo Hee Min: Toward a Sociology of Machine Learning Explainability: Human–Machine Interaction in Deep Neural Network-Based Automated Trading. In: Big Data & Society 9 (2022) H2, p. 20539517221111361.
15 Jenny Bunn: Working in contexts for which transparency is important: A recordkeeping view of explainable artificial intelligence (XAI). In: Records Management Journal, 30 (2020) H(2), pp. 143–153.

sary transparency is *paradata*, defined in the InterPARES Trust AI project at the present as information about the procedure(s) and tools (for example, AI models and systems) used to create and process information resources, or in the broader sense, as information on practices and processes.[16] In the AI context, the research on paradata has inquired so far into the concept and how to frame it, and in a more practical sense, what would need to be documented relating to AI use.[17] On-going studies inquire in detail, for example, in how it is currently done, and what opportunities and challenges have been identified.

An underlying assumption of many of the propositions for advancing human understanding of AI stemming from the technology industry rather than contexts where AI is used is a belief that AI techniques can be comprehensively explained much similarly to how solving issues related to (un)fairness and bias in AI are matters of developing a better mathematical model that is fair and unbiassed.[18] The idea that documenting especially probabilistic AI techniques is a comprehensible solvable problem has been, however, increasingly challenged.[19] Rather, it has become increasingly apparent that documenting and keeping AI accountable requires thorough attention to the adequacy of explanations for specified contexts and purposes – with an obvious repercussion of a need keeping record of AI and AI use.

The diversity of AI techniques and to what extent they are explainable and how means that there is no one size fits all solution to paradata, or in more general terms, to recording AI and AI use. Drawing parallels from the research on documenting complex processes,[20] the problem is not solved by adding a field to a documentation standard or proposing a complex framework for representing chains of inferences and actions with a prompt to document AI systems and their use in sufficient detail. Instead of holding to the current technology driven approach that regularly downplays recordkeeping beyond immediate informational needs, there is a critical need to decide what is required of a record of AI processing and of AI processing of records. With criteria founded on the requirements stemming from recordkeeping rather than specific characteristics of technologies, AI techniques can be implemented and applied to support human processes instead of trying to force non-technical concepts such as accountability, ideals of fairness and lack of bias, transparency and sensitivity to a new mathematical model. If an appropriate record of AI used for a particular high-risk purpose cannot be generated, it could rather be a signal that a particular AI should perhaps not be used for that purpose at all rather than of the inadequacy of how a record is conceptualised.

REIMAGINING AI

What is then an appropriate record of AI, or more specifically, what is an appropriate record of which AI, and what characterizes appropriate use of AI in archives and records management? The difficulty of getting to grips with AI – how it fits specific tasks and how to record its pertinent aspects – might not interestingly be that much a question of what AI is but rather what it is imagined to be. Sociologist Lucy Suchman refers to the problematic assumption of the "thingness of AI"[21] as a critical reason why it is so difficult to discuss AI in a fruitful manner. The term is used as if there was an AI, a perpetrator that makes things happen, while in reality it refers to a vast constellation of existing and imagined techniques, technologies and ideas that have very little in common. Even if conceptual open-endedness can be helpful for catalysing discussion, the present concept of AI can be rightfully criticised of having become so vague that is has all but lost its role as a meaningful point of reference. It is too open, vague and fluid to properly facilitate communication between multiple, both technical and non-technical social worlds as a useful 'boundary object', a popular concept used to describe physical and conceptual things that facilitate communication and collaboration between communities.[22]

Rather than thinking of AI as a thing, Suchman urges to turning it controversial through considering instead of what the specific technologies and techniques under the AI umbrella are and what they do in specific contexts.[23] She refers to decision taken at the Center on Privacy & Technology at Georgetown Law to stop using the term AI.[24] Instead, the director of the centre, Emily Tucker, urges to describe what specific techniques do, for example, how face recognition is about matching images of human faces or hand-written recognition about converting digitised hand-written script to machine-encoded text. The major benefit of such a move is to bring the discussion back to what really can be done with particular technologies, what are their strengths and weaknesses, what is best achieved using other technologies or techniques – and also how to record the pertinent aspects of specific technologies and their use for archival purposes.

A parallel caveat to the purported thingness of AI is in its close association with the neo-liberal ideal of speed and increased efficiency through automation. Admitting that automation has many foreseeable advantages, it is worth considering in the spirit of slow movement[25] what are the effects of automation and speed, and why certain processes previously have been slow and manual – not only because of the lack of appropriate technologies and sophisticated processes. Automation has been advocated as a new archival paradigm,[26] but rather than accepting it as a thing similar to how Suchman describes AI imaginaries,[27] it is worth resisting and seeing as potentially equally controversial as AI. The same applies to speeding up processes that previously have been slow. In many cases, slowness is a 'problem' but it is worth considering whether solving that problem will solve its underpinning challenge,[28] or merely cause a lot of unanticipated and unwanted side-effects.

By stopping to talk about AI and resisting the temptation of overvaluing speed, it is possible to shift discussion from being technology-focused to becoming more people-centric as advocated by as a critical move towards understanding and implementing automation in the everyday practice instead of subjecting human pursuits to the condition of poorly understood technologies.[29] Such a move is crucial to avoid using AI as a system developed to provide unknown answers to questions we are not yet capable of formulating and asking.[30] In archival settings, people-centrism means focusing on archival practice but also turning attention to the implications of specific technologies and techniques for records and recording in the broader societal context.

Rather than starting with the question of what technologies there are to use, it is more relevant to start by identifying what archival problems there are to solve. If the major problem is poor searchability, there are a plenty of useful techniques available from text summarisation based on large language models to relevance-ranking of search results based on matching search strings and target documents and beyond. At the same time, solving that particular problem does not automatically solve the challenge of *finding* all relevant records. Identification of particular types of records for appraisal or release can be enhanced, for example, using models that help to predict if a particular record contains specific type of information. However, solving that problem does not mean that necessarily all pertinent documents would be identified and that the resulting set of records would for sure contain only expected types of records. Much similarly, the use of AI in diverse administrative processes across the society to make decisions and generate records would benefit of careful consideration of what is the problem to be solved and to what extent different techniques actually solve the underpinning challenge.

CONCLUSION

Even if it spontaneously might feel reasonable to think that the most significant challenges posed by AI for archives and records management, both directly and indirectly through how it is entangled in an increasing number of societal processes, are technical by their nature, they are to a large extent elsewhere. At the heart of what makes AI as a 'thingy' concept, and the plethora of technologies and techniques flagged as AI difficult is how they in different ways affect how and what we feel we should keep accountable. Even if AI is purported to be artificial *intelligence* – that is something commonly considered as human but artificial – at least for now, it is only mimicking – even if increasingly often so well that it is difficult to tell the difference without a considerable effort. At the same time neural network-based AI techniques are operating in ways that do not follow the innate logic of how we have learned to understand how technologies work either. This makes it difficult to envisage how a particular AI could be held accountable in terms of how we should judge, criticise, accept, or refuse its outputs. While understanding how AI works in context and how it can be relied is important to different degrees in all areas of life, it is an especially critical concern for archives in how the accountability of all the technologies used fiddle with records is in the very essence of what archives and records management is all about. Currently we are not letting AI loose in archives only in how it is used for archival applications but also through the processes that generate and impact records and archives throughout the entire continuum of recordkeeping. AI is a technology to manage with, but to a far greater extent a technology that requires energetic managing.

Rapidly advancing artificial intelligence (AI) hold great promise to support archival work. At the same time, its diversity complicates grasping its implications through its use in archives and records management tasks and even more so, through its use for generating and processing records in different fields of the society. Understanding AI requires abandoning the simplistic ideas of AI being intelligent like humans and a single monolithic thing.

Isto Huvila
Uppsala University
Department of ALM
Box 625, 751 26 UPPSALA, Schweden
E-Mail: isto.huvila@abm.uu.se

[16] Scott Cameron and Babak Hamidzadeh: Preserving Paradata for Accountability of Semi-Autonomous AI Agents in Dynamic Environments: An Archival Perspective. In: Telematics and Informatics Reports 14 (2024), p. 100135.; Isto Huvila: Improving the Usefulness of Research Data with Better Paradata. In: Open Information Science 6 (2022) H1, pp. 28–48; Huvila, Improving the Usefulness of Research Data with Better Paradata.

[17] E. g., Cameron and Hamidzadeh, Preserving Paradata for Accountability of Semi-Autonomous AI Agents in Dynamic Environments; Patricia C. Franks: The Crucial Role of Paradata in AI Governance". In: SCEaR Newsletter Special Issue: Artificial Intelligence and Documentary Heritage. Paris 2024, pp. 54–60.

[18] Maarten Buyl and Tijl De Bie: Inherent Limitations of AI Fairness. In: Communications of the ACM 67 (2024) H2, pp. 48–55.

[19] E. g., Hans de Bruijn, Martijn Warnier and Marijn Janssen: The Perils and Pitfalls of Explainable AI: Strategies for Explaining Algorithmic Decision-Making. In: Government Information Quarterly 39 (2022) H2, p. 101666.

[20] E. g., in Isto Huvila, Lisa Andersson and Olle Sköld: Perspectives on Paradata: Research and Practice of Documenting Data Processes. Cham 2024.

[21] Lucy Suchman: The Uncontroversial 'Thingness' of AI". In: Big Data & Society 10 (2023) 2, p. 205395172312067 94.

[22] Cf. Susan Leigh Star: The Structure of Ill-Structured Solutions: Heterogeneous Problem-Solving, Boundary Objects and Distributed Artificial Intelligence. In: Proceedings of the 8th AAAI Workshop on Distributed Artificial Intelligence, Technical Report, Department of Computer Science, University of Southern California. Los Angeles 1988, pp. 37–54; Isto Huvila et al.: Boundary Objects in Information Science". In: JASIST 68 (2017) 8, pp. 1807–1822.

[23] Suchman, The Uncontroversial 'Thingness' of AI.

[24] Emily Tucker: Artifice and Intelligence 2022; Suchman, "The Uncontroversial 'Thingness' of AI".

[25] Vincanne Adams, Nancy J. Burke and Ian Whitmarsh: Slow Research: Thoughts for a Movement in Global Health". In: Medical Anthropology 33 (2014) 3, pp. 179–197.

[26] Siham Alaoui: Peut-on Parler de l'automatisation Comme Cinquième Paradigme Archivistique?" In: The Canadian Journal of Information and Library Science 47 (2024) H1, pp. 18–34.

[27] Suchman, The Uncontroversial 'Thingness' of AI.

[28] Birhane, Algorithmic Injustice.

[29] See e. g., Sarah Pink et al.: Everyday Automation: Experiencing and Anticipating Emerging Technologies. London 2022; Lanier, There Is No AI.

[30] Jesse Josua Benjamin: Machine Horizons: Post-Phenomenological AI Studies". Unpublished Doctoral Dissertation. Enschede 2023; Hans-Jörg Rheinberger and Peter McLaughlin: From Experimental Systems to Cultures of Experimentation". In: Concepts, Theories, and Rationality in the Biological Sciences. Ed. by Gereon Wolters and James G. Lennox. Pittsburgh 1995, pp. 107–122.