Documenting AI Use in Humanities Research

Isto Huvila[1]

[1] *Department of ALM, Uppsala University, Thunbergsvägen 3H, Uppsala, Sweden*

Abstract

This paper explores the critical need to document the use of Artificial Intelligence (AI) in humanities research. While AI offers efficiency and analytical power, its application raises concerns about transparency, bias, and reproducibility. Existing documentation frameworks often emphasise technical aspects, overlooking the human and contextual dimensions vital to humanities scholarship. Drawing on cross-disciplinary literature, the paper advocates for integrating paradata (process-related meta-information) to capture both technical and human facets of AI use. It proposes shifting the focus from speculative future needs to documenting the transformation AI is intended to achieve within specific research contexts. Practical strategies include combining automated tools with reflective documentation practices and providing clear explanations of the purpose and expected outcomes of AI use. The paper calls for infrastructural support and a rethinking of documentation sufficiency to enhance understanding, reuse, and accountability in humanities research.

Keywords

paradata, artificial intelligence (AI), documentation

1. Introduction

Adequate documentation of Artificial Intelligence (AI) systems is critical for the transparency and accountability of their use and understandability of their outcomes in humanities research. AI offers multiple benefits, including of being potentially less expensive and more efficient to use than competing techniques of processing data. In addition, they are argued to be less biased and capable of achieving higher levels of accuracy in various tasks [1, 2] many of which are pertinent to humanities scholarship. At the same time, there are, however, many well-known risks. Outcomes of AI systems can be unpredictable, difficult to understand and fallacious [3, 4, 2]. AI systems have also shown tendencies to discriminate, and to replicate and perpetuate historical societal biases [5]. In spite of the critical importance of explainability of AI and AI use [4], documenting AI use in humanities research to avoid production of biased and fallacious data and knowledge, so far the bulk of the emerging research on documenting both technical and social aspects of AI systems and their use has focused on other fields of scholarship (cf. e.g., [6, 7]).

The aim of this paper is to draw attention to the critical importance of documenting AI use and include such documentation both in terms of documents and the practice of documenting AI use in infrastructures that support humanities research. Drawing on a cross-disciplinary reading of the literature with a focus on insights relevant to AI use in humanities context, this paper calls attention to why documenting technical characteristics of AI systems is not enough and proposes measures to improve the usefulness of the documentation of the human aspects of AI use.

2. Documentation of research processes in the AI era

Recent literature has been increasingly emphasising the importance of not merely documenting research outputs but also the process of how research is conducted. Adequate process documentation in terms of diverse forms of paradata (understood broadly as process information

Huminfra Conference 2025, Stockholm, 12-13 November 2025.

isto.huvila@abm.uu.se (I. Huvila)

or a “category of things that can be appropriated as informative of processes and practices” [8]; for a more in-depth discussion of the concept, see [9]) is crucial to understanding and accountability of research practices, outputs and the reusability of research data. At the same time, knowing and eliciting paradata that is useful for sometimes hard to predict future uses and determining how much paradata is enough can be exceedingly difficult [10].

A critical part of paradata is sufficient documentation of the use of any technologies applied in the research process. This applies to, for example, measurement and other equipment used for data capturing but also software and instruments used in data analysis and producing research outputs [10]. Current rapidly advancing AI technologies present a new set of opportunities for paradata generation [11] but also challenges to documenting the technologies themselves when applied in research processes [12]. Much of the work relating to AI use in research has focused partly on debating what types of AI use are ethical and partly on how to properly acknowledge AI use. Much less attention has been directed on how and what to document about the AI in the loop [13, 7] and even fewer studies have addressed so far the issue of AI documentation in the context of humanities research (exceptions e.g. [14, 15]). Even if much remains to be done in developing both theoretical and practical understanding of documenting AI use in research processes, there is an emerging transdisciplinary body of work relating to AI documentation [7], a part of which can help to provide relevant insights for paradata work also in humanities scholarship.

The principal concern of AI regulation and documentation has been on the diverse uses of AI in contexts judged as particularly risky pertaining to health, safety and fundamental human rights [2]. A growing number of frameworks and guidelines provide general guidance and directives of what is expected to be documented relating to AI systems and their use. The existing guidelines have, however, criticised of having multiple shortcomings. Many of the frameworks are focused on technical documentation of tools rather than their use [3] and how they integrate in human processes [16]. Similarly, a dominant theme in earlier observations is how documentation and many documentation guidelines give precedence to accountability, managing risks and liabilities and conformity to rules and regulations (e.g., [3][17]) or technical integration of systems [18] rather than promoting reproducibility and understanding and reuse of the outcomes of AI use. Königstorfer [7] notes further that proposed documentation techniques are seldom evaluated by governance experts or auditors. In parallel to high technical specificity in part of the proposed techniques, the frameworks that focus on outlining general principles of ethical AI implementation risk to be too unspecific to provide actionable guidance [19][10][7]. Privacy, research ethics and lack of permission to share process information are also identified as major impediments especially in proprietary contexts and research with human subjects involving sensitive personal data [20][21].

The inclination of AI documentation techniques to focus on documenting technical artefacts and preserving easily documentable aspects of AI use parallels with observations on research data makers’ similar concern when documenting data making, processing and use [22]. Their focus of interest differs from data reusers’ emphasis of information that makes data understandable and reusable [22]. This parallels with concerns expressed of the usability and usefulness of software-focused AI documentation produced according to many of the currently proposed frameworks [16].

3. Approaches to the documentation of AI use

Regarding the aspects of AI and AI use that needs to be documented, unsurprisingly considering the critique of technology overemphasis, the literature underlines the need to document the technical design and functionality of AI systems, including the AI model [6], and its training data [16]. In addition, the documentation should include information on the application context of the system and its development process [16]. Königstorfer [7] has analysed AI documentation techniques and identified three main aspects of AI that are recurringly in the focus, including the documentation of training data, application domain and design decisions. Moreover, similarly to the literature on research documentation and sharing (e.g., [9][23]), the AI literature underlines the importance of the understandability of AI documentation and balancing the costs and benefits of producing it [16].

Techniques proposed for documenting training data include approaches based on providing summary statistics and visualisations of datasets. Tools exist for analysing diverse aspects of datasets, including bias and fairness, and facilitating the creation of relevant documentation [24, 25, 26]. Unstructured data is generally more difficult to document and often relies on descriptive metadata rather than summary data [7]. The techniques proposed for collecting concise summary documentation of the paradata on the use of AI, including various aspects of collecting and processing training data, application domain, design decisions, and deployment and use of AI are predominantly based on guidelines formulated as templates and checklists. Such frameworks include, for example, such frequently cited techniques as Model Cards [26] and FactSheets [13]. There are also tools like Jupyter Notebooks and workflow tracking systems that allow to diverse extents automatic documentation of computational processes [27][28].

In addition to AI specific documentation techniques, the majority of the methods applicable for general paradata documentation are also useful for documenting AI use. Liu and Huvila [11] provide examples of categories of such methods including methods descriptions in form of formal metadata, narrative descriptions, recording, logging, research plans and prospective workflows. As suggested by Juneström and Huvila [29], such approaches and multiple others can be used to document paradata prospectively before, during, and retrospectively, after AI use.

4. Challenges and ways forward

As with paradata in general, the key challenges with AI use specific paradata are how to elicit adequate documentation and how to determine what to preserve (cf. [10]). Both earlier paradata [30, 22, 31] and AI documentation research [7] point to that while user needs are a critical premise of what documentation is relevant, they are also a fleeting target. Expressed needs vary between different users and uses that are associated with diverging perspectives to what aspects of AI use require explanation and transparency. Humanities research is generally characterised by a greater diversity of research practices than paradigmatically more homogenous disciplines [32, 33]. This accentuates the challenges of documenting AI use. It is unlikely that a single documentation standard could capture all information relevant for all humanities research, or even for a single discipline or discipline-antagonistic "data culture" [21].

Much similarly to process documentation in general [10], undoubtedly the most critical challenge of documenting AI use especially in humanities research context is how to capture enough while adding as little to the effort of documentation as possible. When documenting AI use, it is important to try to strike a balance between focusing on keeping (relatively) easily collectable and preservable technical information directly from and in research infrastructure services and identifying, generating and keeping typically more eclectic information about the use of AI to an extent that is necessary for understanding the systems and its use. Keeping certain technical information, such as summary information of training data and design specifications of models and algorithms, on AI algorithms and uses on the basis of its easy collectability and preservability is sensible. Similarly to collection of paradata in general [11], a part of the documentation of AI systems can be automated [7]. In contrast to keeping such information, assuming that it would automatically be enough is, however, highly problematic [11]. The ease of keeping specific types of information also comes with an immanent risk of keeping too much that makes searching and preserving the information both too laborious and resource-demanding [34]. Instead of assuming that keeping existing information would solve the conundrum, a critical question to consider is the sufficiency of documentation. In cases when sufficiency can be to a reasonable extent specified, evaluated and audited, such measures should be considered. In contexts like the humanities research where a single measure of sufficiency is close to an oxymoron due to the diversity of practices and theoretical perspectives, an approach worth considering is to try flip the perspective. Instead of approaching sufficiency from imagined and sometimes unimaginable user perspectives, a more practical line of action ought to be to approach it from the premises of how a specific instance of AI use is conceived from the perspective of the AI use itself.

First, instead of trying to produce documentation for others and assuming that it provides expected type of transparency, a potentially useful complementary strategy is to try to produce a *brief description of what each piece of documentation is attempting to achieve and how*. For example, such documentation might contain a brief explanation of how a descriptive summary of training data and a narrative description of the principles of how an algorithm works are expected to allow an individual with basic skills in computational methods to understand the logic of how the documented system generates its outputs.

Second, rather than assuming that a meticulous step-by-step description of what was done could capture all relevant information, a potentially powerful complementary approach is to produce a *documentation of what a particular AI system was used for and what it was considered to have achieved in the context of its use*, that is, what was the purpose of AI use and how the outputs and outcomes of an AI system were considered to compare to its inputs. A pertinent critique of present process documentation practices is that it is not always clear what documentation is attempting to achieve [35]. The general contextuality of process documentation [36] increases the risk that acontextual procedural descriptions are especially unclear in this respect. In contrast, when the focus is put on explicating the transformation AI is used to achieve, the likelihood of being able to convey a more comprehensive understanding of the *use* of AI can be expected to be greater.

The crucial step in switching perspective from speculating on future user needs in cases when they are unknown to taking the AI use itself as a starting point is not to produce two new categories of documentation but rather to take a perspective that can be taken and utilising it as a starting point for producing meaningful documentation. This is also a task where documenters could benefit of infrastructural support. In addition to the general need of more support in process documentation [37], eliciting AI use oriented documentation requires that documenters focus on producing a best possible account of their own practices instead of trying to adhere to a set of guidelines. While humans generally have difficulties to explain their actions and rationales in detail, it is still much easier than hypothesising what others might need to know about them. Assessing the sufficiency of documentation is also likely to be easier when it is done from a specific, to the documenter familiar, perspective rather than approached as a measure without an explicit point of reference. Furthermore, articulating the perspective makes it explicit for those consulting the documentation, simultaneously helping them to make sense of it.

5. Conclusion

Documenting AI use in humanities research and conveying such documentation is critical for understanding research processes, their direct outputs from research data to diverse forms of publications and broader outcomes. The fast-evolving AI techniques and their uses mean that the adequacy of documentation is a fast-moving target. There is an urgent need to develop a better understanding how to document AI use in humanities research, what needs to be documented and how documentation and the work of documenting should be incorporated in and taken into account in the development of humanities research infrastructures. Major challenges in the process include how to capture enough of the human side of the AI use processes and how to decide what is sufficient documentation. Rather than trying to speculate on future needs, an approach worth considering is to switch perspectives to focus on documenting the transformation AI is used to achieve rather mere interactions with a particular system, and to reflect and document what the documentation itself is attempting to achieve, for whom and how.

Acknowledgements

This work has been supported by the European Research Council (ERC) Grant number 818210 and InterPARES Trust AI project (SSHRC of Canada).

References

[1] I. Grossmann, M. Feinberg, D. C. Parker, N. A. Christakis, P. E. Tetlock, W. A. Cunningham, AI and the transformation of social science research, Science 380 (2023) 1108–1109. doi:10.1126/science.adi1778.

[2] M. E. Kaminski, Regulating the Risks of AI, Boston University Law Review 103 (2022) 1347–1411. doi:10.2139/ssrn.4195066.

[3] S. Cameron, P. C. Franks, I. Huvila, N. Mooradian, Navigating accountability: The role of paradata in AI documentation and governance, Journal of Documentation 81 (2025) 906–926. doi:10.1108/JD-01-2025-0009.

[4] A. Prescott, Bias in Big Data, Machine Learning and AI: What Lessons for the Digital Humanities?, Digital Humanities Quarterly 017 (2023).

[5] A. Foka, L. Eklund, A. S. Løvlie, G. Griffin, Critically assessing AI/ML for cultural heritage: Potentials and challenges, in: Handbook of Critical Studies of Artificial Intelligence, Edward Elgar Publishing, 2023, pp. 815–825.

[6] T. A. Brereton, M. M. Malik, M. Lifson, J. D. Greenwood, K. J. Peterson, S. M. Overgaard, The Role of Artificial Intelligence Model Documentation in Translational Science: Scoping Review, Interactive Journal of Medical Research 12 (2023) e45903. doi:10.2196/45903.

[7] F. Königstorfer, A comprehensive review of techniques for documenting artificial intelligence, Digital Policy, Regulation and Governance 26 (2024) 545–559. doi:10.1108/DPRG-01-2024-0008.

[8] I. Huvila, A Paradata Reference Model, Cambridge University Press, Cambridge, 2025, pp. 180–210.

[9] I. Huvila, Improving the usefulness of research data with better paradata, Open Information Science 6 (2022) 28–48. doi:10.1515/opis-2022-0129.

[10] Y.-H. Liu, I. Huvila, Methods for Generating and Documenting Paradata, Cambridge University Press, Cambridge, 2025, pp. 75–115.

[11] I. Huvila, Future Directions: Making Paradata Matter, Cambridge University Press, Cambridge, 2025, pp. 211–220.

[12] M. Arnold, D. Piorkowski, D. Reimer, J. Richards, J. Tsay, K. R. Varshney, R. Bellamy, M. Hind, S. Houde, S. Mehta, A. Mojsilovic, R. Nair, K. N. Ramamurthy, A. Olteanu, FactSheets: Increasing trust in AI services through supplier's declarations of conformity, IBM Journal of Research and Development (2019). doi:10.1147/JRD.2019.2942288.

[13] E. Frontoni, M. Paolanti, T. P. Lauriault, M. Stiber, L. Duranti, A.-M. Muhammad, Trusted Data Forever: Is AI the Answer?, in: M. Ramanath, T. Palpanas (Eds.), Proceedings of the Workshops of the EDBT/ICDT 2022 Joint Conference, CEUR-WS.org, Edinburgh, 2022, p. paper 1. arXiv:2203.03712.

[14] J. Edmond, J. Lehmann, Digital humanities, knowledge complexity, and the five äporiasöf digital research, Digital Scholarship Humanities 36 (2021) ii95–ii108.

[15] F. Königstorfer, S. Thalmann, Software documentation is not enough! Requirements for the documentation of AI, Digital Policy, Regulation and Governance 23 (2021) 475–488. doi:10.1108/DPRG-03-2021-0047.

[16] F. Königstorfer, S. Thalmann, AI Documentation: A path to accountability, Journal of Responsible Technology 11 (2022) 100043. doi:10.1016/j.jrt.2022.100043.

[17] R. Isdahl, O. E. Gundersen, Out-of-the-Box Reproducibility: A Survey of Machine Learning Platforms, in: 2019 15th International Conference on eScience (eScience), 2019, pp. 86–95. doi:10.1109/eScience.2019.00017.

[18] R. M. Srinivasan, E. Denton, J. J. Famularo, N. Rostamzadeh, F. Diaz, B. Coleman, Art sheets for art datasets, in: Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks, 2021.

[19] H. Semmelrock, T. Ross-Hellauer, S. Kopeinik, D. Theiler, A. Haberl, S. Thalmann, D. Kowald, Reproducibility in machine-learning-based research: Overview, barriers, and drivers, AI Magazine 46 (2025) e70002. doi:10.1002/aaai.70002.

[20] I. Huvila, L. S. Sinnamon, When data sharing is an answer and when (often) it is not: Acknowledging data-driven, non-data, and data-decentered cultures, Journal of the Association for Information Science and Technology 75 (2024) 1515–1530. doi:10.1002/asi.24957.
[21] I. Huvila, L. Andersson, O. Sköld, Patterns in paradata preferences among the makers and reusers of archaeological data, Data and Information Management 8 (2024) 100077. doi:10.1016/j.dim.2024.100077.
[22] I. Huvila, L. Andersson, Z. Friberg, Hs. Liu, O. Sköld, Paradata: Documenting Data Creation, Curation and Use, Cambridge University Press, Cambridge, 2025. doi:10.1017/9781009366564.
[23] O. Azeroual, T. Koltay, Research information in the light of artificial intelligence: Quality and data ecologies, 2024. doi:10.48550/arXiv.2405.12997. arXiv:2405.12997.
[24] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. III, K. Crawford, Datasheets for datasets, CACM 64 (2021) 86–92. doi:10.1145/3458723.
[25] O. E. Gundersen, Y. Gil, D. W. Aha, On Reproducible AI: Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications, AI Magazine 39 (2018) 56–68. doi:10.1609/aimag.v39i3.2816.
[26] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, T. Gebru, Model Cards for Model Reporting, in: Proceedings of the Conference on Fairness, Accountability, and Transparency, ACM, Atlanta GA USA, 2019, pp. 220–229. doi:10.1145/3287560.3287596.
[27] S. Ames, L. Havens, Exploring National Library of Scotland datasets with Jupyter Notebooks, IFLA Journal (2022). doi:10.1177/03400352211065484.
[28] M. Orzechowski, Ł. Opioła, I. L. Martínez, M. Ioannides, P. N. Panayiotou, Ł. Dutka, R. G. Słota, J. Kitowski, Integrated data, metadata, and paradata management system for 3D Digital Cultural Heritage objects: Workflow automation, federated authentication, and publication, Future Generation Computer Systems 174 (2025) 107964. doi:10.1016/j.future.2025.107964.
[29] A. Juneström, I. Huvila, Categorizing methods and approaches for generating and identifying paradata, Journal of Documentation (forthcoming). doi:10.1177/09610006251342811.
[30] I. Huvila, L. Andersson, O. Sköld, Researchers' data processing descriptions—Understanding paradata creation practices and their underpinning instrumentalities, Journal of the Association for Information Science and Technology (2025). doi:10.1002/asi.70003.
[31] L. Börjesson, I. Huvila, O. Sköld, Information needs on research data creation, Information research 27 (2022) isic2208. doi:10.47989/irisic2208.
[32] H. Lönnqvist, The research processes of humanities scholars, in: E. D. Garten, D. E. Williams, J. M. Nyce, S. Talja (Eds.), Advances in Library Administration and Organization, Emerald, Bingley, 2007, pp. 175–202.
[33] M. Bates, D. N. Wilde, S. Siegfried, Research practices of humanities scholars in an online environment: The Getty Online Searching Project report no. 3, Library and Information Science Research 17 (1995) 5–40. doi:10.1016/0740-8188(95)90003-9.
[34] T. Cook, Evidence, memory, identity, and community: Four shifting archival paradigms, Archival Science 13 (2013) 95–120. doi:10.1007/s10502-012-9180-7.
[35] I. Huvila, L. Andersson, O. Sköld, Concluding Discussion: Paradata for Information and Knowledge Management, in: I. Huvila, L. Andersson, O. Sköld (Eds.), Perspectives on Paradata: Research and Practice of Documenting Process Knowledge, Springer International Publishing, Cham, 2024, pp. 249–264. doi:10.1007/978-3-031-53946-6_14.
[36] S. Cameron, P. Franks, B. Hamidzadeh, Positioning Paradata: A Conceptual Frame for AI Processual Documentation in Archives and Recordkeeping Contexts, Journal on Computing and Cultural Heritage 16 (2023) 1–19. doi:10.1145/3594728.
[37] Z. Friberg, I. Huvila, Methods for managing paradata, in: I. Huvila (Ed.), Paradata: Documenting Data Creation, Curation and Use, forthcoming.