In the Pursuit of Archival Accountability: Positioning Paradata as AI Processual Documentation

PATRICIA C. FRANKS
San José State University

Abstract: Archivists and records managers accustomed to documenting the activities of human agents now find themselves responsible for documenting the activities of artificial intelligent agents. This is a seemingly impossible task due to the variety of AI-enabled technologies employed today and the difficulty of understanding the methods by which they interpret an input and return an output. In response to the uncertainty around accountability (and its twin concept of transparency) when employing AI techniques, the InterPARES Trust^{AI} research group launched a study to identify the unique documentation needs that emerge when information objects are created using AI tools. Our research leads us to the conclusion that a concept familiar in social science research, digital cultural heritage, and archeology could be adapted to meet the challenges of documenting the AI Process. That concept is *Paradata*.

Introduction

Artificial Intelligence in its many forms (e.g., machine learning, natural language processing, neural networks, computer vision, deep learning) can make our lives more comfortable, our businesses more efficient, and our government more responsive to our needs. Organizations in every sector are increasing their investment in AI tools and techniques to automate processes and better serve customers. Many in government view AI as a means to increase security and promote economic prosperity.

As the use of artificial intelligence continues to grow and algorithms and models become more complex, so do the challenges of explaining, justifying, and providing evidence of the actions and decisions carried out with little or no human intervention. It is, therefore, incumbent upon archivists and records managers, long accustomed to documenting actions and decisions of humans, to lend their expertise to documenting the actions and decisions of AI systems.

In 2021, a group of researchers from the InterPARES Trust^{AI} multinational, interdisciplinary project embarked upon a study to explore the need for documentation of the AI process and the suitability of adapting a concept from allied fields to understand and address the challenges. The result of our work to date suggests the concept, *Paradata*, can be used to document, explain, and provide evidence of the AI processes employed.

Problem Statement

As AI tools and techniques become more sophisticated and opaque, trust in AI systems (inclusive of data, algorithms and applications) continues to decline. A global study undertaken in 2023 by The University of Queensland Australia and KPMG revealed that three out of five

Published by Society of American Archivists, March 2024.

people (61%) are wary about trusting AI systems. The percentages differ among countries, with 40 percent of respondents in the United States indicating they are willing to trust AI systems somewhat, mostly, or completely but only 24 percent were highly or completely willing to accept them. The term accept in this case is equivalent to the concept of adopt for use. The percentages in Canada are lower, with 32 percent of respondents willing to trust AI systems to the same extent, but only 18 percent willing to accept them. This study recognized three characteristics of trustworthy AI systems: ability, humanity, and integrity.[1]

Other definitions of trustworthy AI provide more detailed lists of characteristics. Deloitte, for example, developed an AI framework that considers seven dimensions of trustworthy AI: private, transparent and explainable, fair and impartial, responsible, accountable, robust and reliable, and safe and secure. Most of the components are related to the systems themselves, but two—responsible and accountable—are often equated with the human element within the system. Responsible, however, can be attributable to AI systems if they are designed and implemented in a way that ensures fairness, interpretability, privacy, and safety for end users. But that leaves us with the question of who will be accountable for the outcome of the AI implementation if the other elements of AI trustworthiness are not met? And how does the accountable party provide evidence that the other elements have been satisfied?

Method

A literature review was conducted to understand the AI process and the ways in which the activities conducted during the process are recorded. From the archival and records management perspective, the researchers wished to understand what records needed to be retained to provide an explanation of the process and evidence that the process was carried out in an ethical manner. The well-known value of metadata (data about the information object) and the newer concept of Explainable AI (XAI or description of the AI model, its expected impact, and potential biases) were considered. Emerging guidance for the use and governance of AI was explored, including the proposed *EU AI Act*,[2] the *NIST AI Risk Management Framework*,[3] and the *White House's Blueprint for an AI Bill of Rights*[4].

In addition, we examined the tools recommended for documenting various phases of the AI process. Among them were Datasheets for Datasets endorsed by Microsoft,[5] Google Model Cards adopted by Salesforce,[6] and IBM Factsheets that recognize the documentation needs of various stakeholders including business user, data scientists, and validators.[7] While each will be useful in certain circumstances, we determined none provide a complete picture of the AI process from planning through implementation, monitoring, and validation.

Further review of the literature was undertaken to determine if and how information about "a" process is documented in other fields. We found the term *paradata* referred to in publications related to social science research, digital cultural heritage, and archeology to describe documentation of various processes. For example, as early as 2010 paradata was captured automatically as part of computer assisted data collection and used to provide transparency in virtual heritage projects. Among the paradata captured were all records, interviewer observations, time stamps, key stroke data, travel and expense information, and other data produced during the process.[8] In 2012, Martin J. Turner, acknowledged the distinction between metadata and paradata

in his article, "Lies, damned lies and visualizations: Will metadata and paradata be a solution or a curse?"[9] Today, paradata is collected by the U.S. Census Bureau as a by-product of the data collection process. Paradata helps the bureau identify potential areas of improvement, implement changes, and evaluate the effect of those changes.[10] The usefulness of paradata is also the focus of ongoing research. For example, CAPTURE (CApturing Paradata for documenTing data creation and Use for the REsearch of the future), a study led by Professor Isto Huvila at Uppsala University, is currently exploring ways in paradata can provide information about the creation and use of research data in the fields of archaeology and cultural heritage.[11]

Additional research was conducted to better understand all phases of the AI process and the actions and decisions that should be documented. This involved investigating different types of AI tools including machine learning, deep learning, computer vision, and more recently generative AI. This paper is the result of our efforts during this first phase of the paradata project.

Importance of Accountability

Organizations are accountable to their stakeholders (e.g., employees, shareholders, customers, citizens, and society) for any harm caused by their decision to implement artificial intelligence tools and technologies. *Accountability* in general terms is "the ability to answer for, explain, or justify actions or decisions for which an individual, organization, or system is responsible."[12]

AI accountability is expressed as the ability to explain and provide evidence of the actions or decisions made by a system of interlocking elements that accept input, process data using an algorithm and model, and produce a result. It is also the ability to take responsibility for the outcomes and impacts of an AI system. It further involves the ability to monitor, audit, and correct the system if it deviates from its intended purpose or causes harm.

Archivists and records managers understand that accountability regarding archival materials, *archival accountability*, "may be supported through provision of, or access to, records created and maintained through the normal course of a creator's activities."[13] It is this understanding we suggest must be introduced into the conversation around documenting the AI process.

Therefore, we contend that a broad systemic approach is needed to define and investigate: 1) multiple facets of Accountable AI systems (e.g., technical, legal, policy, governance, cybersecurity, ethical), 2) technological means to provide evidence needed to operationalize accountability within organizations implementing AI solutions, and 3) the development of a framework for educating and training employees to work effectively at the human-machine interface.

Results and Findings

The first stage of this study is exploratory. The outcomes will provide the foundation for the activities to be conducted during phase 2.

Defining the term: Paradata for the AI Process

Once the AI paradata team determined paradata would be useful to explain the AI Process from planning through implementation, definitions of the term used in statistical science, virtual heritage visualization, and archeology were examined. The result was the development of the following definition for Paradata as used in the AI context.

> *Paradata* as related to the AI process is defined as "information about the procedure(s) and tools used to create and process information resources, along with information about the persons carrying out those procedures."[14]

Distinguishing between Metadata, XAI, and Paradata

The team proceeded to distinguish between paradata and metadata based on the relationships and purpose for each type of "data." As shown in Figure 1, metadata is about the information resource while paradata is about the AI Process.

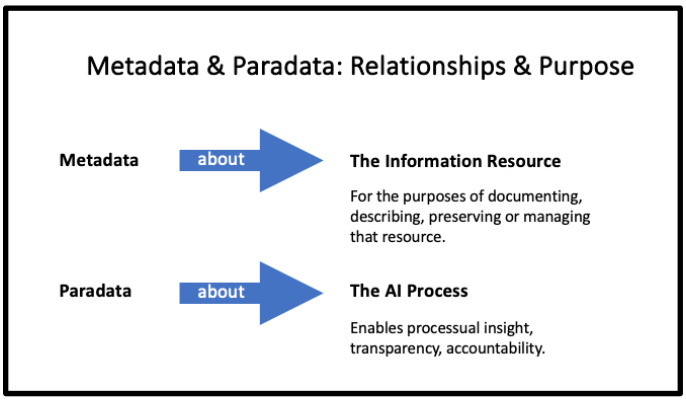

Figure 1. Distinction between Metadata and Paradata.

The team then compared Explainable AI (XAI) with Paradata. We determined that XAI is about the tools and that Paradata can be about that and everything else (see Figure. 2).

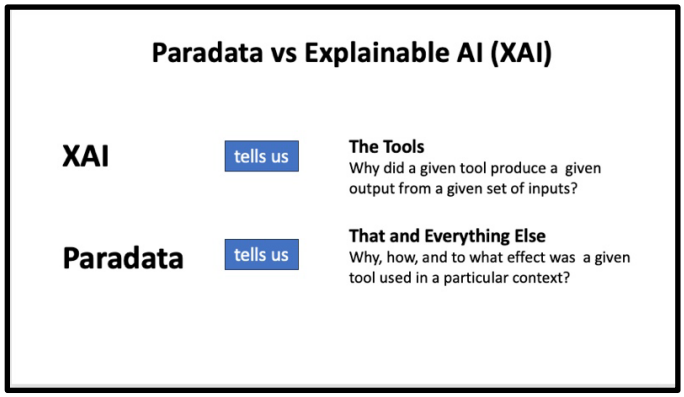

Figure 2. Distinction between XAI and Paradata.

Because of the various types of AI tools and the ways in which they are implemented in different circumstances, there is no standard way of documenting the actions they take and the decisions they make. This situation is also complicated by the needs of various stakeholders involved in each AI process from data scientists through end users.

One way to illustrate potential paradata is to use one model, such as the Machine Learning Lifecycle Model shown in Figure 3.

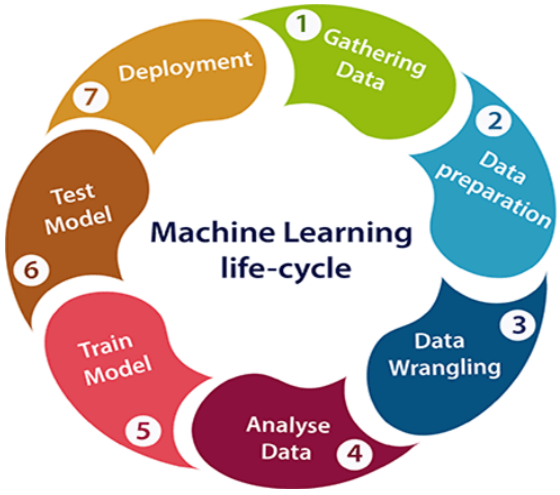

Figure 3. Machine Learning (ML) Lifecycle Model.

Actions taken during the Machine Learning AI process may include the following:

- Obtaining and formatting the dataset,
- Developing or obtaining the ML model,
- Training the model with the dataset that was prepared,
- Evaluating the model performance,
- Implementing the model, and
- Monitoring and possibly continuously improving the model with new data.

Jenny Bunn, fellow InterPARES[AI] researcher and Head of Archives Research at the National Archives of the UK offers a few practical questions that can be asked and answered related to documentation of the AI process.

- What records are created within AI research teams to document their processes?
- What records are created of the decisions to procure or deploy systems utilizing AI?
- What records are created of the decisions and impact of such systems?
- Are the records sufficient to meet existing legal provisions?
- Do the created records meet the required standards of quality?[15]

Paradata that might be collected to satisfy the need for documentation can be categorized into system paradata and operational paradata as shown in table 1.

Table 1. Examples of paradata to describe the AI/ML process.

System Paradata	Operational Paradata
AI Model (tested and selected)	AI policy
Evaluation and performance metrics	Design plans
Logs generated	Employee training
Model training dataset	Ethical considerations
Training parameters for model	Impact assessments
Vendor documentation	Implementation process
Versioning information	Regulatory requirements

Some of the paradata may be collected as part of the AI system's operations, such as versioning information and logs generated. Other paradata will be available as part of the organization's governing process, such as AI policy and regulatory requirements. Other paradata may need to be acquired or developed during the AI process, such as through the acquisition of vendor documentation or completion of impact assessments prior to development of the AI model as well as after implementation. One long-term goal of this study is to provide guidance for the collection of paradata on both the system and operational levels based on the levels of risk presented as described in the EU AI Act and the NIST AI Risk Management Framework. A second is to develop a method to aggregate and manage the necessary AI paradata so that it can be referenced along with the information object produced and any related metadata.

Example of AI Risk, Consequences, and Potential Role of Paradata

Best practice requires taking a risk-based approach to AI implementation. Figure 4 is based ongoing deliberations over the EU proposed Regulation on Artificial Intelligence (the EU Act) likely to be passed by the end of 2023.

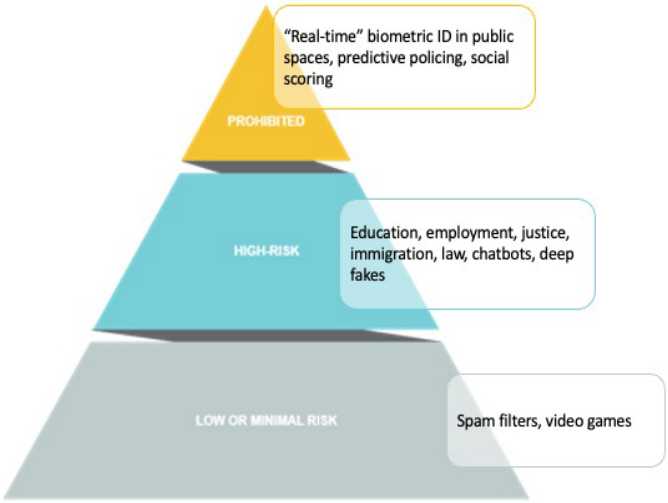

Figure 4. A layered risk-based approach to AI implementation.

Unacceptable uses of AI may be banned by governing entities. One example is social scoring, and another is predictive policing. Low or minimal risks, while frustrating, may be borne by the

user without resorting to action for damages on their part. Examples of this level or risk are the use of spam filters and video games. The high-risk category in the middle of the chart is where organizations must pay the most attention. AI used to make decisions or take actions related to education, employment, and immigration, for example, demand that the developers and users consider the impact of those AI systems on the individual, the organization, and society.

One example of the harm that may result from the US of AI in law enforcement is overreliance on AI.

> **The event:** On February 16, 2023, a young 8-month pregnant black mother of two was at home around 8 a.m. helping her 6- and 12-year-olds get ready for school when six Detroit police officers arrived at her door with an arrest warrant for carjacking and robbery.
>
> **The basis:** The arrest was based on the use of facial recognition technology that compared a photo of her with a video taken during the robbery.
>
> **The result:** The accused was falsely arrested and filed a lawsuit against both the police department and the detective ordering the arrest. The outcome of the lawsuit will reveal whether the police department, the detective, or both will be held accountable.
>
> **Contributing factors:** The police used a 2015 photo rather than a more recent 2021 driver's license photo in a photo lineup. Facial recognition algorithms are known as more likely to misidentify racial minorities. And the detective exhibited an overreliance on AI to make the arrest.

After the lawsuit was filed the police chief stated, "It was not an AI failure but an investigative failure."[16] However, not covered in the article from which the previous information was taken is the fact that this is the second lawsuit filed by a person of color against this city for false arrest related to the use of facial recognition. As a result of this second lawsuit, the police department stated they would update their policy about the use of AI. But if we want to understand the use of facial recognition in this case, we should explore the complete AI process.

Depending on which side of the lawsuit you are, you may want to gather documentation to determine if the AI system was implemented in a responsible manner. Answers to questions like the following can be found in the paradata collected:

- What was the AI facial recognition product used?
- What dataset was used to train it?
- What is the accuracy rate of the model?
- What type of vendor documentation was provided?
- Were impact assessments done before and after use?
- What ethical considerations were made?
- Was employee training provided?

This example underscores the number of actions taken and decisions made with one AI use case—facial recognition in law enforcement. While it may seem overwhelming to consider the

totality of AI implementations and the paradata that may be collected, we can begin to see questions and answers relating back to the system and organizational paradata examples provided earlier. The first is the system paradata: What do we know about the model, the data used to train it, its impact on an individual, organization or society if employed? What information (including accuracy rate) is provided by the vendor if this a purchased product or service. The second relates to the organizational paradata. What ethical considerations were made when employing the AI tool? Is there an AI policy? Does the AI policy relate to the organization's Ethics Policy? Were employees properly trained to use the AI tool?

Disseminating Information and Gathering Feedback

Based on our research, we are convinced that the use of paradata for the AI process will assist organizations in explaining, justifying, and providing evidence of actions taken by artificial intelligence systems. Since various stakeholders are involved in developing, implementing, and monitoring AI implementations—including data scientists, AI engineers, business executives, information governance professionals, and users—we began to disseminate our findings to and solicit feedback from diverse populations. The methods of dissemination include publication of peer-reviewed journal articles, conference presentations, webinars, panels, seminars, and workshops. The audiences for these presentations have included archivists, records managers, information governance professionals, attorneys, information managers, and security and privacy experts. The feedback gathered is positive as to the use of paradata to document data about the AI process. Exactly what paradata and how it can be collected and managed need further investigation.

Conclusion

Due to the growing interest in the application of AI tools and technologies, the risks presented, and the increased regulations under which organizations will operate, the time for the introduction of paradata to explain, justify, and providing evidence of the actions and decisions carried out with little or no human intervention has come.

Phase one of the AI and Paradata study described in this paper reveals a need for documentation of the AI process in the form of paradata to promote transparency and accountability. During phase two of this study, we will conduct case studies to evaluate the AI process and determine the paradata that must be collected and managed and how those tasks can be automated.

Endnotes

1. KPGM and The University of Queensland Australia. (2023). Trust in Artificial Intelligence, accessed November 9, 2023, https://assets.kpmg.com/content/dam/kpmg/au/pdf/2023/trust-in-ai-global-insights-2023.pdf

2. European Parliament. (2023, July 14). *EU AI Act: first regulation on artificial intelligence*, accessed November 11, 2023, https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

3. NIST. (n.d.). AI Risk Management Framework, accessed November 11, 2023, https://www.nist.gov/itl/ai-risk-management-framework

4. The White House. (n.d.) Blueprint for an AI Bill of Rights, accessed November 11, 2023, https://www.whitehouse.gov/ostp/ai-bill-of-rights/

5. Timnit Gebru et. al. Datasheets for Datasets, v8, last revised 2021, December 1). Cornell University, accessed November 12, 2023, https://arxiv.org/abs/1803.09010

6. Kathy Baxter. (2023, May 26). Model Cards for AI Model Transparency, accessed November 12, 2023, https://blog.salesforceairesearch.com/model-cards-for-ai-model-transparency/

7. IBM. (2023, November 9). Using AI Factsheets for AI Governance, accessed November 12, 2023, https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/factsheets-model-inventory.html?context=cpdaas

8. Franke Kreuter, Mick Couper, Lars Lybert. (2010) "The use of paradata to monitor and manage survey data collection."

9. Martin J. Turner, "Lies, damned lies and visualizations: Will metadata and paradata be a solution or a curse?" In *Paradata and Transparency in Virtual Heritage*, eds. Anna Bentkowska-Kafel, Hugh Denard and Drew Baker. New York: Routledge, 2012, p. 12.

10. U.S. Census Bureau (n.d.) Paradata, accessed November 10, 2023, https://www.census.gov/topics/research/paradata.html

11. Uppsala University. (n.d.) CAPTURE, accessed November 10, 2023, https://www.abm.uu.se/forskning/pagaende-forskningsprojekt/capture/

12. SAA. (n.d.). Dictionary of Archives Terminology, accessed November 8, 2023, https://dictionary.archivists.org/entry/accountability.html

13. Lara Wilson. "Accountability." In: *Encyclopedia of Archival Science*, eds. Duranti and Franks, New York: Rowman & Littlefield, 2015, pp. 3-5.

14. Jeremy Davet, Babak Hamidzadeh, Patricia Franks, Jenny Bunny. (2022). "Tracking the functions of AI as paradata and pursuing archival accountability," In: *Archiving 2022: Final Program and Proceedings*, 7-10 June 2022. Society for Imaging Science and Technology, Springfield, VA, USA, pp. 83-88.

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

16. Jennifer Henderson. (2023, August 8). “Black mom sues city of Detroit claiming she was falsely arrested while 8 months pregnant by officers using facial recognition technology,” *CNN*, accessed November 12, 2023, https://www.cnn.com/2023/08/07/us/detroit-facial-recognition-technology-false-arrest-lawsuit/index.html

Bibliography

Baxter, Kathy. (2023, May 26). Model Cards for AI Model Transparency, accessed November 12, 2023, https://blog.salesforceairesearch.com/model-cards-for-ai-model-transparency/

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

Davet, Jeremy, Babak Hamidzadeh, Patricia Franks, Jenny Bunny. (2022). “Tracking the functions of AI as paradata and pursuing archival accountability,” In: *Archiving 2022: Final Program and Proceedings*, 7-10 June 2022. Society for Imaging Science and Technology, Springfield, VA, USA, pp. 83-88.

Gebru, Timnit et. al. Datasheets for Datasets, v8, last revised 2021, December 1). Cornell University accessed November 12, 2023, https://arxiv.org/abs/1803.09010.

Henderson Jennifer. (2023, August 8). “Black mom sues city of Detroit claiming she was falsely arrested while 8 months pregnant by officers using facial recognition technology,” *CNN*, accessed November 12, 2023, https://www.cnn.com/2023/08/07/us/detroit-facial-recognition-technology-false-arrest-lawsuit/index.html

KPGM and The University of Queensland Australia. (2023). Trust in Artificial Intelligence, accessed November 9, 2023, https://assets.kpmg.com/content/dam/kpmg/au/pdf/2023/trust-in-ai-global-insights-2023.pdf

Kreuter, Franke, Mick Couper, Lars Lybert. (2010) “The use of paradata to monitor and manage survey data collection,” accessed November 12, 2023, https://www.academia.edu/29582626/The_use_of_paradata_to_monitor_and_manage_survey_data_collection

Turner, Martin J. “Lies, damned lies and visualizations: Will metadata and paradata be a solution or a curse?” In *Paradata and Transparency in Virtual Heritage*, eds. Anna Bentkowska-Kafel, Hugh Denard and Drew Baker. New York: Routledge, 2012, p. 12.

Wilson, Lara. “Accountability.” In: *Encyclopedia of Archival Science*, eds. Duranti and Franks, New York: Rowman & Littlefield, 2015, pp. 3-5.