All.Net Analyst Report and Newsletter

Welcome to our Analyst Report and Newsletter

Paradata and Forensics in Emerging AI used in Archives[1]

Paradata might be well described as ‘everything else’. That is, we have the data associated with records, we have the metadata like filenames, location in the fonds, inherent properties of the digital forms the records take, protection settings, authorship information, date and time stamps, and the rest of the elements of the Chain of Preservation (COP)[2] table, and yet, even with all of this information, we may find it hard to do the work required for examination.[3]

For example...

Reconstruction requires the ability to reproduce the results that reasonably reflect what took place. All of the records and metadata combined almost never include, for example, the data used to produce the models used by large language models, the models themselves, the software used to operate the models, their state of adaptation (often called learning) at the time the model was used, and the specific basis for the response generated to a query. So when trying to examine the evidence associated with a claim about what the archive did or did not do in response to a request for documents, there is typically no way to determine whether a claimed result was the actual result or whether that result accurately reflected a correct response to a query.

The everything else

In this case, we would have to capture all of the relevant information somewhere to do an accurate reproduction and claim that something did or did not, or that it could or could not be produced by the mechanism used by the archive to produce the result. And that is almost certainly beyond the capacity of most archives using these sorts of emerging AI for these purposes today.

The juridical context of testimony

The issue is further exacerbated by the fact that, in most juridical systems, records and other results produced by public archives is presumed reliable and authentic, and overcoming those presumptions requires a qualified expert to be able to testify in a manner that demonstrates those records are not reliable and authentic. Without the ability to reproduce what was done, it becomes far more difficult to opine on the process that took place and provide adequate proof of what actually took place.

It’s not just reproduction that is an issue

The process of examination of digital forensic evidence (called traces) generally includes one or more of; analysis, interpretation, attribution, and reconstruction; based on whatever was collected and retained reliably. Without access to the models and content forming the basis

1 This work was completed as part of our affiliation with the InterPARES Trust AI efforts performed in conjunction with the team at the University of British Columbia and archival, educational, governmental, and professional institutions from around the world.
2 See https://all.net/SoP/Archives/Metadata.html - a summary that reflects results of InterPARES 1, 2, 3, and Trust efforts at the University of British Columbia.
3 F. Cohen, “Digital Forensic Evidence Examination”, 2014 edition, ISBN # 1-878109-49-9

 Copyright (c) Fred Cohen, 2025 - All Rights Reserved

for the traces at issue, it may be arbitrarily difficult to make specific statements about what was produced, as opposed to general statements about hos such systems work and may produce results. This problem is made worse because current generative AI using large language models has a tendency to provide wrong answers[4], sometimes called "hallucinations".

Particularly disconcerting is the difficulty of dealing with:

- Inversion (forms of 'not'),
- Counting things,
- Production of reliable and authentic results from random applications of statistical 'next word' prediction from a vast corpus including large amounts of incorrect, inaccurate, or intentionally false information,
- Inconsistent results to sequences of identical queries from the same LLM system,
- Use of content not from the records at issue to generate results that seem reasonable but do not reflect the content at issue.
- Use of 'safety' protocols and other biasing elements in the production of results, and
- Reflection of biases of model information in results produced.

These produce:

- Incorrect analysis results, such as identifying the wrong operating environment associated with the production of traces, leading to incorrect interpretation of content;
- Over- or under-interpretation in the reconciliation of possible event sequences that could have produced the traces at issue;
- False attributions, as demonstrated in repeated instances of legal filings in courts using generative AI where the citations are non-existent or incorrectly analyzed in filings;
- Inaccurate reconstructions that fail to reflect the processes originally taking place.

Using paradata to mitigate these issues for archives

The design of paradata requirements for archives should presumably be driven by the juridical requirements of the archives. It is obvious that this includes the legal processes the archives support, and as public records, this includes the requirements of the relevant agencies and other sources of information the archives collect, preserve, and make available for use. Defining the uses and requirements of the sources should normally be part of the design of the archival processes, and this will generally be reflected in the applicable elements of the COP table selected for use, the descriptions of the archives and their functions, and the laws and regulations governing them.

Somehow these elements have to be translated into the specific activities of the archives and associated with the mechanisms that provide them and the level of certainty associated with the methods applied. If and to the extent the methods provide adequate certainty that the objectives they support will be accomplished under the design basis threat (which presumably

4 See https://all.net/Analyst/2025-09.pdf "Evaluating Generative AI for Business Applications" for some detailed examples.

 Copyright (c) Fred Cohen, 2025 - All Rights Reserved

has to be defined based on the same information), additional requirements for attaining unmet or inadequately met protection objectives for the desired level of certainty and that cannot reasonably be provided based on the other available information in the archives, paradata is the ‘bucket’ for the ‘everything else’ required.

Based on the otherwise requirements…

Parada becomes, in some sense, the last hope for meeting the requirements of surety for the archives. By identifying the information required for meeting the otherwise unmet objectives and identifying the methods by which that information may be used to achieve the objectives the set of information and methods may then be provided as paradata.

Of course as this is done, the paradata presumably becomes codified in records in the archives as well. In essence, paradata may be thought of as a set of records associated with the archives (and perhaps contained within them) required to meet juridical requirements not otherwise met by the archives.

What paradata should be included for AI?

Obviously this has to be based on the requirements as identified above, however, as a starting point, the following list is notionally identified:

- Software capable of performing the AI functions in the context of systems also available in the archives. Note this means that the physical archives will presumably need to contain appropriate hardware or that the software will have to include adequate emulation capabilities to operate in modern hardware and that preservation then requires conversion over time to maintain that capability as obsolescence occurs.
- Data not otherwise in the archives and used by that AI software to perform its function, in a form and format usable by that software and converted as required to meet changes in the software.
- Underlying factual content forming the basis for the software and data, including documentation, mathematical or engineering data required for understanding it construction and operation.

More generally, the archives should include how the AI and results it produces and produced came to be, how they came to the archives, and what the archives did with them.

But there are problems with this approach

The biggest problem, and the things that has substantially changed with the use of large language models, is that:

- The training data is very large, approximately 45 Tbytres for ChatGPT 3[5] before filtering, and reduced to 570Gbytes after ‘filtering’.
- The models are on the order of 800 Gbytes.

Models are updated fairly often today, and the result of using a service provider instead of doing everything internally is feasibility (it’s too expensive to do it yourself) and lack of transparency (in some cases). But even with complete transparency, the storage requirements are substantial for keeping copies of the training data.

5 https://community.openai.com/t/what-is-the-size-of-the-training-set-for-gpt-3/360896

 Copyright (c) Fred Cohen, 2025 - All Rights Reserved

Alternative approaches

Some different approaches to the requirement for paradata in terms of the archival environment is to use the AI to produce information adequate to not require as much (or any) paradata. For example, and depending on the juridical requirements:

- The operations of archives may be able to log enough information and associate enough of the provided results to reduce or eliminate the need for reproducibility.
  - Logging each request and response associated with searches may be adequate to provide evidence of what happened regardless of how it happened.
  - Providing explanations in proper language in the output, such as “based on X, this record was included in collection Y” along with the generic “records from A, B, and C as determined by D were not included in collection Y”, and so forth.
  - Instead of Yes or No decisions, the AI could produce high likelihood Yes and No decisions along with the metrics used and a 3rd (or more) category(ies) for less certain results that could be manually examined or reviewed by another method.
- Statistical information many be usable with samples of different categories of results to allow for subsequent understanding of the capabilities and limitations of these systems.
  - A study comparing the actual archive results from previous methods and the newer AI method might be performed to show the level of accuracy of the AI (and the previous method) in making the relevant decision(s). This then gets approved by the juridical body through some process, and the paradata is the analytical process and results, while the juridical decision is reflected in record(s) in the archives.

Other similar approaches may be adequate substitutes for processes that cannot otherwise be codified for strict reproducibility.

It’s should also be understood that humans make mistakes too, so using human performance as a baseline might be a critical part of the paradata associated with the AI in use.

Conclusions

This article is a bit nebulous and theoretical in the sense that it doesn’t define specifics for any particular archival environment. But that is only natural since the requirements are driven by juridical systems that vary widely.

However, the sciences involved are universal in the sense that the underlying principles of how digital systems operate, the nature of the traces they produce, and the methods available for examination don’t lose their validity based on the laws and regulations at issue.

Different processes and acceptance criteria are used for what can be done in practice, and for that reason, what can and should be included as paradata in archives for AI remains an issue.

As a general rule, absent any other constraints, it is a good idea to preserve the software and mechanisms, data, and factual content required to reproduce the required functions of the archives. Of course there are always other constraints present. Based on those constraints a number of other approaches may be applied to gain many of the advantages of cost and performance associated with emerging AI while mitigating the uncertainties associated with their use through properly generated and selected paradata.

 Copyright (c) Fred Cohen, 2025 - All Rights Reserved