I Trust AI

Luciana Duranti and Muhammad Abdul-Magee
Co-Directors, I(nterPARES) Trust AI
The University of British Columbia

International Symposium, Lanzarote
October 27, 2022

I(nterPARES) Trust AI

- I Trust AI is the 5th phase of the InterPARES project, directed by myself and Muhammad Abdul-Mageed, and funded, like the previous phases, by the Social Sciences and Humanities Research Council of Canada (SSHRC).
- Like the previous phases, I Trust AI focuses on
  - maintaining the trustworthiness of digital records overtime, and on
  - digital means of trustworthy access to and preservation of records in all media and form.
- What is different among the various phases of InterPARES is the technology that each phase examines for such purposes.

Trustworthiness

Reliability

The trustworthiness of a record as a **statement of fact**,

based on:

- the competence of its author
- the controls on its creation

Accuracy

The **correctness and precision** of a record's data

based on:

- the competence of its author
- the controls on content recording and transmission

Authenticity

The trustworthiness of a record that **is what it purports to be,** untampered with and uncorrupted

based on:

- identity
- integrity

InterPARES
TrustAI

Trustworthiness Issues with Digital Records

- In digital records, **content, structure, and form are not inextricably linked**
- The record as a stored entity is distinct from its manifestation on a computer screen, and its **digital components** have to be considered in addition to its **documentary form**
- Digital records are **vulnerable** (easy to destroy, lose, corrupt, tamper with, or become inaccessible if not protected) yet **persistent** (forever there, if not purposefully destroyed)
- When we save a record, we take it apart in its digital components. When we retrieve it, we generate a copy: there are **no originals** in the digital environment
- **Hence, it is not possible to preserve digital records: we can only preserve the ability to re-produce or re-create them**
- **Digital preservation** is the process of generating and maintaining **authentic copies** of digital materials and keeping them accessible during and across different generations of technology over time, irrespective of where they are stored
- **Authenticity is the major issue when it comes to digital records**

Diplomatic Authenticity

- **Diplomatics** has long been concerned with the authenticity of records and, since first developed in 1681, it has aimed to establish a scientific methodology for determining the authenticity of any record.
- This methodology examined the **form** of the record, that is, the rules of representation used to convey a message (those characteristics of a record that can be separated from the determination of the particular subjects, persons, or places that the record is concerned with) and the record's **degree of perfection** (whether it is a draft, a copy, or an original).
- Form is physical, i.e. the external make-up of a records (e.g. medium, ink), and intellectual, i.e. its internal articulation (e.g. salutation, preamble). If both correspond to the practice of the presumed or declared time, place, and author, then the record is authentic.
- The analytical approach of diplomatics aims to **establish on the record itself that the record is what it appears to be**, or what the person who submits it as evidence of a fact or an act claims it to be.

InterPARES
TrustAI

Archival Authenticity

- Archival science includes authenticity among the qualities that characterize every record, together with naturalness, impartiality and interrelatedness, and links it to them.
- **All records are authentic with respect to their creator**, that is, to the natural or juridical person who makes or receives them, and keeps them for further action or reference, that is, for its own legitimate purposes, even when, *diplomatically*, they are forgeries.
- Archives are authentic when they are made or received and kept **for the need to act through them,** and when they are preserved as **faithful witness of facts and acts** by the creator and its legitimate successors.
- Archival science, by **linking the record to its context of creation and preservation**, extended authenticity from being a property of the record itself to being **a property of procedures** and further tied it to **unbroken custody**

Authenticity in the Digital Environment

- There was no question in archival science that the identity of a record, and therefore its authenticity, resided in the provenance and documentary context of the record, but **this fact turned out to be linked to the immutability of a record affixed to a permanent medium, that is to its integrity.**
- In the late 1990s, we (the InterPARES Research Project—1998-2027) understood that, **in the digital environment, authenticity could no longer be assessed only on the basis of the records' context.**
- In fact, even if the relationships between and among the records established at creation remained intact throughout time, **the documentary component of the entity record could lose integrity** (a quality of the record that was never before part of the equation when establishing authenticity), because—as mentioned—its content, structure and form are no longer inextricably linked (content data, composition data, and form data are separate stored digital components).
- Thus, InterPARES returned to diplomatic authenticity and looked separately to identity and integrity.

Identity

Identity refers to the attributes of a record that uniquely characterize it and distinguish it from other records. These attributes include:

- the **names** of the persons concurring in its creation (i.e., author, addressee, writer, originator, creator);
- its **date(s)** of creation (i.e. making, receipt, filing) and transmission;
- the matter or **action** in which it participates;
- the expression of its **relationships** with other records (e.g. classification code); and
- an indication of any **attachment(s)**

Integrity

Integrity refers to the quality of **being complete and unaltered in all essential respects.**

We were never fussy about it. What if a document had holes, was burned on a side or the ink passed through?

The same definition of integrity was used with respect to data, documents, records, copies, records systems

As long as it was good enough to understand it, it had integrity...but how good is good enough in the digital environment?

Assessing Authenticity

The **fundamental difference** between the authenticity of analogue and digital records is in the fact that, while the authenticity of analogue material can be proven and verified on its face and only exceptionally is circumstantial or extrinsic evidence necessary, the authenticity of digital material cannot.

The assessment of the **authenticity of digital material**

- **is always an inference** based on extrinsic elements such as significant properties included in identity and integrity metadata, and
- **relies on circumstantial evidence** such as
  - the integrity of the system hosting it at any given moment in time,
  - the policies and procedures controlling such system, and
  - the technology encrypting the record or securing the access to it.

Could we use Artificial Intelligence to verify authenticity?

Artificial IntelligenceSystems

Artificial Intelligence Systems (AIS) are computing systems using algorithms capable of carrying out complex tasks that were once believed to be the sole domain of natural intelligence:

processing large quantities of information,

calculating and predicting,

learning and adapting responses to changing situations,

recognizing and classifying objects.

Research question:

Can we develop AIS for carrying out competently and efficiently all records and archives functions all the while respecting the nature and ensuring the continuing trustworthiness of the records?

AIS Issues

Artificial Intelligence Systems provide

- Inconclusive Evidence (based on probabilities)
- Inscrutable Evidence (no interpretability or transparency)
- Misguided Evidence (as good as the data provided)
- Unfair Outcomes (disproportionate impact on one group of people)
- Transformative Effects (challenges for autonomy and privacy)
- Non Traceability (hard to assign responsibility)

Plus

- The decisions AIS make are based on past decisions, and
- when it comes to human affairs, tomorrow rarely resembles today, and data and numbers can't say what has a moral value, nor what is socially desirable

Background

There have been several projects looking at AI in archives: they typically look at a particular tool in a specific context or even a single set of records.

- recurrent neural networks for classification of the content of large aggregations of records
- recommendation systems that connect relevant images to digitized letters, by using handwritten text recognition (HTR) to make old documents searchable
- chatbots that emulate human conversation through voice commands or text chats or both to help knowledge seekers find connected information
- a combination of Named Entity Recognition (NER), entity relations tools, and topic modeling to create visualization tools for the types of data stored on disk images

The Archival Problem

- Relying on existing off the shelf tools, as all the past studies on AI in archives have done, limits what challenges can be met, as it makes the needs of archives subservient to the larger field of machine learning
- It may be practical, but many tangible instances of bias have been found in modern machine learning models, often driven by *laissez faire* data collection practices
- This raises the questions of a) whether off the shelf tools are the best solution for the archival field and b) what AI could look like if this power relationship between AI and archives were reversed, with archival theory informing the creation of AI tools

I Trust AI Project Goal

The overall goal of I Trust AI is to design, develop, and leverage Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records by forming a sustainable, ongoing partnership producing original research, training students and other highly qualified personnel (HQP), and generating a virtuous circle between academia, archival institutions, government records professionals, and industry, a feedback loop reinforcing the knowledge and capabilities of each party.

Objectives

- Identify specific AI technologies that can address critical records and archives challenges;
- Determine the benefits and risks of using AI technologies on records and archives;
- Ensure that archival concepts and principles inform the development of responsible AI; and
- Validate outcomes from Objective 3 through case studies and demonstrations.

Studies

- Studies are **all international and interdisciplinary**
- Focus on all aspects of archival functions

1. Creation and use of trustworthy records
2. Appraisal and acquisition of archival material
3. Arrangement and description
4. Retention and preservation
5. Management and administration of records and archives
6. Reference and access

InterPARES TrustAI

Expected Outcomes

The project will improve upon existing tools and create new Machine Learning tools that will address archival needs, such as

- machine translation,
- image recognition and description,
- optical character recognition (OCR) and handwritten text recognition,
- text summarization and classification, and
- text style transfer for language civilization (e.g., removal of bias, hate, and sexism)

Indirect Outcomes

- New Professionals: by the end of the project, there will be well over 100 professionals who will have worked as student research assistants on case studies with test-bed organizations and who will spread the acquired knowledge, without counting all the future professionals taught such knowledge during their course of study
- Students from other disciplines, computer scientists, lawyers, etc. will understand and value the archival perspective in their work and the impact of records and recordkeeping on the broader society
- Knowledge co-creation: the project will enrich research in archival science, records management, AI, cybersecurity, information science, law, and ethics, through knowledge exchange and uptake between scholars and practitioners within and among those disciplines.
- Sensitizing AI developers, scholars, and other members of that community to the role of AI in record keeping and archival preservation and to the role of archival concepts and principles in AI design and development.

Participants

From 30 countries in 5 continents:

- 83 universities
- 22 organizations (businesses, international organizations)
- 16 regional, state, or national archives
- 118 academics
- 102 professionals
- 41 student researchers

Find Us

www.interparestrustai.org

@itrustai

www.facebook.com/interparestrust

InterPARES
TrustAI