I Trust AI

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

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.
- I Trust AI, as the name says, focuses on Artificial Intelligence Systems

InterPARES Trust

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, <u>tomorrow rarely resembles today</u>, and <u>data and numbers can't say what has a moral value</u>, nor what is socially desirable

Montreal Declaration Principles (2018)
Canada and the Association of Southeast Asian Nations

- Respect for Well-being principle
- Respect for Autonomy Principle
- Protection of Privacy Principle
- Solidarity Principle
- Democratic Participation Principle
- Equity Principle
- Diversity and Inclusion Principle
- Caution Principle
- Responsibility Principle
- Sustainable Development Principle

InterPARES Trust

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

- <u>Relying on existing off the shelf tools</u>, as all the past studies on AI in archives have done, <u>limits what challenges can be met</u>, as it <u>makes the needs of archives subservient to the larger field of machine learning</u>

- 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

InterPARES Trust

I Trust AI Project Goal

The overall goal of I Trust AI is to <u>design, develop, and leverage</u> Artificial Intelligence <u>to support the ongoing availability and accessibility of</u> trustworthy <u>public records</u> 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.

InterPARES Trust

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.

InterPARES Trust

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

InterPARE
TrustA

Studies

Studies address **common concerns across domains** (cultural heritage, government, education, private sector)

- Privacy, personal and sensitive information
- Ethical issues about records and about use of AI
- Education, teaching, and AI literacy
- Cultural heritage: digitization, access and retrieval, analysis and data mining
- Records management and metadata
- Municipal, regional, and national digital strategies
- Users PoV

InterPAR
Trust

Expected Outcomes

The project will <u>improve upon existing tools and create new Machine Learning tools</u> 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 <u>student research assistants</u> 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 <u>understand and value the archival perspective</u> in their work and the impact of records and recordkeeping on the broader society

- **Knowledge co-creation**: the project will <u>enrich research in archival science, records management, AI, cybersecurity, information science, law, and ethics</u>, 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 <u>role of AI in record keeping and archival preservation</u> and to the <u>role of archival concepts and principles in AI design and development</u>.

InterPARES Trust

Participants

- 88 partner public and private organizations in 5 continents: universities, archives, businesses, corporations
- 110 co-applicants
- 78 Collaborators
- Many graduate research assistants

InterPARES
Trust

Find Us

www.interparestrustai.org

@itrustai

www.facebook.com/interparestrust

InterPARES
Trust