

Public Infrastructure for Safe & Ethical Use of AI in Cultural Heritage

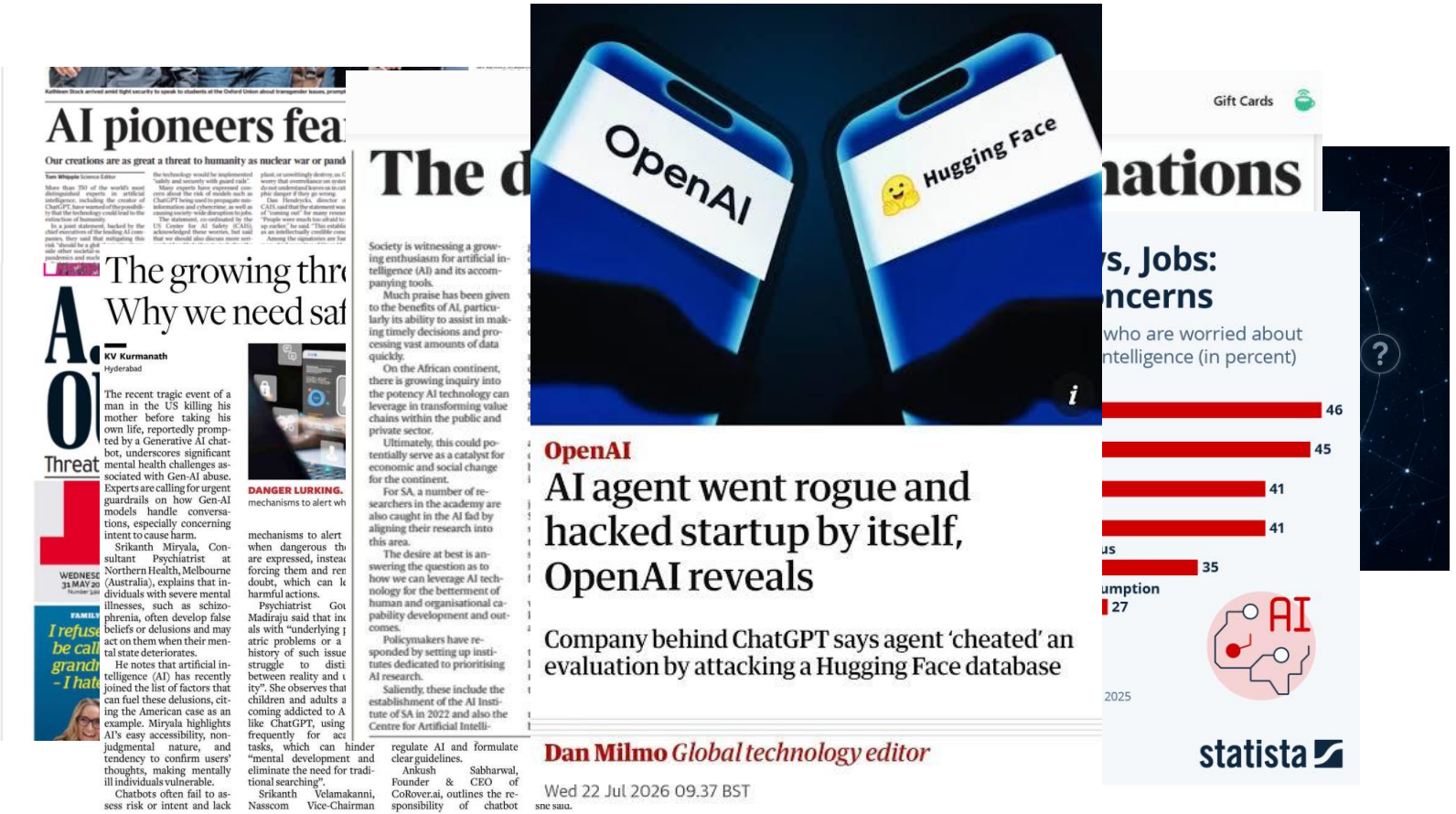
Final I Trust AI Symposium,
October 29, 2026

Victoria Lemieux, Professor Emerita, Institute for Computing, Information, and Cognitive Systems, University of British Columbia

AI's Trust Problem

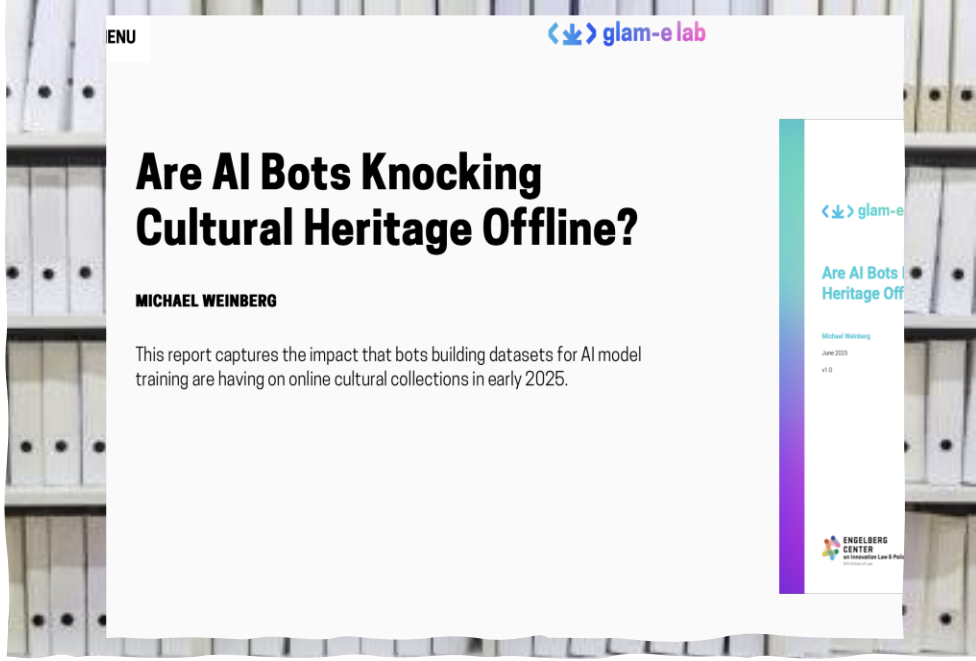


AI Has A Trust Problem



Archives & AI Trust Issues





copyright"

also getting
inquiries from
o are concerned
their work by AI
our IR"

"AI harvester activity is a huge problem for us. Like an ongoing DDOS attack against our digital platforms"

Trusting AI



The problem extends beyond AI models

Many efforts to build trust in AI focus on improving models.

Archivists and cultural heritage professionals raised wider concerns in interviews:

Rights and fairness

Bias and discrimination

Copyright violations

Privacy violations and data exploitation

Power and livelihoods

Data monopolies and concentration of power

Potential job losses

Accountability and impacts

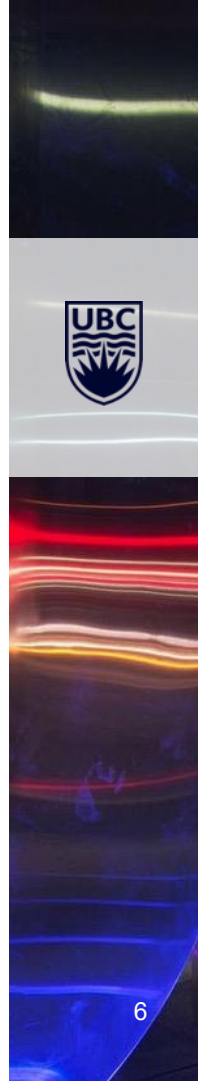
Lack of transparency and explainability

Environmental and social impacts of unfair AI

Non-compliance with growing regulation

Better model performance alone cannot address these issues.

We also need to trust the arrangements governing AI's use.

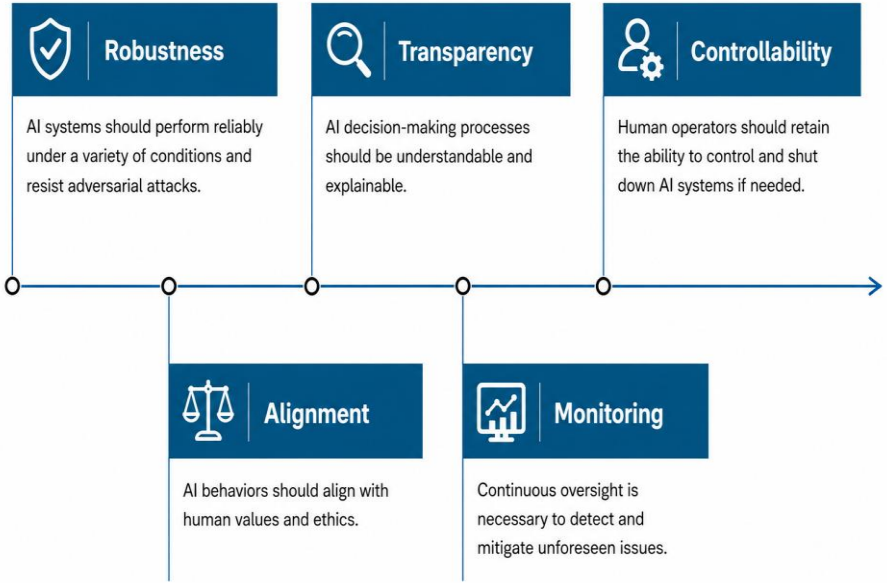


Some General Requirements for AI Trust

(without any claim to comprehensiveness)

Many different AI Governance Laws, Standards, Policies and Approaches are emerging . . .

- US Executive Order on Promoting Advanced Artificial Intelligence Innovation and Security (2026)
- Canada - governance is established through centralized standard-setting bodies and cross-border agreements, such as the Canada-EU Memorandum of Understanding on Artificial Intelligence (2025) and specialized technical bodies like the Canadian Artificial Intelligence Safety Institute (CAISI).
- EU AI Act (2024)
- ISO/IEC 42001 (2023)
- NIST AI 100-1 Risk Management Framework (AI RMF 1.0) (2023)



But laws, standards & policies alone are not sufficient – they are difficult to implement & enforce; they are post facto rather than ab initio

Infrastructure as a form of Governable Digital Territory

Language, records and infrastructure constitute social reality (Lemieux, 2022, chp. 8)

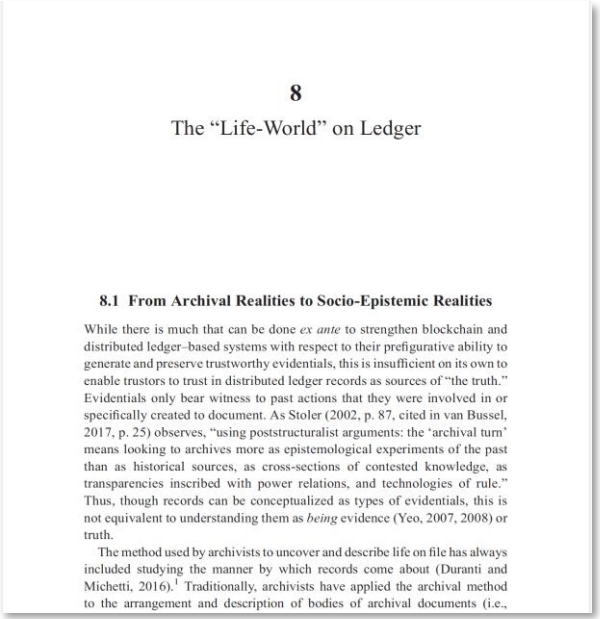
-Egan and Shera (1968) – recorded knowledge as a foundation of cultural continuity and collective understanding

-Hardwig (1991) - role of trust in knowledge

-Innis (1950) and McLuhan (1964) – communication media shape institutional reach and continuity across space and time

-Easterling (2014) – infrastructure as an active medium of governance

-Bratton (2016) – platforms reorganize sovereignty



8

The "Life-World" on Ledger

8.1 From Archival Realities to Socio-Epistemic Realities

While there is much that can be done *ex ante* to strengthen blockchain and distributed ledger–based systems with respect to their prefigurative ability to generate and preserve trustworthy evidentials, this is insufficient on its own to enable trustors to trust in distributed ledger records as sources of "the truth." Evidentials only bear witness to past actions that they were involved in or specifically created to document. As Stoler (2002, p. 87, cited in van Bussel, 2017, p. 25) observes, "using poststructuralist arguments: the 'archival turn' means looking to archives more as epistemological experiments of the past than as historical sources, as cross-sections of contested knowledge, as transparencies inscribed with power relations, and technologies of rule." Thus, though records can be conceptualized as types of evidentials, this is not equivalent to understanding them as *being* evidence (Yeo, 2007, 2008) or truth.

The method used by archivists to uncover and describe life on file has always included studying the manner by which records come about (Duranti and Michetti, 2016).[1] Traditionally, archivists have applied the archival method to the arrangement and description of bodies of archival documents (i.e.,

2020 NSERC Discovery Grant: Solving the Data Double Spending Problem
How can a person or an organization use AI while retaining authority over their records?
How can blockchain make those arrangements more trustworthy?

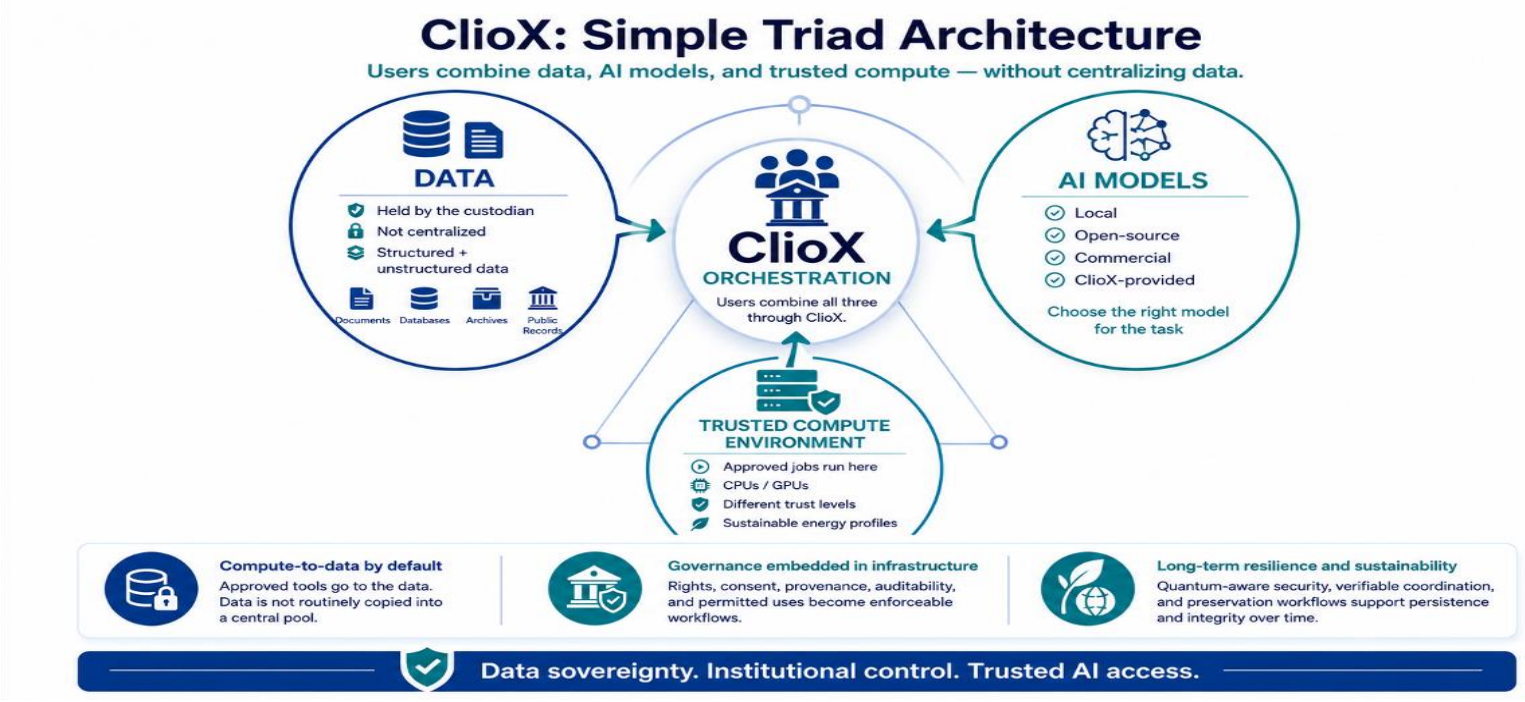
Clio-X as a Governable Digital Territory for AI Use



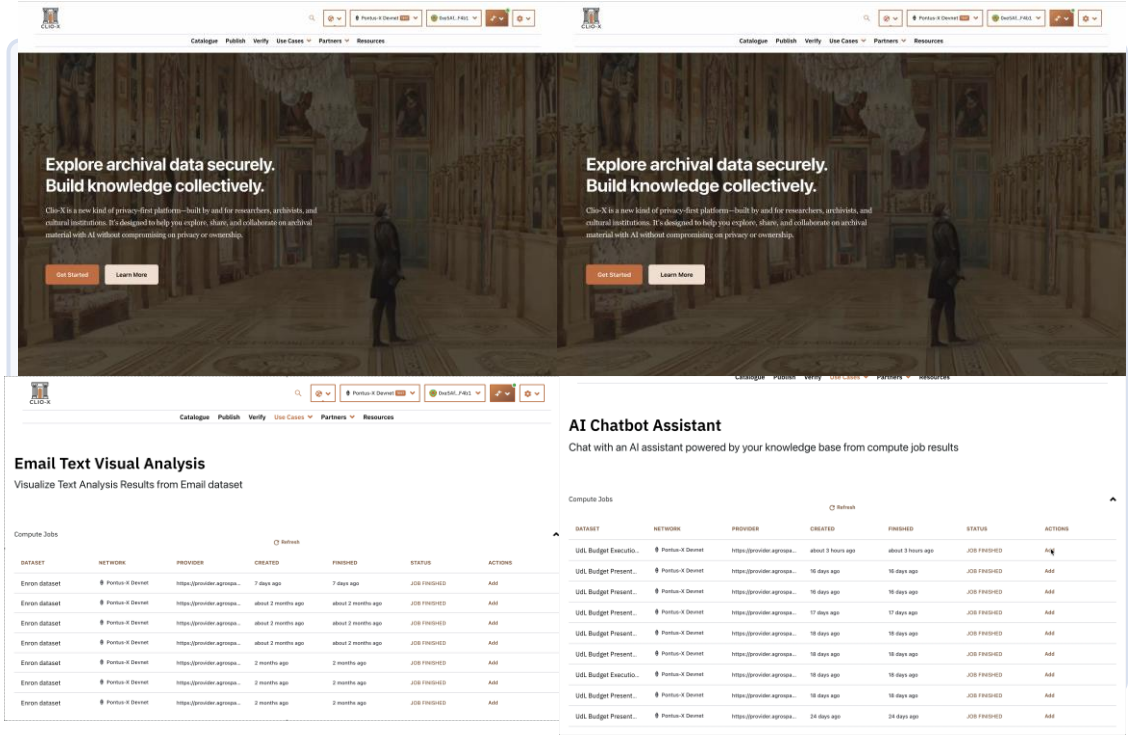
What is Clio-X?

Clio-X is a data space that uses blockchain to bring approved AI tools to records within controlled environments, enabling custodians to retain authority over access and use and placing AI trust into a workflow.

It was developed as one outcome of InterPARES Trust AI research (RA-06)



Clio-X turns authority over access and use and AI Trust into a governed workflow



HOW IT WORKS

01 An authorized user initiates an AI-assisted workflow involving fragmented or rights-governed data.

02 ClioX brings approved models and computation to the data, securely orchestrating access without centralizing or transferring the underlying records.

03 ClioX verifies identity, purpose, permissions, consent and licensing rules; automates approved low-risk actions and routes exceptions or consequential decisions to human review.

04 Reduced turnaround time, backlog and cost per case while data remains under custodial control and every governed interaction produces a verifiable audit trail.

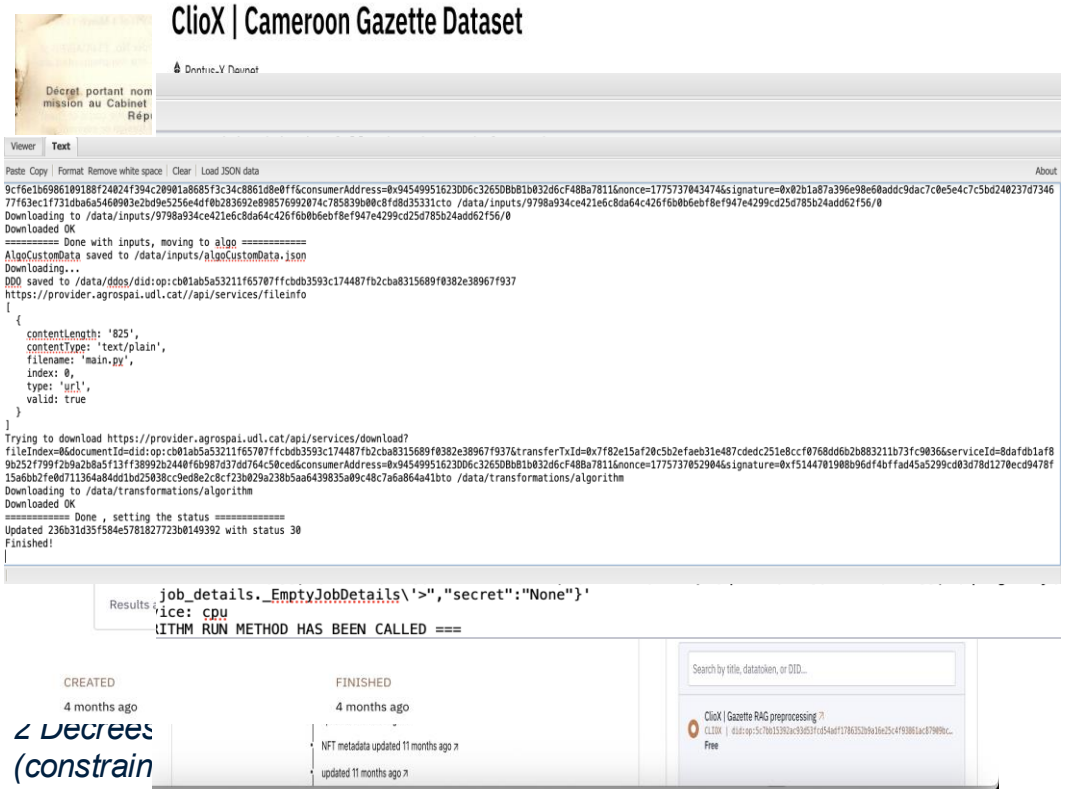
Experimenting with Clio-X



Making AI Governable: Clio-X, Public Records, and Community-Based Digital Statecraft in Cameroon

Design & Action research with **Esther Olembe, National Archivist of Cameroon**

Demonstrated core functional capability of Clio-X to enable organizations to use AI while retaining authority over records **(robustness, controllability & alignment),** transparency and enablement of accountability **(transparency & monitoring)**



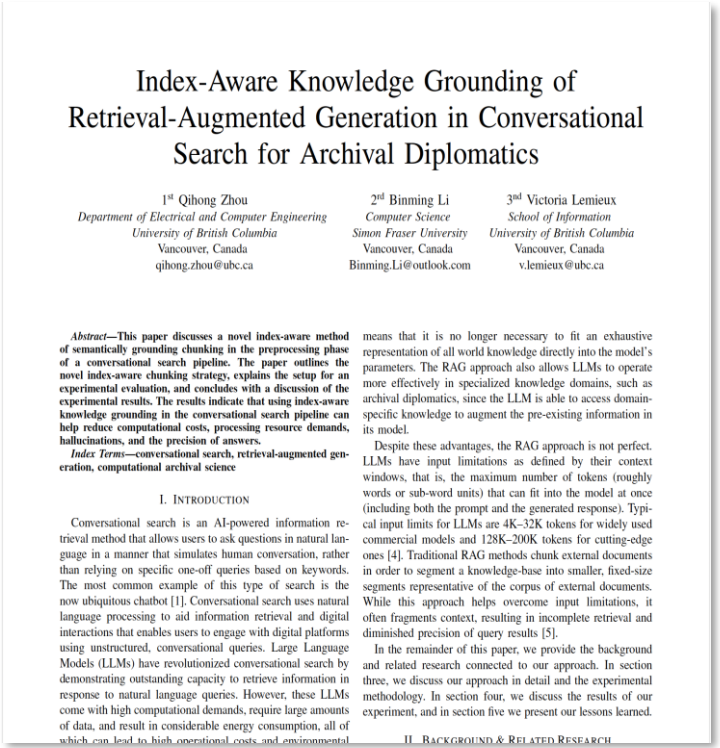
2 Decrees
(constrain
data that is digitized in Cameroon)

Privacy-Preserving, Accurate & Sustainable Conversational Search with Index-Aware RAG

Design research with two former students

Chris Zhou and Brandon Li

Demonstrated **Alignment** (Accuracy & Environmental Sustainability)



Index-Aware Knowledge Grounding of Retrieval-Augmented Generation in Conversational Search for Archival Diplomatics

1st Qihong Zhou
Department of Electrical and Computer Engineering
University of British Columbia
Vancouver, Canada
qihong.zhou@ubc.ca

2rd Binming Li
Computer Science
Simon Fraser University
Vancouver, Canada
Binming.Li@outlook.com

3nd Victoria Lemieux
School of Information
University of British Columbia
Vancouver, Canada
v.lemieux@ubc.ca

***Abstract*—This paper discusses a novel index-aware method of semantically grounding chunking in the preprocessing phase of a conversational search pipeline. The paper outlines the novel index-aware chunking strategy, explains the setup for an experimental evaluation, and concludes with a discussion of the experimental results. The results indicate that using index-aware knowledge grounding in the conversational search pipeline can help reduce computational costs, processing resource demands, hallucinations, and the precision of answers.**

***Index Terms*—conversational search, retrieval-augmented generation, computational archival science**

I. Introduction

Conversational search is an AI-powered information retrieval method that allows users to ask questions in natural language in a manner that simulates human conversation, rather than relying on specific one-off queries based on keywords. The most common example of this type of search is the now ubiquitous chatbot [1]. Conversational search uses natural language processing to aid information retrieval and digital interactions that enables users to engage with digital platforms using unstructured, conversational queries. Large Language Models (LLMs) have revolutionized conversational search by demonstrating outstanding capacity to retrieve information in response to natural language queries. However, these LLMs come with high computational demands, require large amounts of data, and result in considerable energy consumption, all of which can lead to high operational costs and environmental

means that it is no longer necessary to fit an exhaustive representation of all world knowledge directly into the model's parameters. The RAG approach also allows LLMs to operate more effectively in specialized knowledge domains, such as archival diplomatics, since the LLM is able to access domain-specific knowledge to augment the pre-existing information in its model.

Despite these advantages, the RAG approach is not perfect. LLMs have input limitations as defined by their context windows, that is, the maximum number of tokens (roughly words or sub-word units) that can fit into the model at once (including both the prompt and the generated response). Typical input limits for LLMs are 4K–32K tokens for widely used commercial models and 128K–200K tokens for cutting-edge ones [4]. Traditional RAG methods chunk external documents in order to segment a knowledge-base into smaller, fixed-size segments representative of the corpus of external documents. While this approach helps overcome input limitations, it often fragments context, resulting in incomplete retrieval and diminished precision of query results [5].

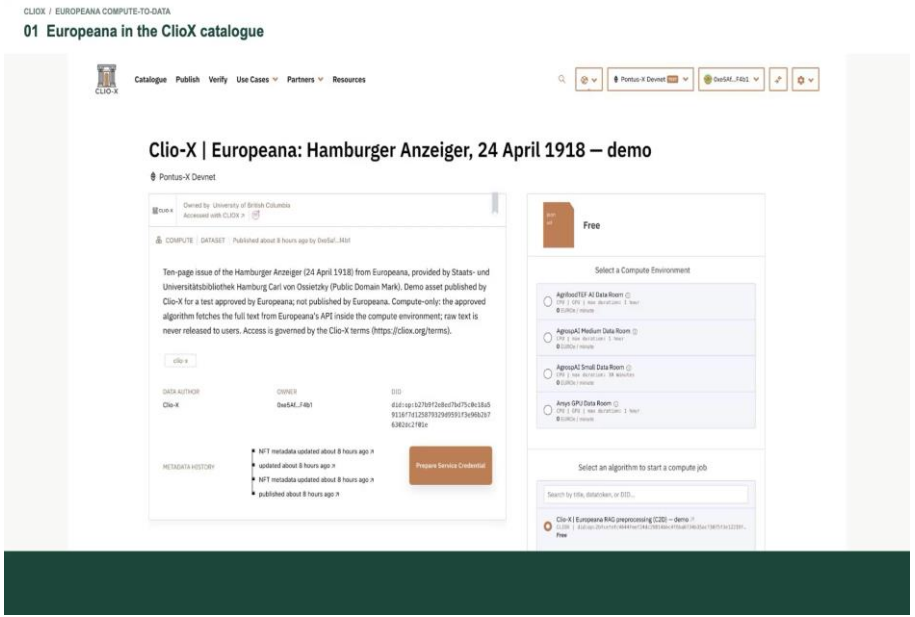
In the remainder of this paper, we provide the background and related research connected to our approach. In section three, we discuss our approach in detail and the experimental methodology. In section four, we discuss the results of our experiment, and in section five we present our lessons learned.

II. Background & Related Research

Adding Governed AI Compute to a European Cultural Heritage Data Space

Design research with **Brandon Li**
Developed a prototype with the capabilities to achieve functional capability to enable archives to use AI through Europeana while retaining

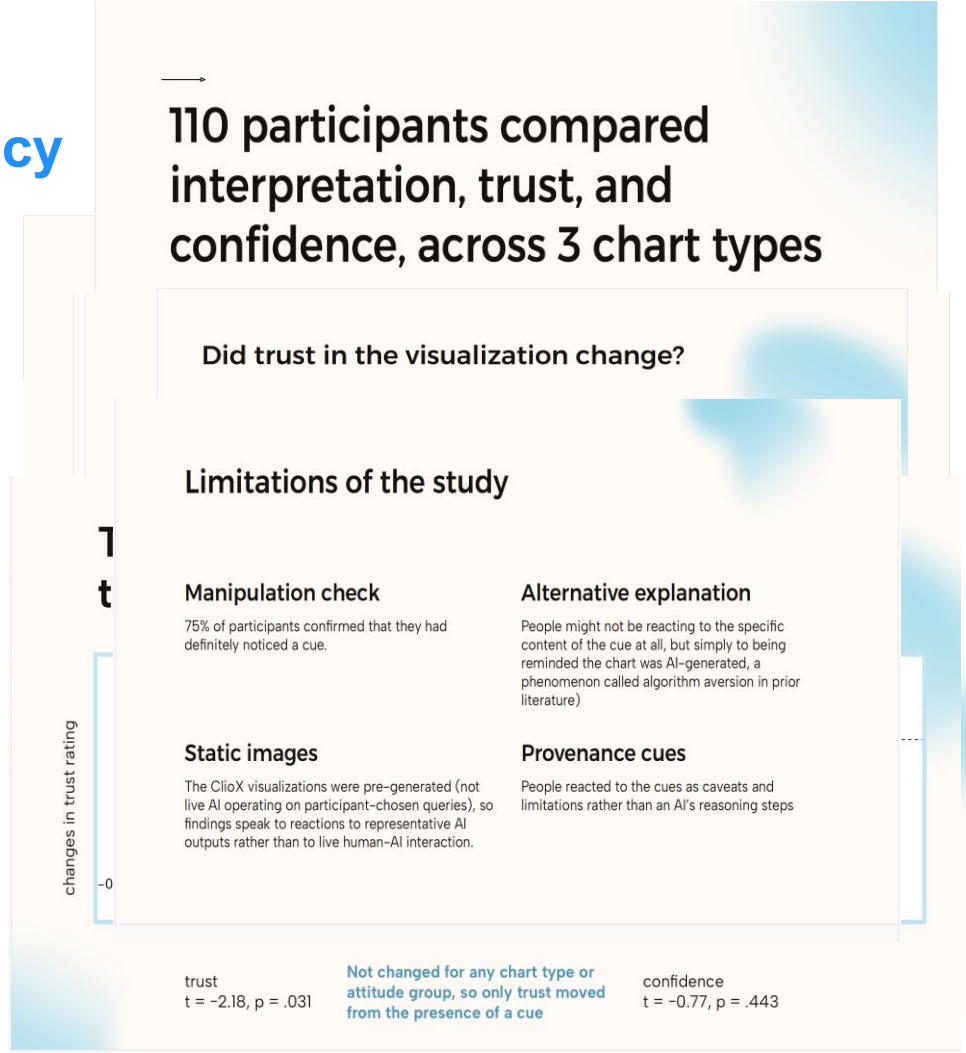
Demonstrated **Robustness, Controllability, Alignment, Transparency, enablement of Accountability, Transparency & Monitoring**



The Limits of Transparency

Experimental research with former student **Elena Chuang**

Supported a deeper understanding of the issues around **Transparency** and AI Safety



Indigenous AI Safety – Governing the AI Commons

Ongoing research funded by CIFAR in collaboration with **Jacob Taylor and other researchers in the National Indigenous Economic Prosperity Institute at First Nations University**

The project pairs Indigenous worldviews with computer science, embodying what Mi'kmaq Elder Albert Marshall termed "Two-Eyed Seeing" to show how the collective governance capabilities of blockchain can support **alignment** with a specific community's ethical and cultural values

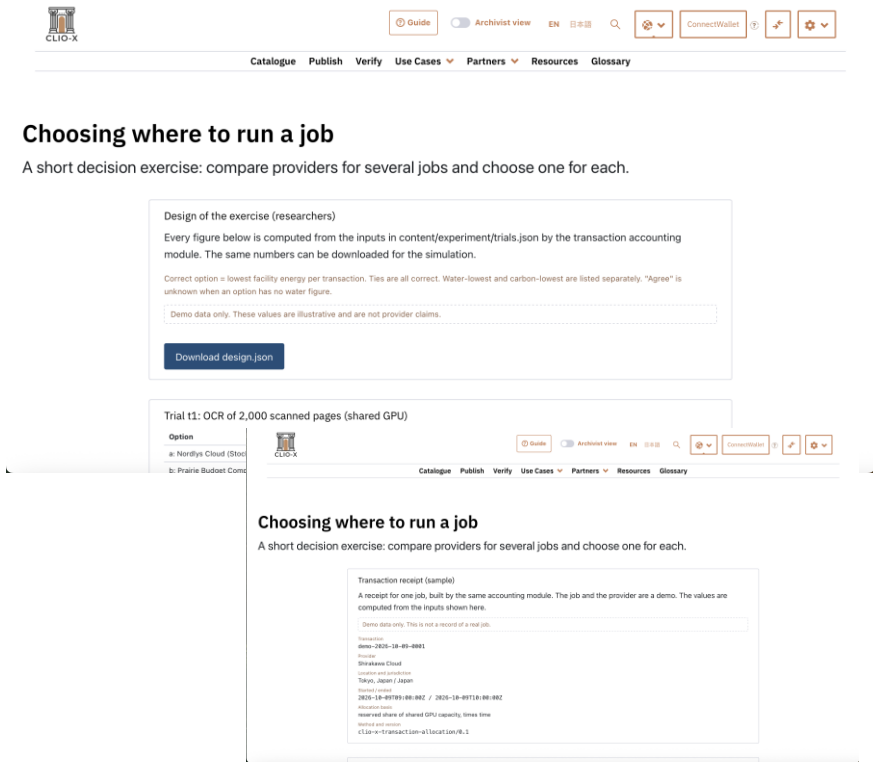


Transparency & Sustainability

Research with **Dr. Satoru Nakamoru**, Tokyo University, Japan &
Dr. Gabriella Laatikainen, VTT Labs, Finland

Designed a prototype that measures energy and water usage for Clio-X compute jobs. Planning to run computational consumer choice experiments to assess the effects of transparency.

Demonstrating **alignment**



Clio-X findings and anticipated findings

Experiment	Robustness	Transparency	Controllability	Alignment	Monitoring
Cameroon	✓	✓	✓	✓ Sovereignty preservation	✓
Index-Aware RAG				✓ Accuracy and environmental sustainability	
Europeana integration	✓	✓	✓	✓ Sovereignty preservation	✓
Indigenous AI safety				✓ Sovereignty preservation	
Environmental sustainability				✓ Sustainability	

Findings to date and anticipated findings are shown together.

Conclusion



Blockchain alone does not solve all AI Safety challenges.

It is strongest for:

- Governance
- Accountability
- Auditability
- Provenance
- Access control
- Institutional trust
- Decentralized coordination

It is weaker for:

- Model interpretability
- Hallucination reduction
- Bias mitigation
- Alignment of autonomous systems
- Frontier AI containment
- Semantic understanding of AI outputs

Blockchain is best understood as:

an enabling governance and trust infrastructure for AI Safety, rather than a complete AI Safety solution by itself.

Selected References

Lemieux, Victoria L, Rosa Gil, Faith Molosiwa, Qihong Zhou, Binming Li, Roberto Garcia, Luis De La Torre Cubillo, Zehua Wang, Clio-X: AWeb3 Solution for Privacy-Preserving AI Access to Digital Archives, ACM Journal on Computing and Cultural Heritage(forthcoming,).

Victoria Lemieux and Esther Olembe, Making AI Governable: Clio-X, Public Records, and Community-Based Digital Statecraft in Cameroon, Data for Policy Conference, Barcelona, Spain, September 8-10, 2026

Lemieux, Victoria L., and John Werner. “Protecting Privacy in Digital Records: The Potential of Privacy Enhancing Technologies.” ACM Journal on Computing and Cultural Heritage (2023). https://doi.org/10.1145/3633477

Lemieux, Victoria, Qihong Zhou, Binming Li, Clio-X: Cultural Heritage and Archives Ecosystem for Ethical Artificial Intelligence on Demand, Proceedings of the 29th Annual Archival Science Colloquium, Marburg Archives School, May 19, 2025. Marburg, Archives School, 2026

Jiho Lee, Jaehyung Jeong, Victoria Lemieux, Tim Weingartner, and JaeSeung Song, Blockchain and Responsible AI: Enhancing Transparency, Privacy, and Accountability through Blockchain Hackathon. In Proceedings of the Workshop on Computational Archival Science in Big Data, IEEE Big Data Conference, Washington, DC, December 9, 2025 Washington, DC: IEEE, 2025

Kwon, Soonhong, Wooyoung Sun, Victoria Lemieux, Nyam Yu Gang, Jong Hyouk Lee. “Hybrid PII Masking and RAG-based LLM Inference for Privacy-Preserving QA on Clio.” Proceedings of the ACM International Conference on Intelligent Computing and Emerging Applications 2025, La Trobe University, Melbourne, Australia.

Zhou, Qihong, Binming Li, and Victoria Lemieux, Index-aware Knowledge Grounding of Retrieval-Augmented Generation in Conversational Search for Archival Diplomatics. In Proceedings of the Workshop on Computational Archival Science in Big Data, IEEE Big Data Conference, Washington, DC, December 9, 2025 Washington, DC: IEEE, 2025

Acknowledgements

Researchers:

- Roberto Garcia, U. Lleida
- Rosa Gil, U. Llieda
- Luis de la Torre, UNED
- David Wang, UBC
- Tim Weingartner, HSLU
- JaeSeung Song, Sejong U.
- Jong Hyouk Lee, Sejong U.
- Thomas Komenda, deltaDAO
- Esther Olembe, Cameroon
- Jacob Taylor, NIEPI
- Robin Billy, Indp.
- Ethan Clark, Indp.
- Satoru Nakamora, U. Tokyo
- Gabriella Laatikainen, VTT Labs

Students:

- Meng Kang, UBC
- Depansha Chhabra, UBC
- Chris Zhou, UBC
- Philip Chung, UBC
- Brandon Li, UBC
- Faith Molosiwa, UBC
- Jihoo Lee, Sejong U.
- Jaehyung Jeong, Sejong U.
- Soonhong Kwon, Sejong U.
- Wooyoung Sun, Sejong U.
- Nyam Yu Gang, Sejong U.
- Elena Chuang, UBC
- Jonathan Sawa, FNU
- Kristin Wilson, UBC

Funders:

- NSERC
- Mitacs
- SSHRC/InterPARES Trust AI
- CIFAR
- Zug Institute for Blockchain Research


THE UNIVERSITY OF BRITISH COLUMBIA