FUTURE OF ARCHIVES AND LIBRARIES:
HOW TECHNOLOGY AND AI ARE (RE)SHAPING HERITAGE INSTITUTIONS

EXPLAINABLE AI: IMPLICATIONS FOR LIBRARIES AND ARCHIVES

SEPTEMBER 5, 2024
LIBRARY AND ARCHIVES CANADA

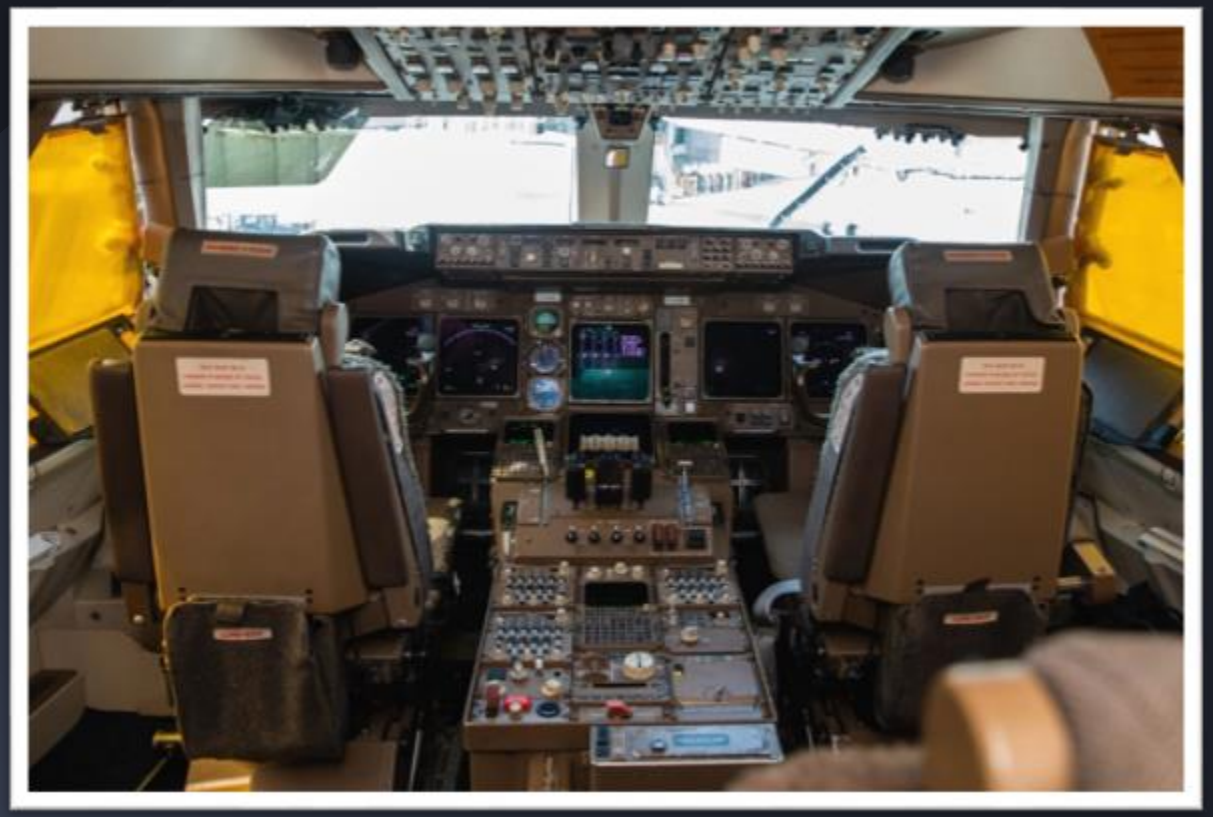
DR. MICHAEL RIDLEY

LIBRARIAN EMERITUS, UNIVERSITY OF GUELPH

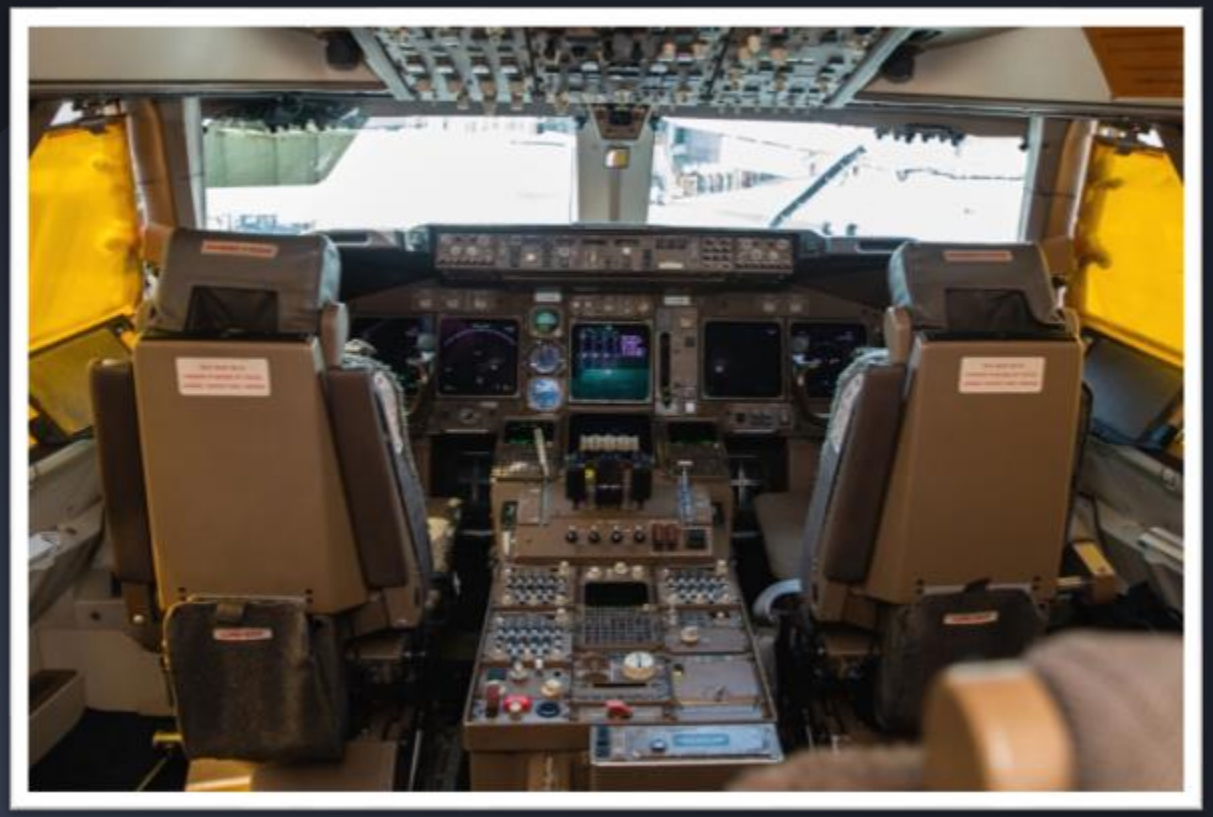
MRIDLEY@UOGUELPH.CA – WWW.MICHAELRIDLEY.CA

If artificial intelligence is so smart, why doesn't it explain itself?

Activists for Explainability



Q:
What is the best flight crew?



A:

A computer, a pilot and a dog.

TRUST & ACCOUNTABILITY

TRUSTWORTHINESS & ACCOUNTABILITY

EXPLAINABILITY

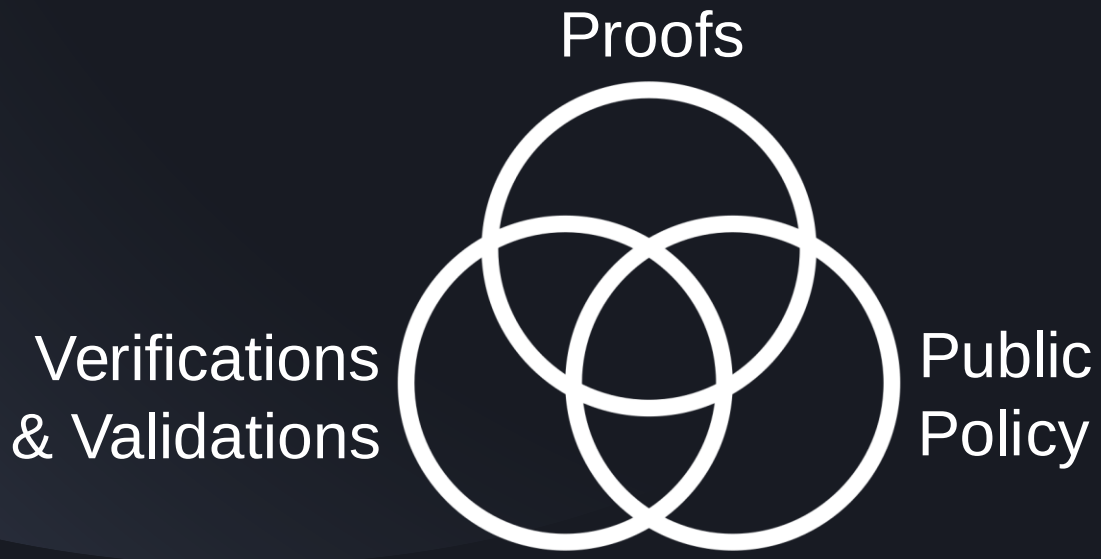
fosters

TRUSTWORTHINESS & ACCOUNTABILITY

Explainable AI (XAI)

XAI is “concerned with developing approaches to explain and make artificial systems understandable to human stakeholders”

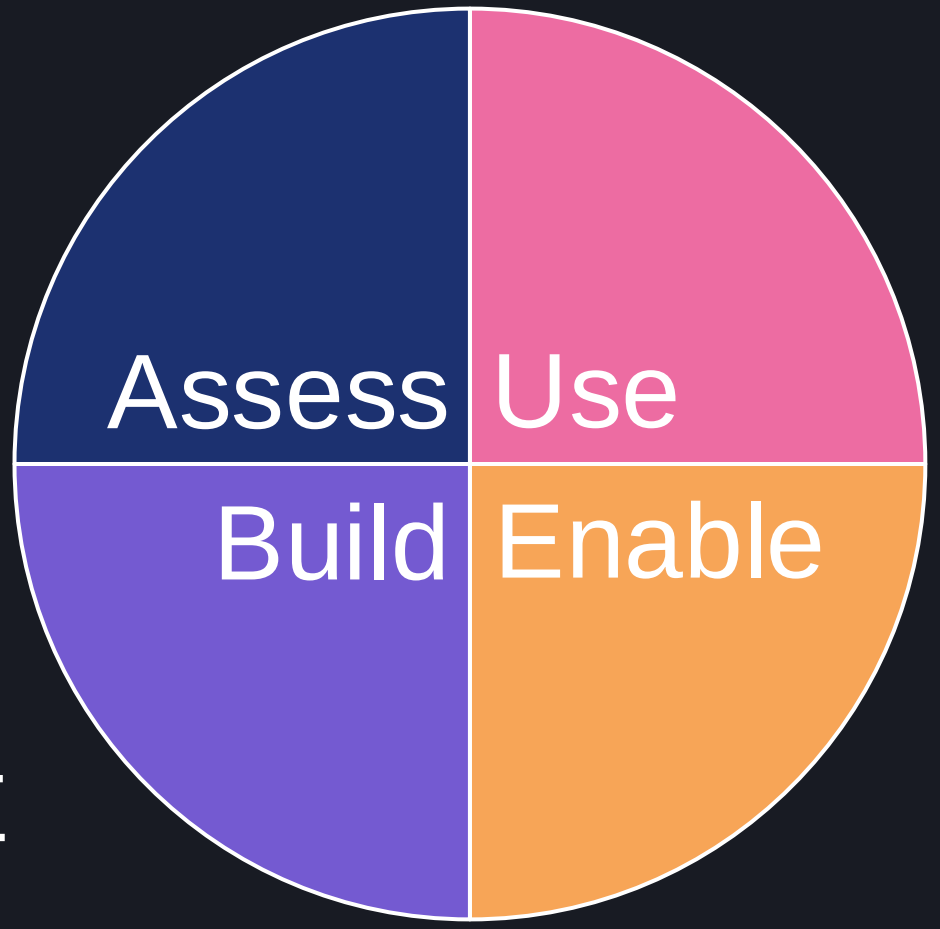
(Langer et al., 2021).



ARTIFICIAL INTELLIGENCE

Ubiquitous
Powerful
Opaque
Invisible
Consequential

ARTIFICIAL INTELLIGENCE



“What happens to libraries and librarians
when machines can read all the books?”

Feral Librarian. March 16, 2017

“I think we would be wise
to start thinking now
about machines and algorithms
as a new kind of patron.”

Chris Bourg
Director of Libraries, MIT

GENERATIVE AI is…

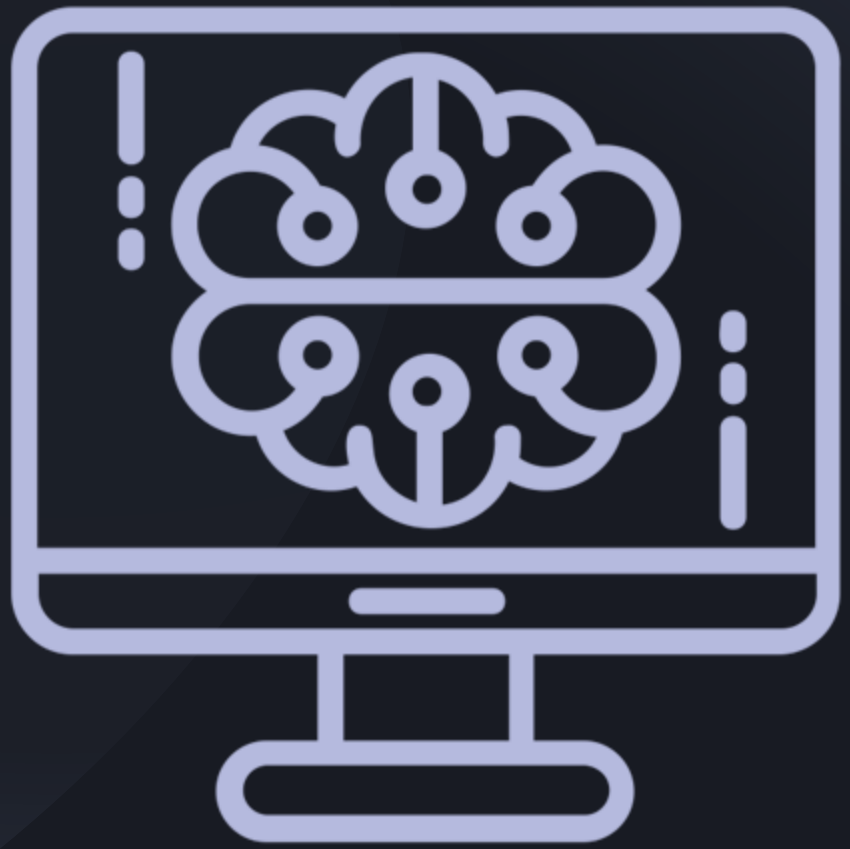
…amazing
…disturbing
…a proof of concept
…seriously flawed
…just a tool
…much more than a tool
…an opportunity
…a call to action

GENERATIVE AI is… like Taylor Swift

AI is Still Evolving

GOFAI

Generative AI



Neurosymbolic AI?

ARTIFICIAL INTELLIGENCE

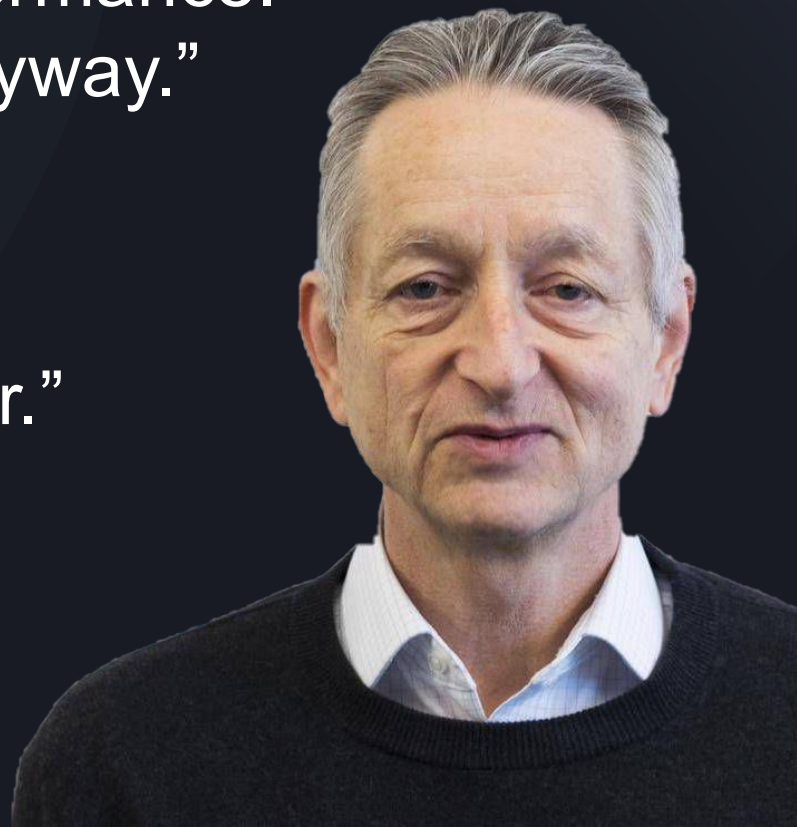
“Assume this is the worst AI you will ever use.”

Ethan Mollick
Co-Intelligence (2024)

Why Isn’t Explainability the Default?

“Machine learning is complex, people won’t understand.”
“We don’t really know what’s happening ourselves.”
“If we had to explain it would hurt performance.”
“People don’t want explanations anyway.”

Besides, Geoff Hinton said
it would be “a complete disaster.”



Emeritus Professor
University of Toronto
“Godfather of AI”

Geoff Hinton is wrong.

Explanations
"are more than a human preoccupation – they are central to our sense of understanding, and the currency in which we exchange beliefs."

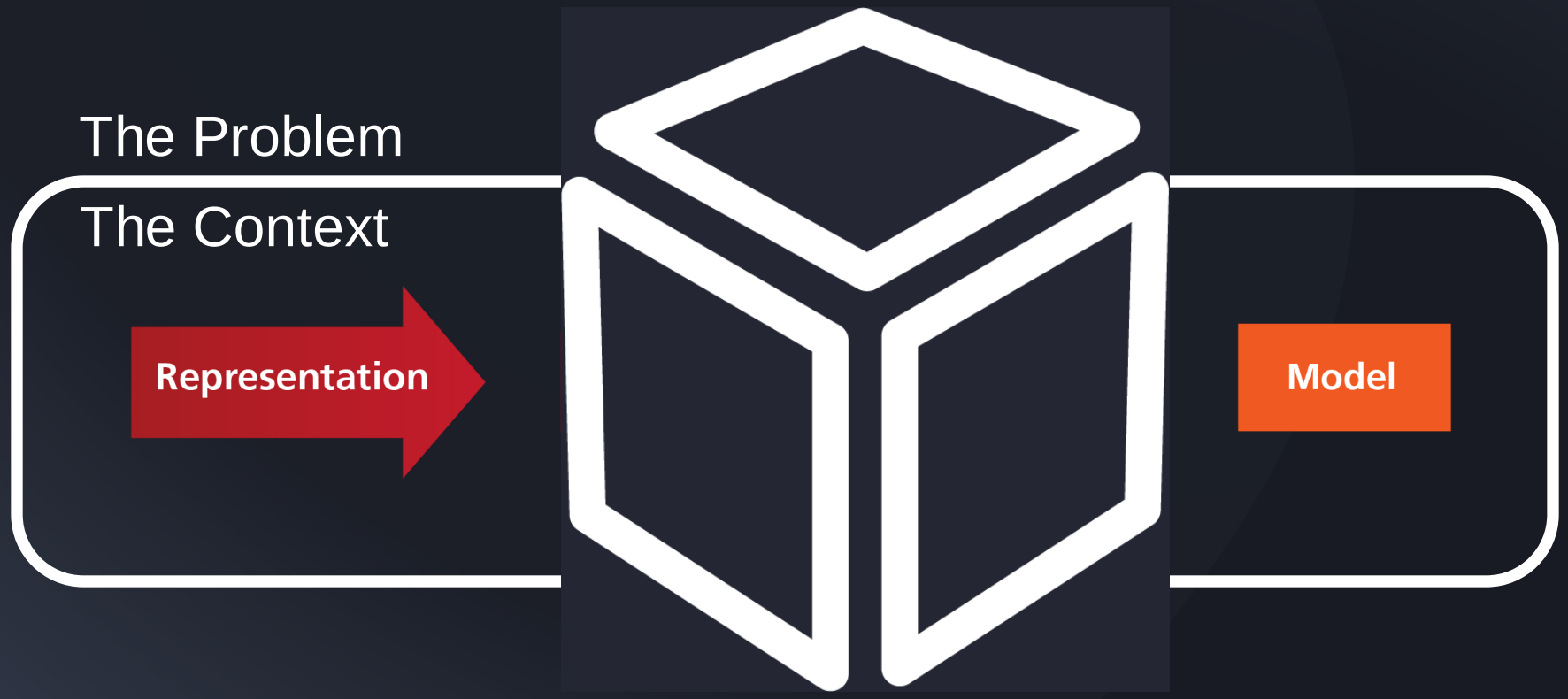
Tania Lombrozo
Director, Concepts & Cognition Lab
Princeton University

A Machine Learning System Framework

Explainable AI (XAI)

XAI is focused on “opening the black box.”

Human Centered Explainable AI



"Not everything that is important lies inside the black box of AI.
Critical insights can lie outside it. Why?
Because that's where the humans are."

Ehsan & Riedl, 2023

“If the machine is involved, then people are too.”

Taina Bucher, 2018

AI is “deeply, inescapably human.”

Jenna Burrell & Marion Fourcade, 2021

“The black box is full of people.”

Nick Seaver, 2021

What XAI & HCXAI Tell Us

XAI
"Inside the Black Box"

HCXAI
"Where the Humans Are"

Implies

Responses:
Technology & Policy

What XAI & HCXAI Tell Us

XAI

“Inside the Black Box”

Features
Examples
Approximations
Counterfactuals

What XAI & HCXAI Tell Us

HCXAI

“Where the Humans Are”

Context (who, when & why)
Actionability & Contestability
Performance Metrics
Pitfalls & Seams
Design Objectives

What XAI & HCXAI Tell Us

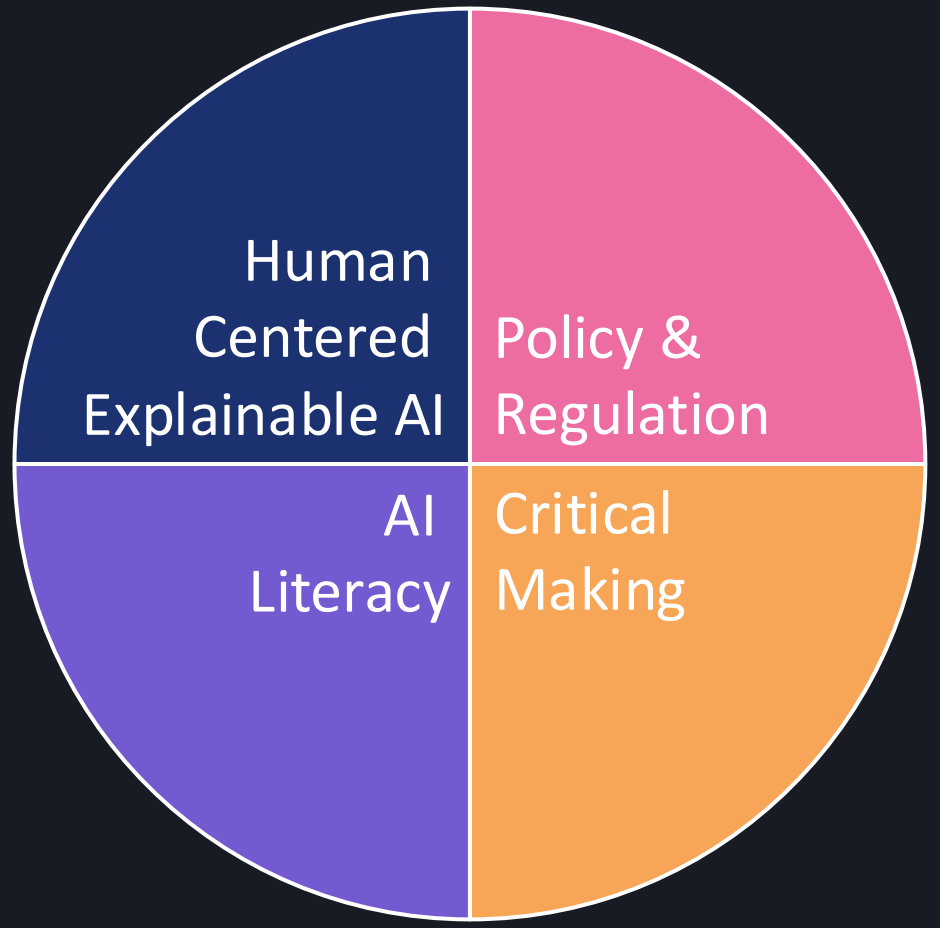
XAI
"Inside the Black Box"

HCXAI
"Where the Humans Are"

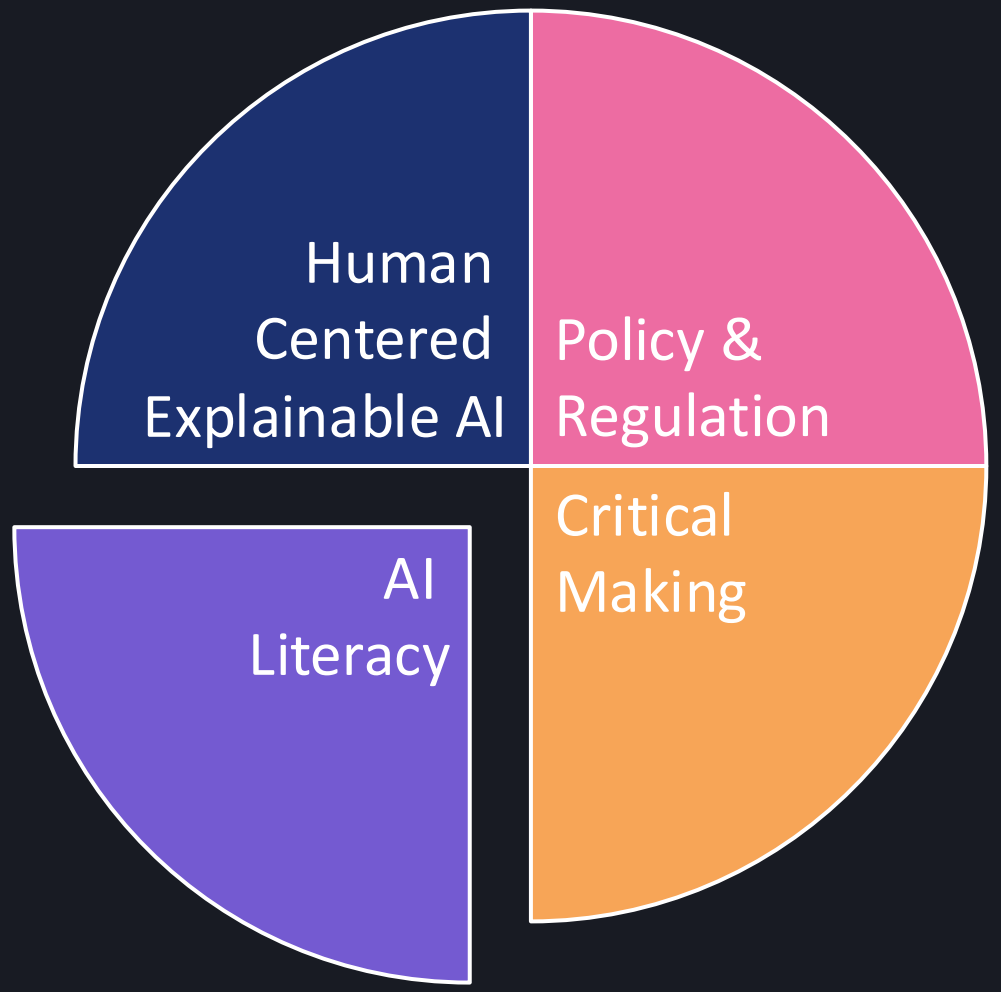
Implies

Interactivity
Personalization
Audit & Oversight
Explanatory Systems

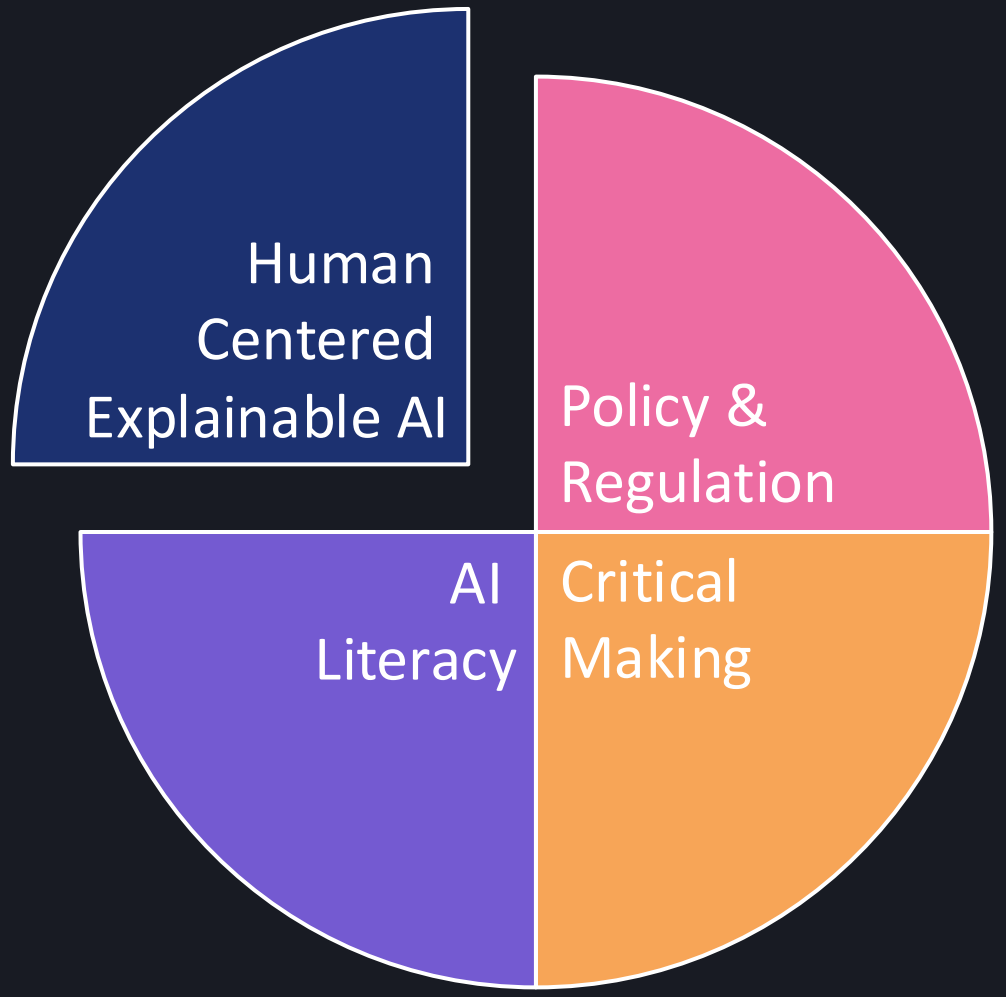
Explainability Activists



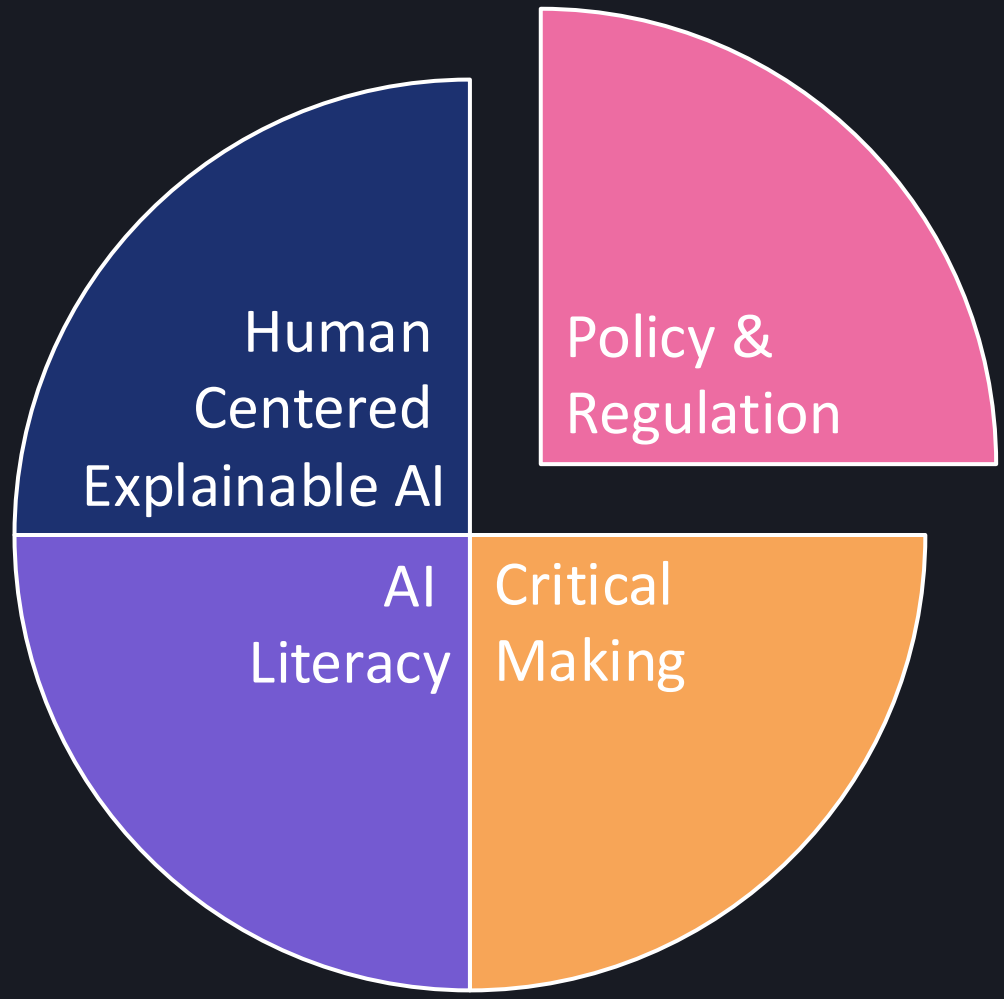
Explainability Activists



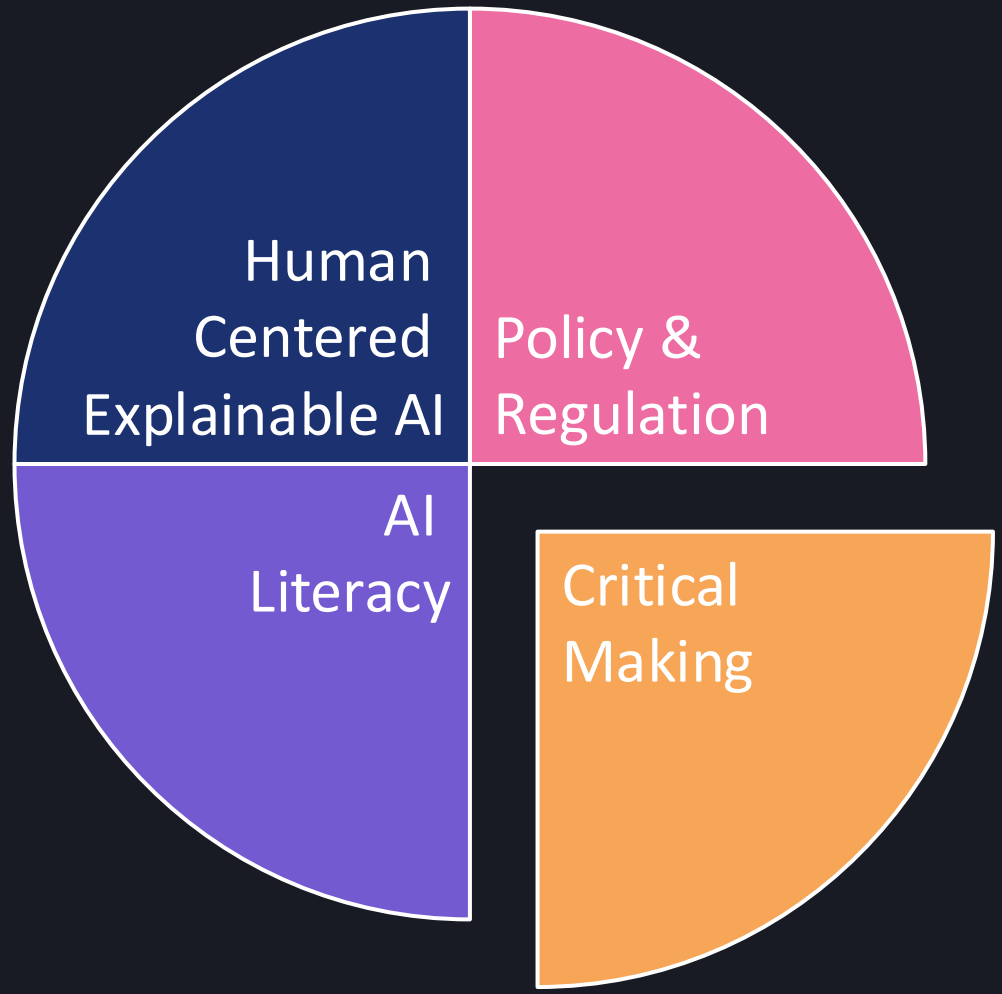
Explainability Activists



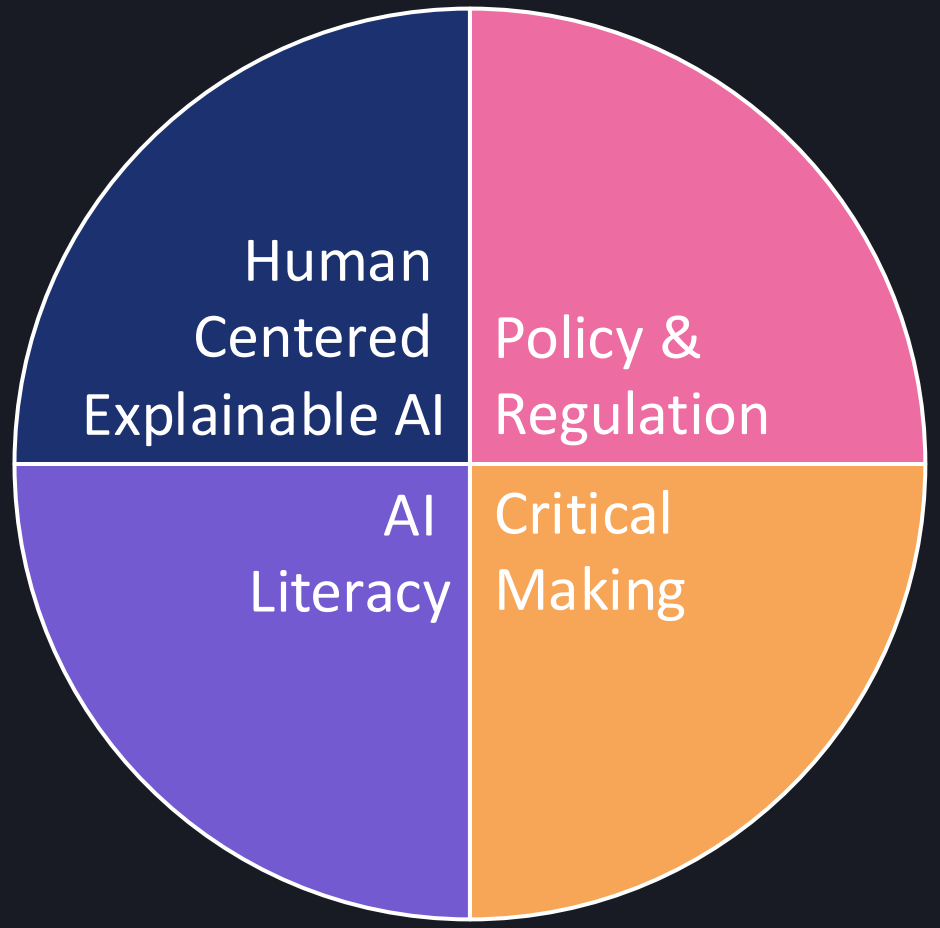
Explainability Activists



Explainability Activists



Explainability Activists



FUTURE OF ARCHIVES AND LIBRARIES:
HOW TECHNOLOGY AND AI ARE (RE)SHAPING HERITAGE INSTITUTIONS

EXPLAINABLE AI: IMPLICATIONS FOR LIBRARIES AND ARCHIVES

SEPTEMBER 5, 2024
LIBRARY AND ARCHIVES CANADA

DR. MICHAEL RIDLEY

LIBRARIAN EMERITUS, UNIVERSITY OF GUELPH

MRIDLEY@UOGUELPH.CA – WWW.MICHAELRIDLEY.CA