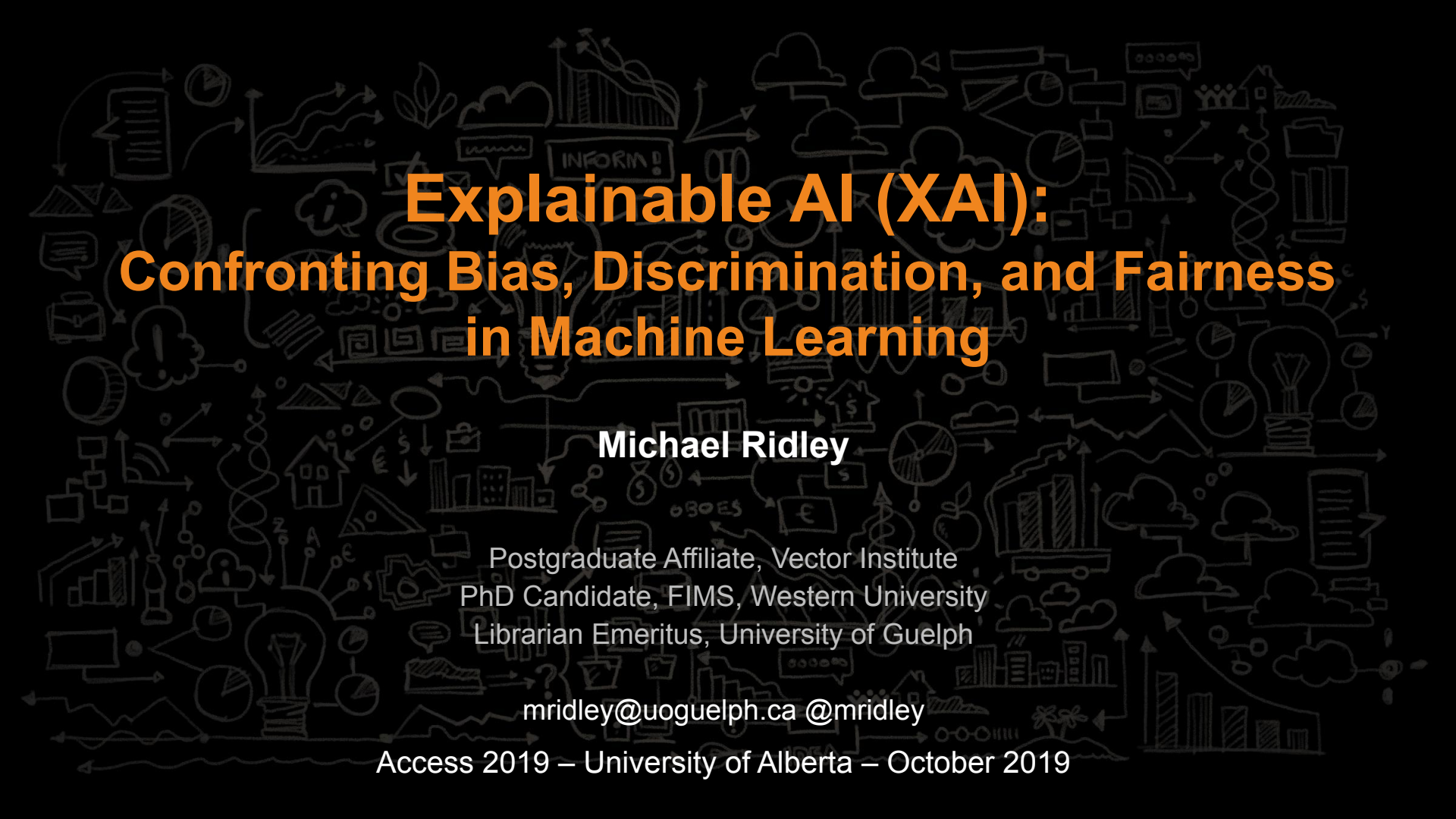


Explainable AI (XAI): Confronting Bias, Discrimination, and Fairness in Machine Learning

Michael Ridley

Postgraduate Affiliate, Vector Institute
PhD Candidate, FIMS, Western University
Librarian Emeritus, University of Guelph

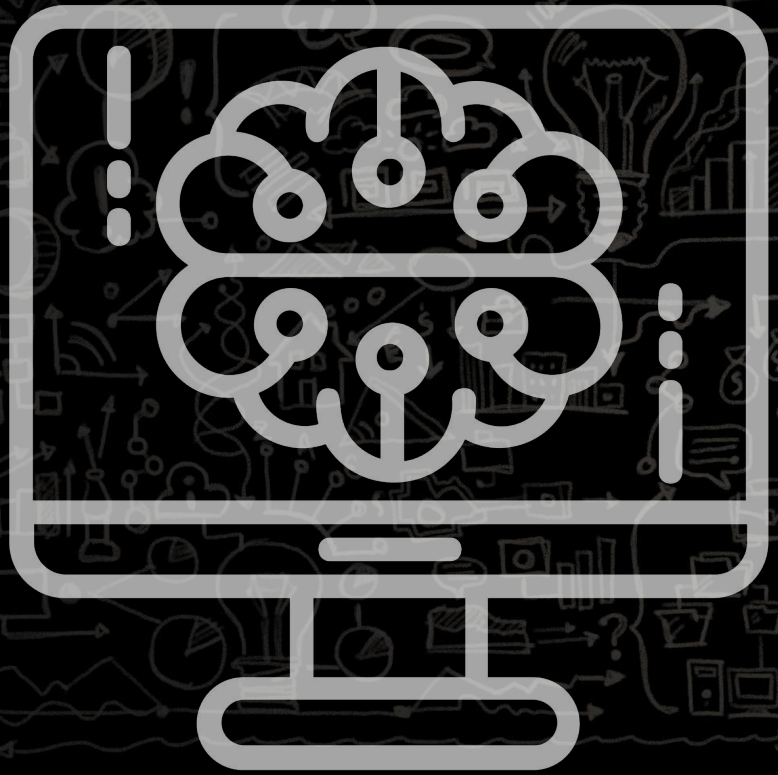
mridley@uoguelph.ca @mridley

Access 2019 – University of Alberta – October 2019

“I believe that artificial intelligence will become a major human rights issue in the twenty-first century.”

Safiya Noble,
Algorithms of Oppression (2018)

Machine Learning



Ubiquitous

Powerful

Opaque

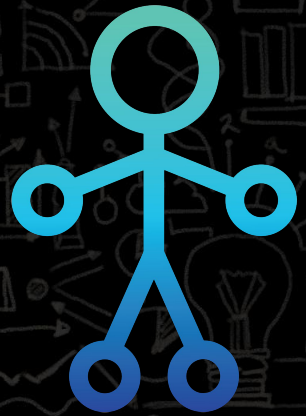
Invisible

Consequential

"The danger is not so much
in delegating cognitive tasks,

but in distancing ourselves from
– or in not knowing about –

the nature and precise mechanisms
of that delegation"



de Mul & van den Berg (2011). Remote control: Human autonomy in the age of computer-mediated agency.

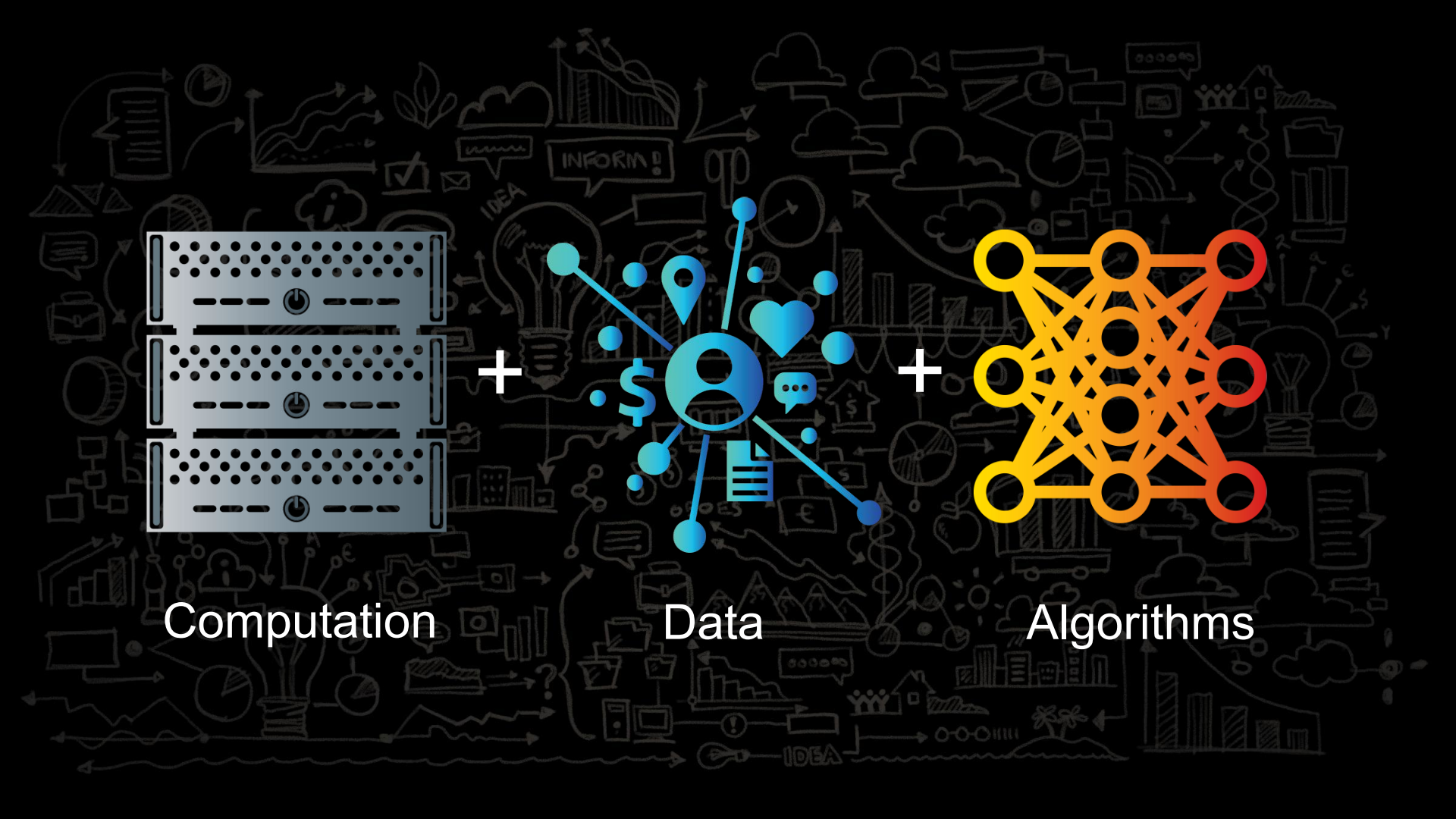
What do people want
in an explanation?

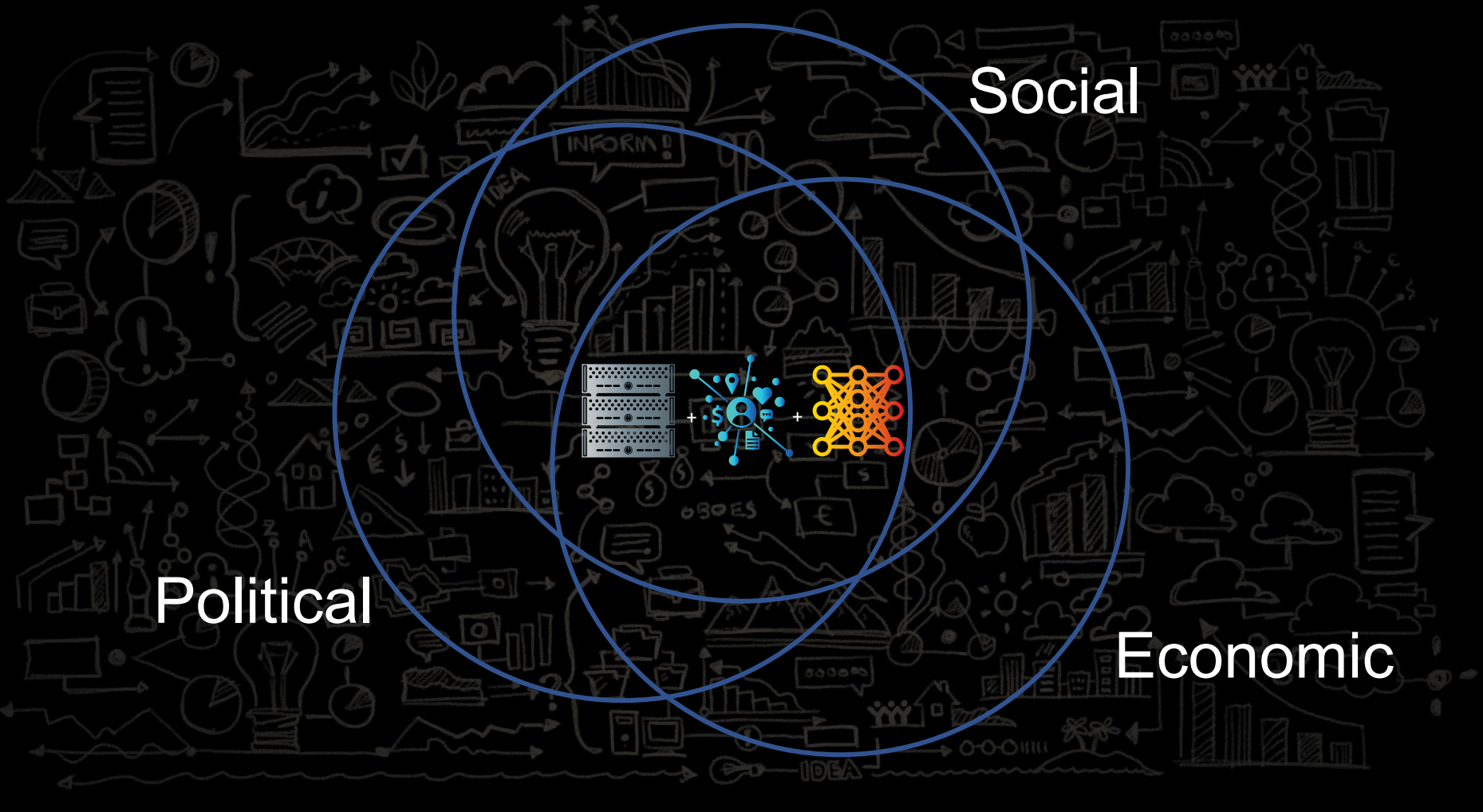
How can AI provide such an explanation?

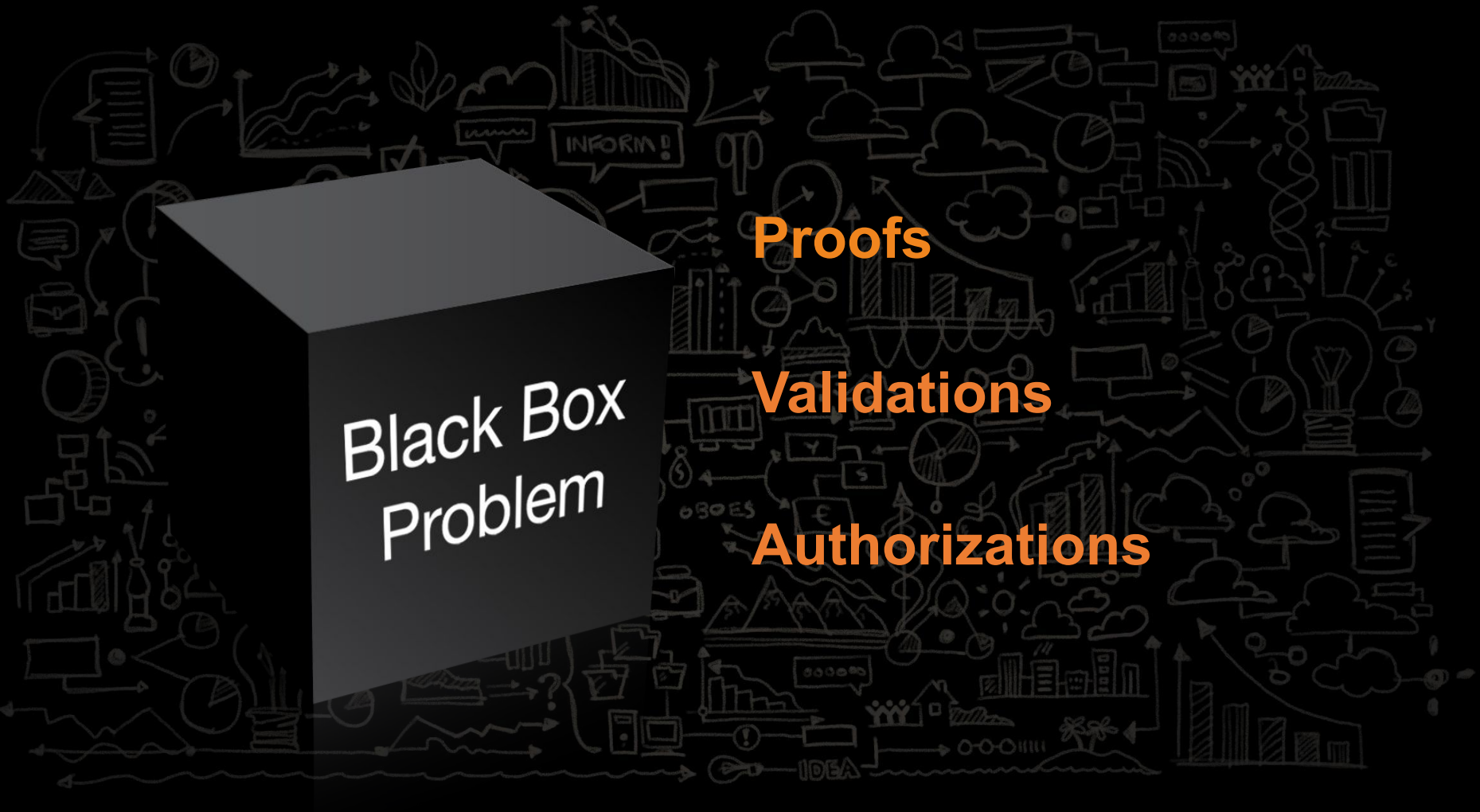

General Data
Protection Regulation
(GDPR)

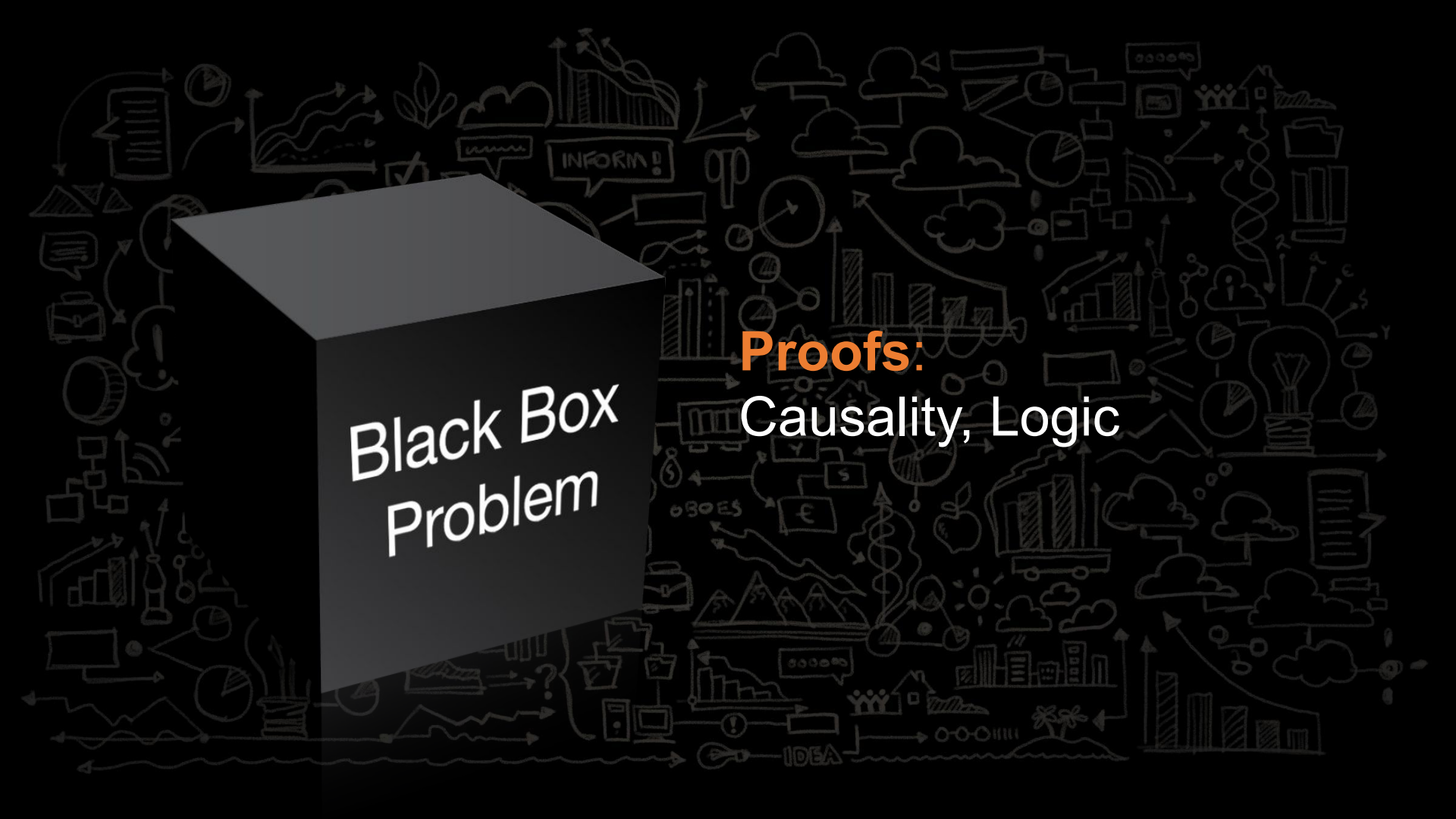

“right to explanation”


Context
Who
Focus
Governance
Regulators
Policy
Application
Users
Trust
Functionality
Developers
Performance

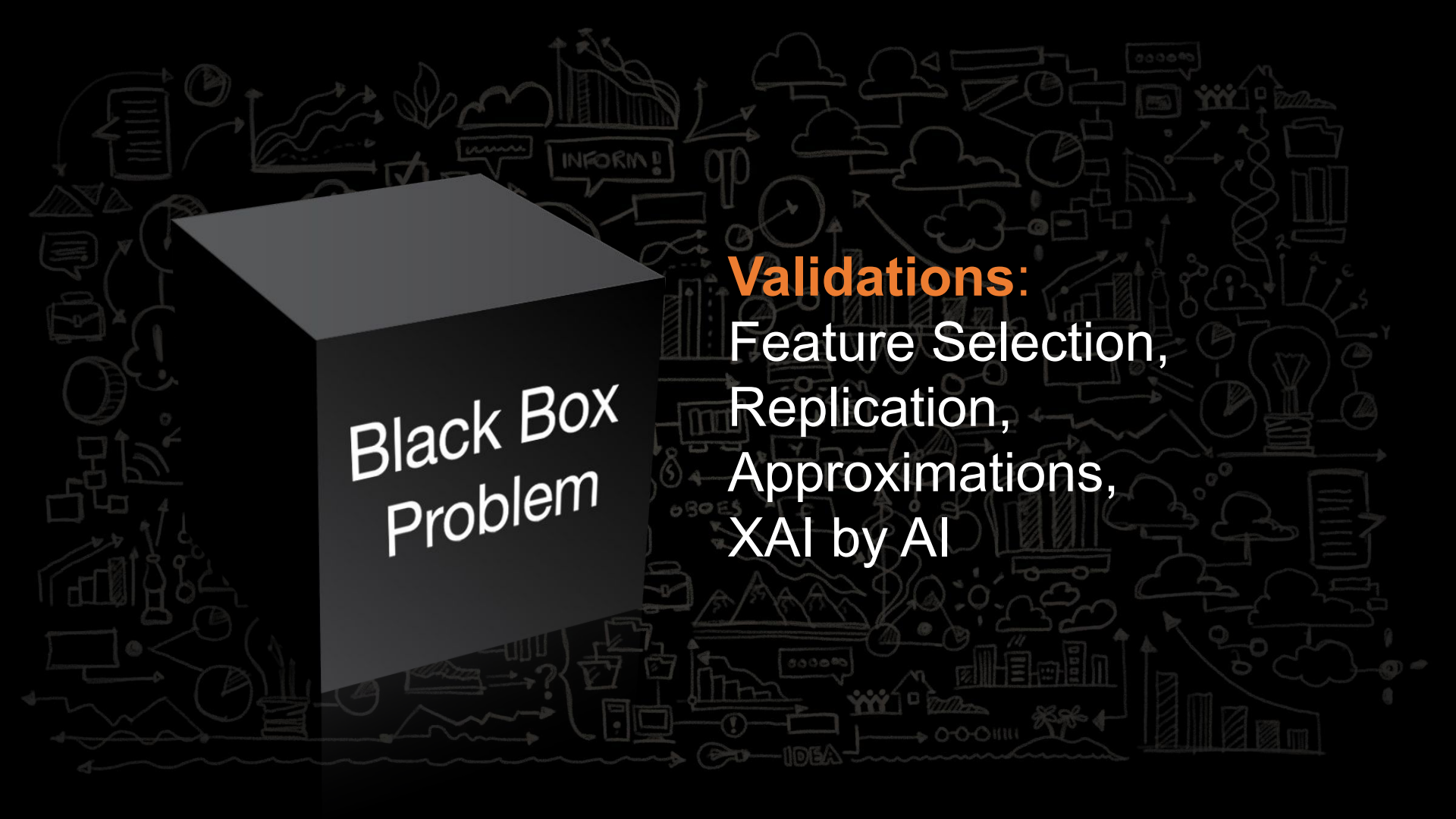

Computation
Data
Algorithms


Social
Political
Economic


Black Box
Problem
Proofs
Validations
Authorizations



Proofs:
Causality, Logic



Validations:
Feature Selection,
Replication,
Approximations,
XAI by AI

Authorizations:
Codes & Standards,
Legislation, Due Process,
Audit, Regulation

Roles for Libraries

Algorithmic Literacy

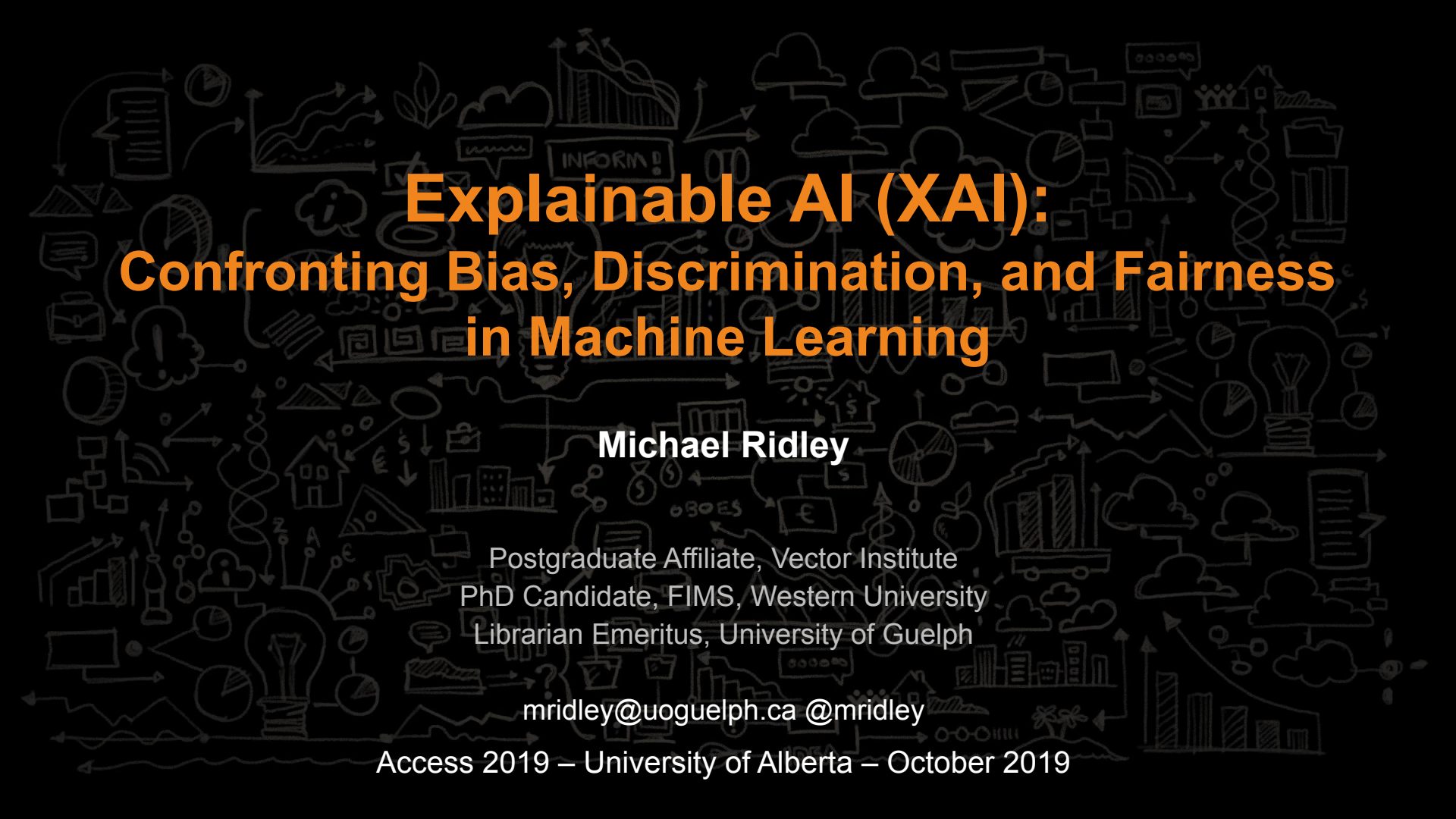
ML Partnerships & Development

Advocacy

The New Digital Divide

"A class of people who can use algorithms and a class used by algorithms"

David Lankes, Director
School of Library and Information Science
University of South Carolina



Explainable AI (XAI): Confronting Bias, Discrimination, and Fairness in Machine Learning

Michael Ridley

Postgraduate Affiliate, Vector Institute
PhD Candidate, FIMS, Western University
Librarian Emeritus, University of Guelph

mridley@uoguelph.ca @mridley

Access 2019 – University of Alberta – October 2019