

TRUSTWORTHY AI AND CORPORATE DIGITAL RESPONSIBILITY (CDR)

Prof. Karen Elliott
Birmingham University Business School
k.elliott@bham.ac.uk



UNIVERSITY OF
BIRMINGHAM

Contents



Image Credit : Andy Kelly on Unsplash

Permission use ©2024 The Business Analytics Institute (Grenoble, Fr)



Public Controversies

Misuse of data/digital technologies



Socio-tech dimensions

It's all about people – ML/AI risks



CDR framework

*What is CDR? Why should you care?
Whose responsibility?*



Identify outstanding questions

What are we missing?

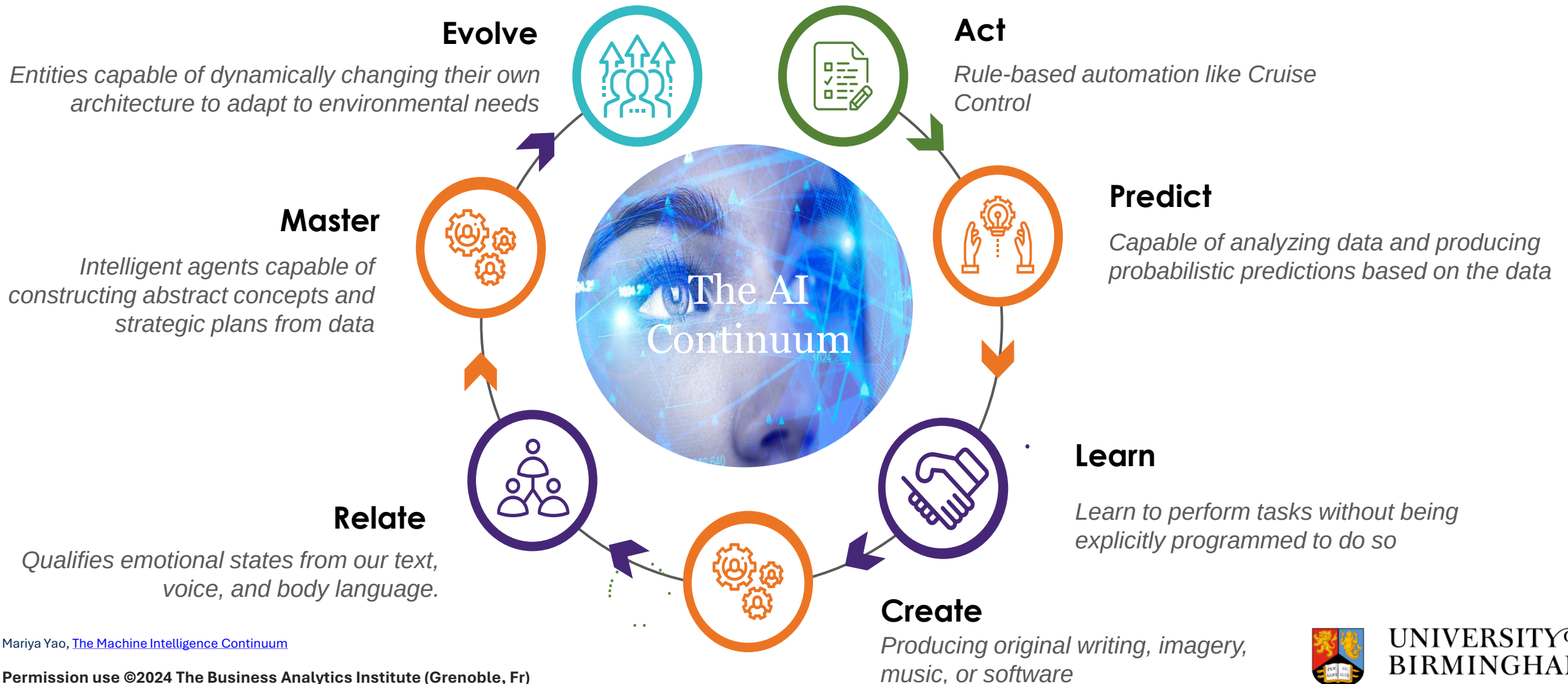


UNIVERSITY OF
BIRMINGHAM

Starts & Ends with PEOPLE



What is AI?



Five Levels of Concern



Personal Privacy

The relationship between the use of personal data and the public expectation of privacy.



Implicit Bias

A bias or prejudice that is present but not consciously held or recognized



Accountability

Assuming responsibility for decisions, actions, and project outcomes



Human Agency

The ability to shape one's life choices



Innovation

The process of creating value by applying novel solutions to meaningful problems

Personal Privacy



Image Credit : Jesse K.

Rahnama and Pentland, [The New Rules of Data Privacy](#), HBR

Permission use ©2024 The Business Analytics Institute (Grenoble, Fr)



Data is eating the World

Digital Curtains - Data protection vs Personal Privacy.



Surveillance Capitalism

Zuboff - Those that collect and control these “transactions” control us



The Asymmetry of Information

The commercial use of the private and sensitive data – can we leverage insights without transferring data?



A Digital Bill of Rights

GDPR, The Digital Services Act, The Digital Markets Act. The right to know, to modify, to delete, to be forgotten



UNIVERSITY OF
BIRMINGHAM

Implicit Bias



Data Bias

Sample, prejudicial, and measurement bias



Algorithmic Bias

Bias and variance are interdependent algorithmic properties that influence prediction



Business Logic

Coding the real-world business rules that determine how data can be created, aggregated, and modified



Ethics

Frameworks for determining governance, fair practices, and corporate digital responsibility



Image Credit : Oladimeji Odunsi

Managerial Accountability (CDR)



Image Credit Roland Samuel

Sanford S., [How to Build Accountability into Your AI](#), HBR

Permission use ©2024 The Business Analytics Institute (Grenoble, Fr)



Profiling

Extrapolating information about a person based on known traits or tendencies



The Black Box

The challenges of interpretability, fairness, and robustness



Automated Decision-Making

Assess governance structures, understand the data, define performance metrics, review monitoring plans



The AI Life Cycle

Design, development, deployment, monitoring



UNIVERSITY OF
BIRMINGHAM

Human Agency



Freedom of Thought

The freedom to hold or consider a fact, viewpoint, or idea independently



Freedom of Action

The ability to shape one's individual choices



Singularity

The point in time at which technological growth becomes uncontrollable and irreversible



HI vs AI

How can human and machine intelligence work best together?



Image Credit : Antoine Beauvillain

Organizational Innovation

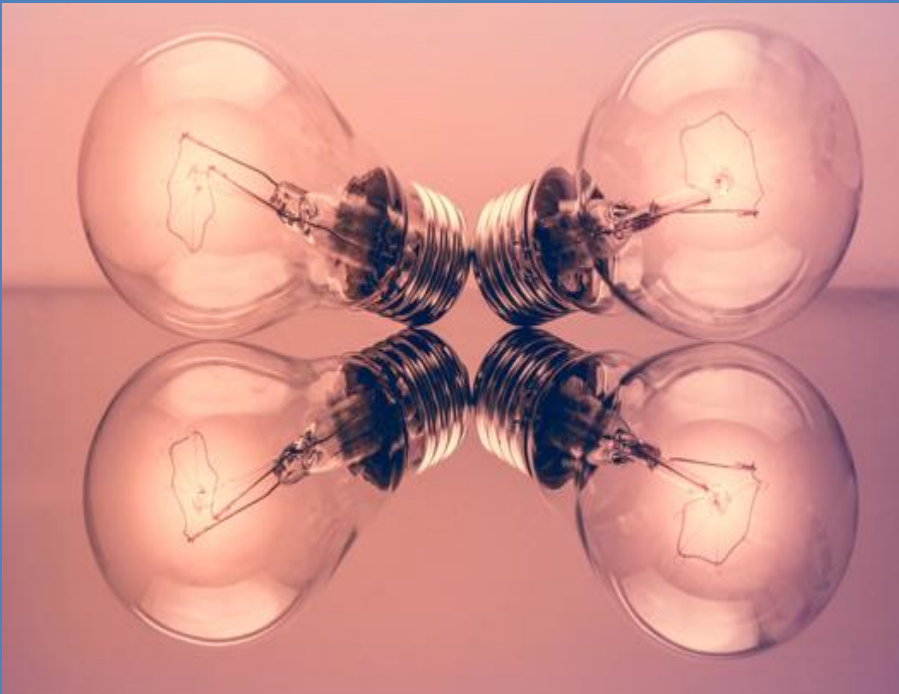


Image Credit : Dragos Gontariu on Unsplash

Davis, N. , *Technology and the power of narratives in times of uncertainty*, Unicef

Permission use ©2024 The Business Analytics Institute (Grenoble, Fr)



Invention vs Innovation

the Obscure Features Hypothesis - “Novel and without precedent” vs “change that creates value”



Deep Learning

Imitating how the human brain processes information and takes decisions



Processes and Paradigms

“Changing the way you address customer challenges” vs “changing the way you think about business”



Narrative Economics

How can we maximize the value of digital innovation to benefit individuals, societies, and environments by leveraging value and governance?



UNIVERSITY OF
BIRMINGHAM

Building socio-technical trust



Computer Science concept

“accepted dependence[1]”, Trusted Platform (TPM), etc.



ML/DL

cyber security experts: secure and privacy preserving // activists: ethical // machine learning experts: accurate and efficient



Concepts of trust

Different terminology for trust-related concepts: ethical vs. trustworthy machine learning



Reality

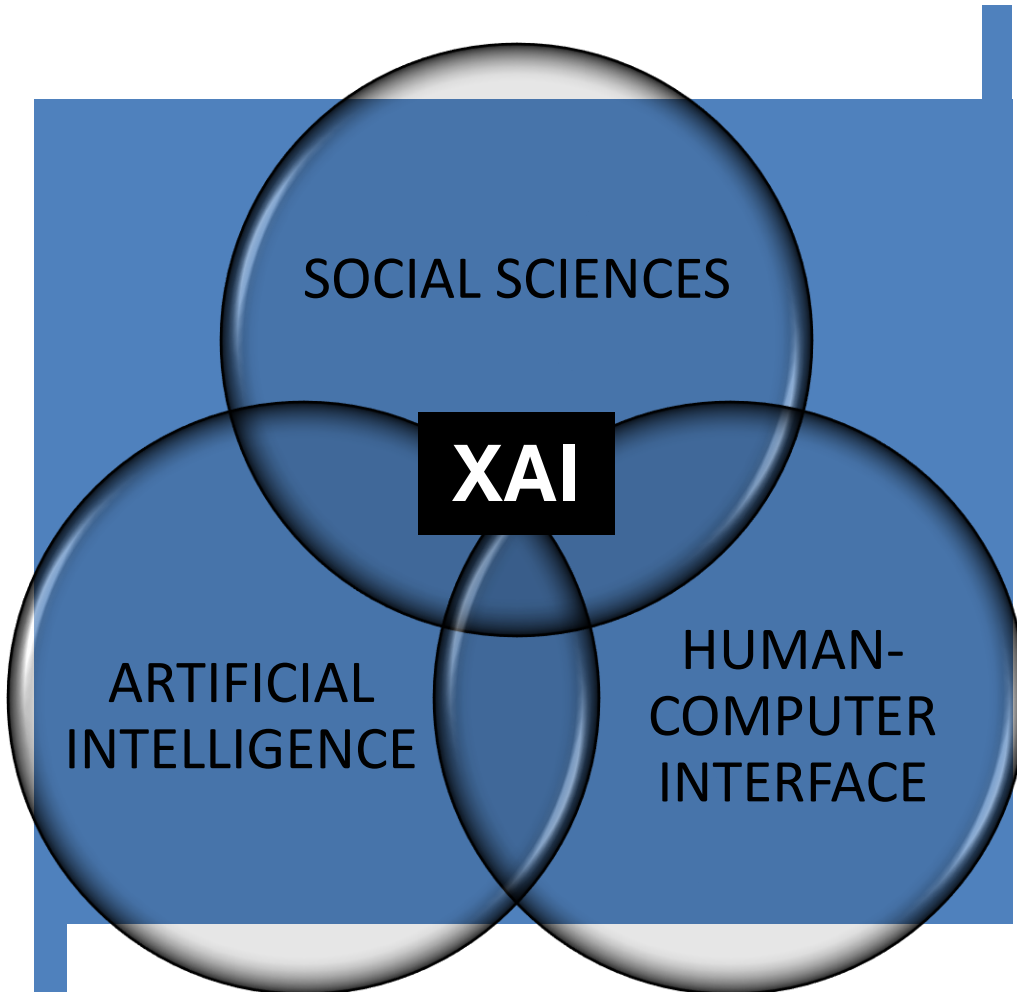
Everyone has their own interpretation and perception of trust whether technical or social in nature.

(Algirdas, et al, 2004)^[1]



UNIVERSITY OF
BIRMINGHAM

Explainable AI (XAI)



Explainability

should also ensure **Legibility** – technical explanations increase transparency only for those with technical expertise.



Transparency

Informational transparency: providing information about technical functions;
Participatory transparency: providing opportunities for public scrutiny or engagement;
Accountability transparency: providing mechanisms for the public to hold decision-makers to account.

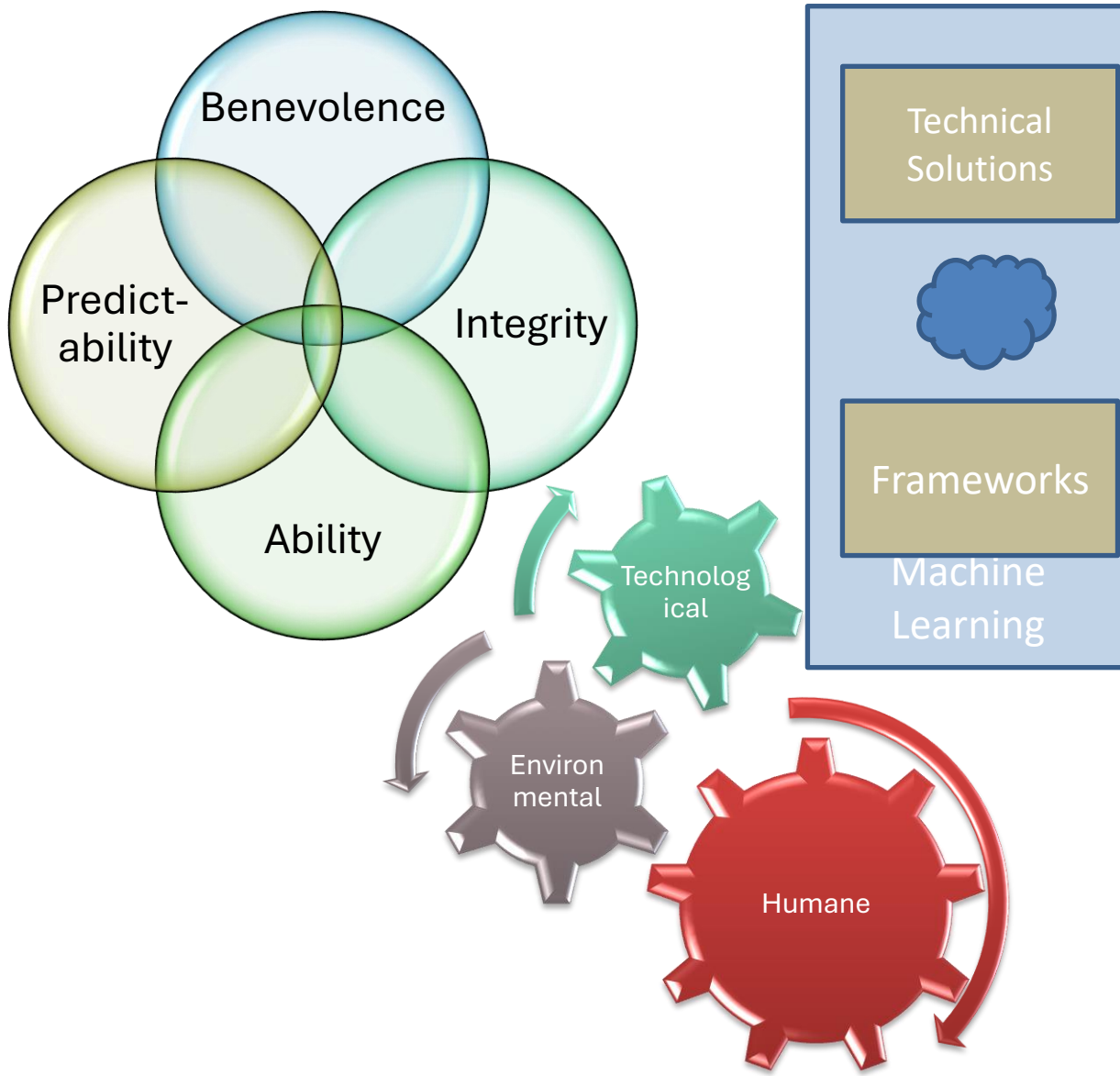


Reality

We move beyond purely technical approaches to XAI.



Trustworthiness: in reality



2 Approaches in ML

Principled and Technological solution to meet 'trustworthy' criteria.

Principled Framework High Level

- **Informational transparency:** providing information about technical functions;
- **Participatory transparency:** providing opportunities for public scrutiny or engagement;
- **Accountability transparency:** providing mechanisms for the public.

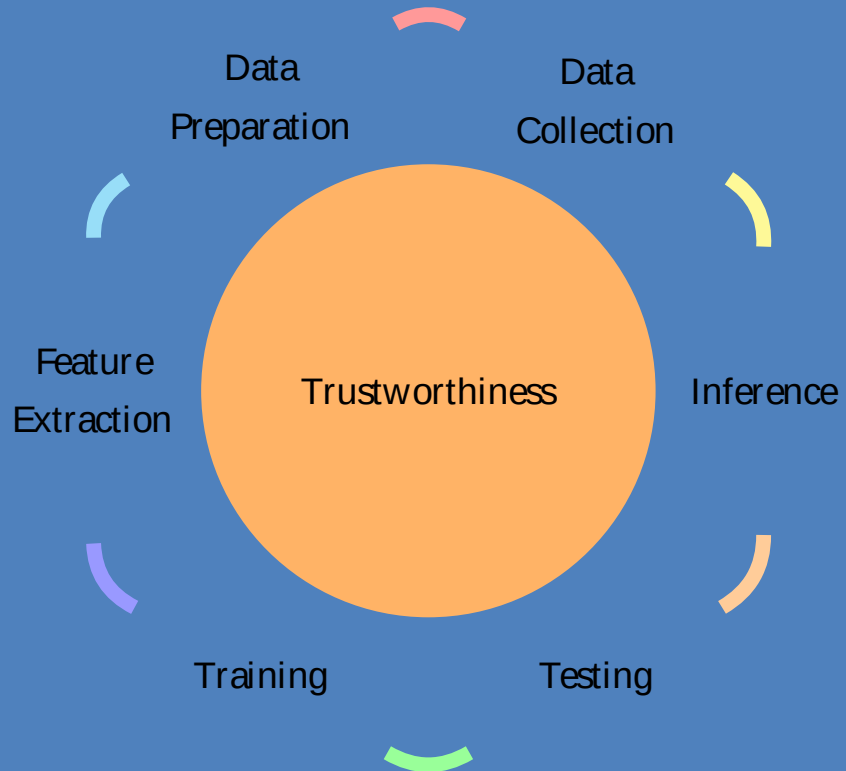
Reality

We move beyond purely technical approaches to XAI.



UNIVERSITY OF
BIRMINGHAM

Chain of Trust



Chain of trust



implementing *Fairness, Explainability, Auditability & Safety (FEAS)* technologies . Trust propagates gradually in the machine learning pipeline.

Two trust focused sections



Data Related Trust Solutions
Model Related Trust Solutions

Benefits in building trust



Different Stages impact on each other
Algorithm iterates through the stages during its lifecycle
Technology decisions in all stages impact others.
Opportunity to respond to accidents, sudden breakdowns of trust or failures effectively.



Recap & Wicked Problems

“FEAS”

Fairness, Explainability, Auditability & Safety
underpins Data governance concerns to generate trustworthy technologies:
Data-related trust solutions
Model-related trust solutions

TECHNICAL TRUST

Data governance surrounds regulatory compliance issues.

For technicians: Accepted dependence Trusted Platform (TPM), etc.

SOCIAL TRUST

In case of machine learning:
secure and privacy preserving

activists:
ethical
machine learning experts:
accurate and efficient
ethical vs. trustworthy
machine learning

FUTURE - CONTAINMENT

Today, AI and GAI are rapidly evolving.

RISK MANAGEMENT SAFETY
HOW TO CONTAIN?
SKILLS GAP?



Why care?



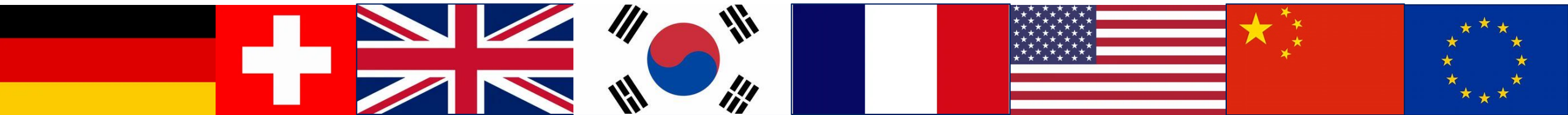
**CAPITALISM TENDS TO DESTROY
ITS TWO SOURCES OF WEALTH:
NATURE AND HUMAN BEINGS**

Karl Marx



What is Corporate Digital Responsibility (CDR)

Corporate Digital Responsibility (CDR) is a set of practices and behaviours that help an organisation use data and digital technologies in ways that are perceived as socially, economically, and environmentally responsible.



Environmental
Sustainability



Digital
Ethics

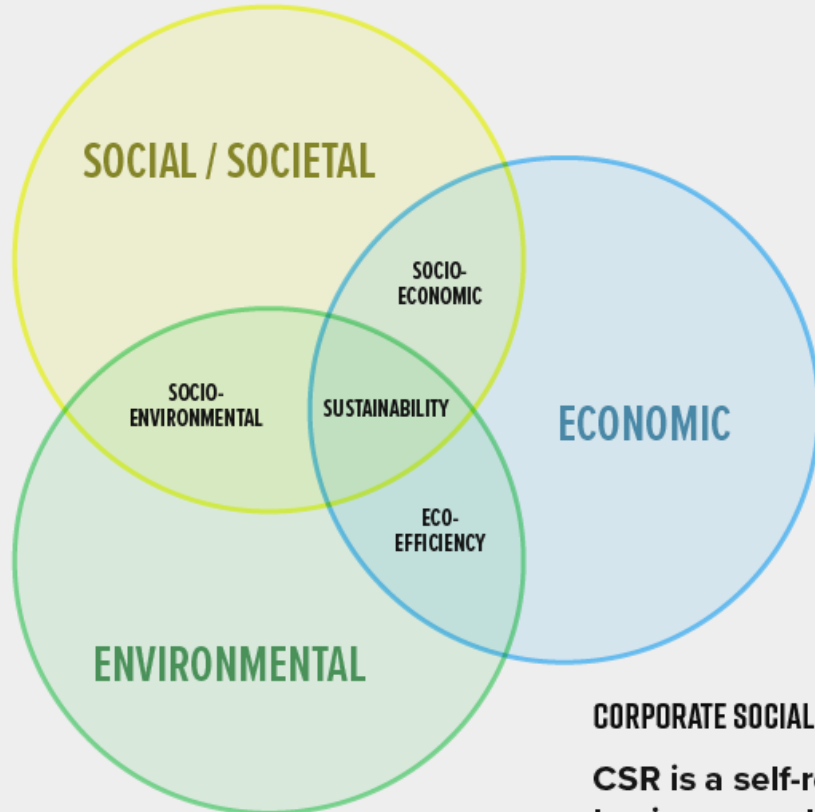


Inclusion,
Accessibility & Diversity



Privacy,
Safety Online

CSR vs CDR



CORPORATE SOCIAL RESPONSIBILITY

CSR is a self-regulating business model that helps a company be socially accountable

PHYSICAL (CSR)

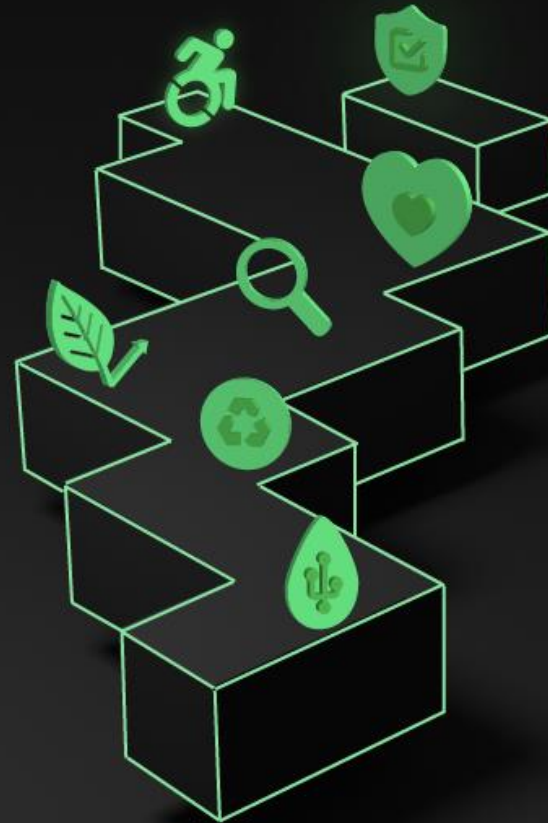


DIGITAL (CDR)



CDR is the digital of the same triangle of Economic, Social (Societal) and Environmental, existing as an additional layer across all aspects of CSR

CDR: 7 core principles/manifesto

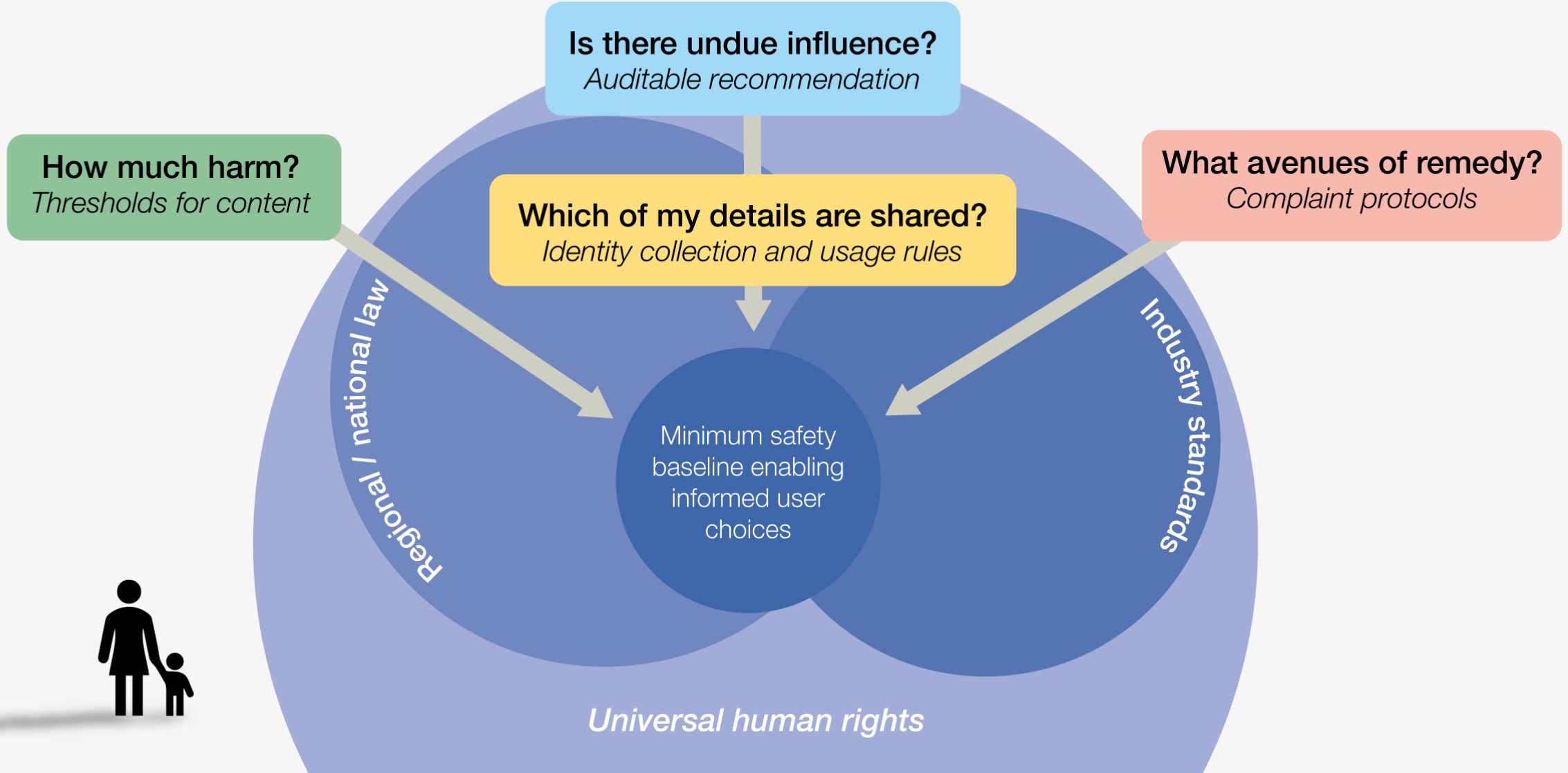


CDR PRINCIPLES

- 1 PURPOSE AND TRUST
- 2 FAIR AND EQUITABLE ACCESS FOR ALL
- 3 PROMOTE SOCIETAL WELLBEING
- 4 CONSIDER ECONOMIC AND SOCIETAL IMPACT
- 5 ACCELERATE PROGRESS WITH IMPACT ECONOMY
- 6 CREATING A SUSTAINABLE PLANET TO LIVE
- 7 REDUCE TECH IMPACT ON CLIMATE AND ENVIRONMENT

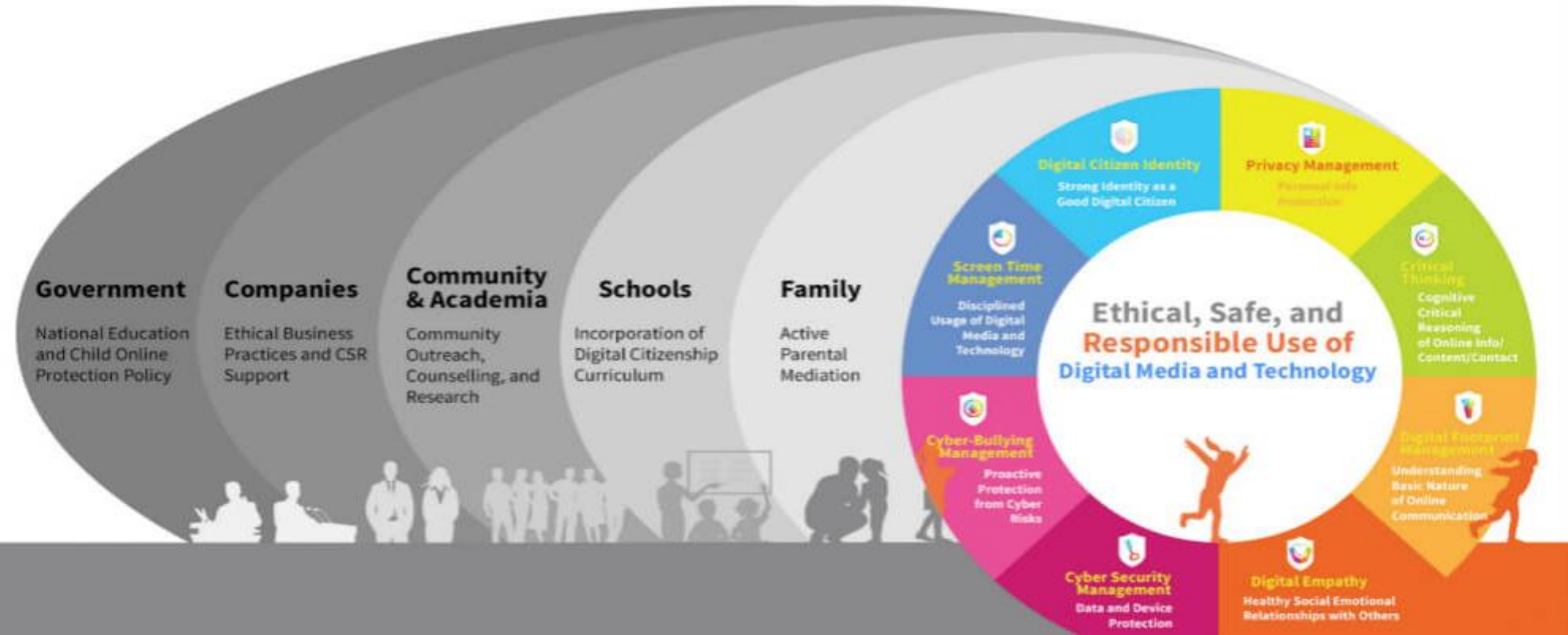


Data/digital = a human right

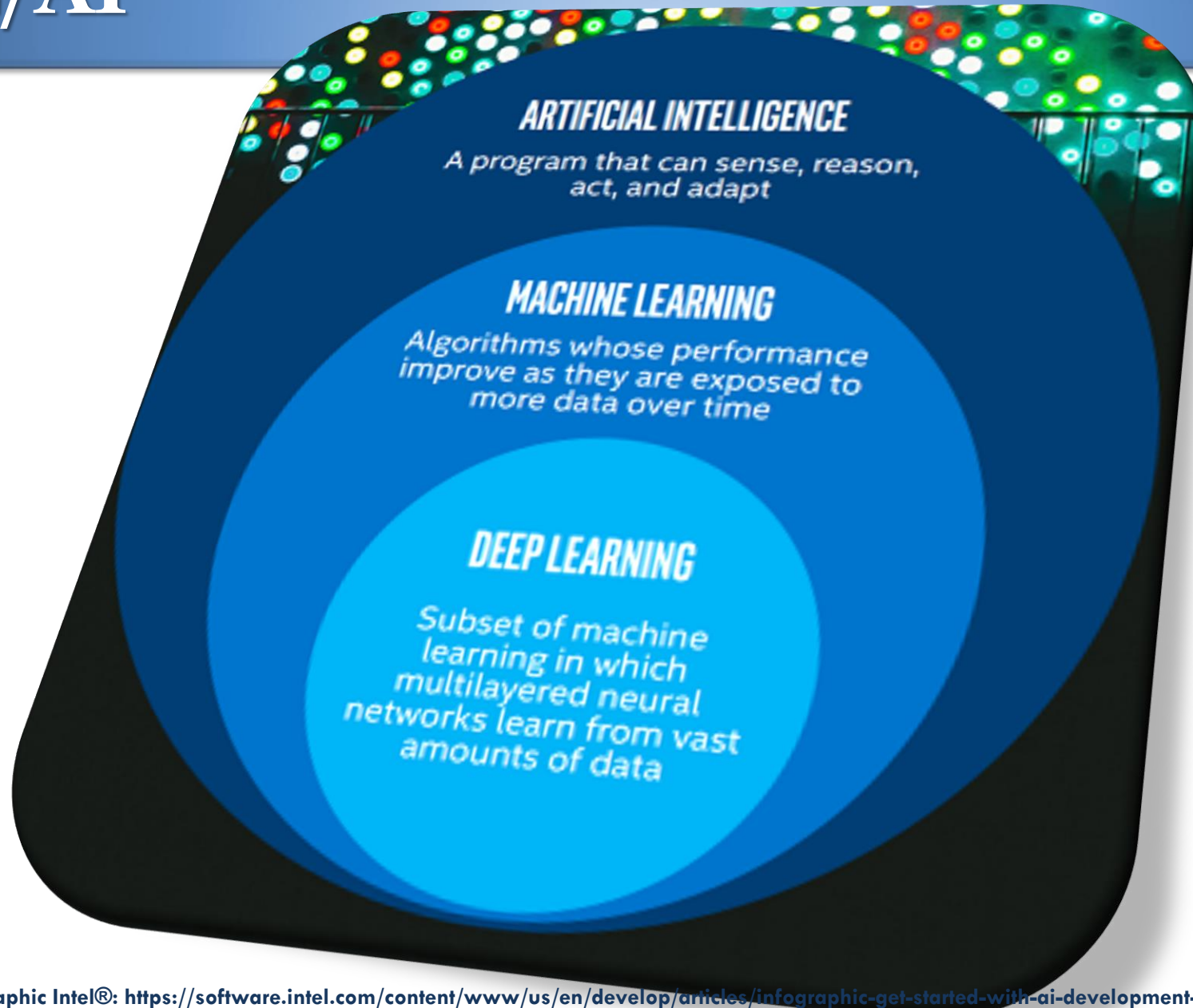


Collective Responsibility (CDR)

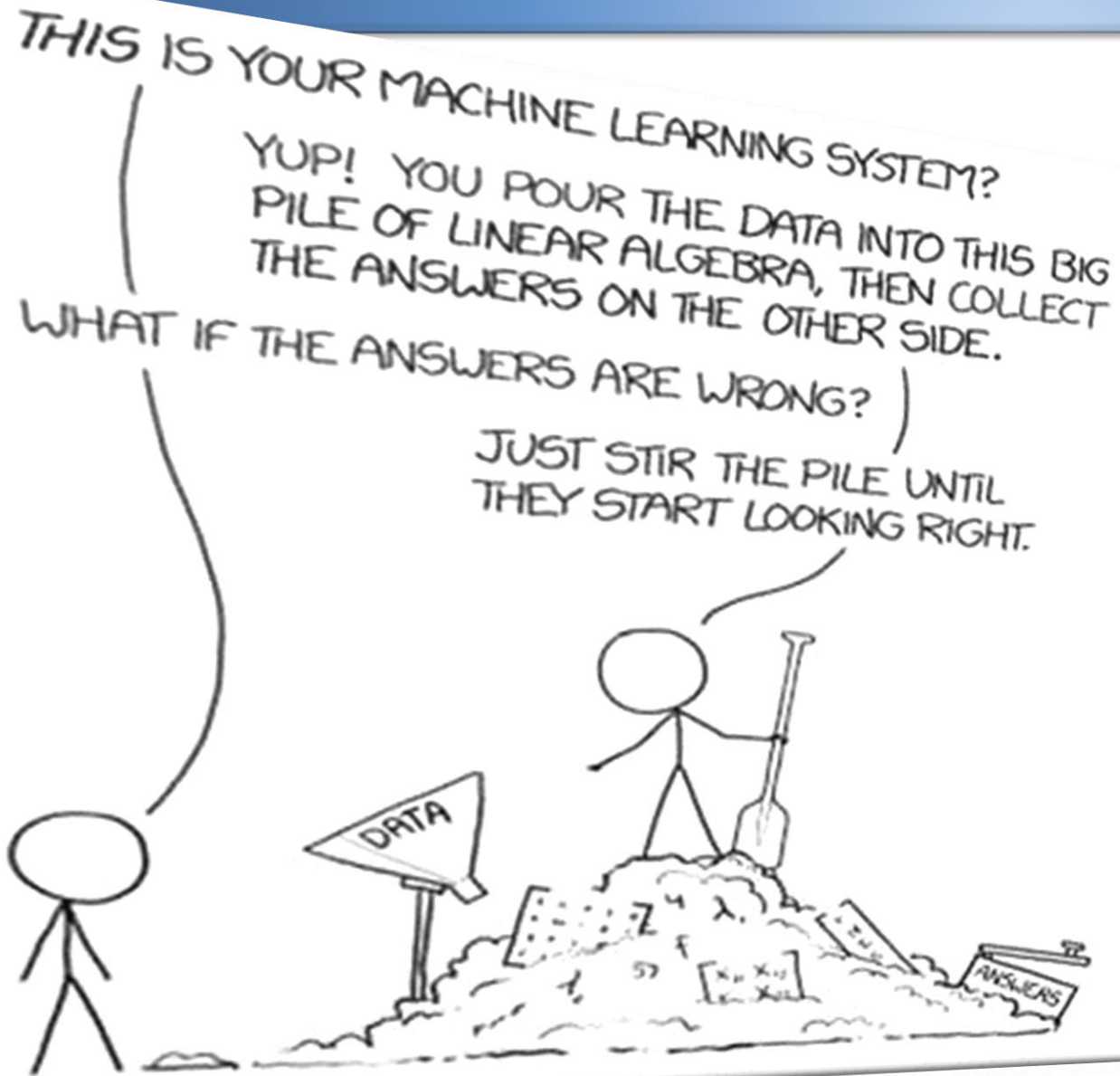
Building an Ethical Digital Ecosystem as a Collective Responsibility



DL/ML/AI



ML – an objective process?



**“TRUST =
CURRENCY OF
INNOVATION”**

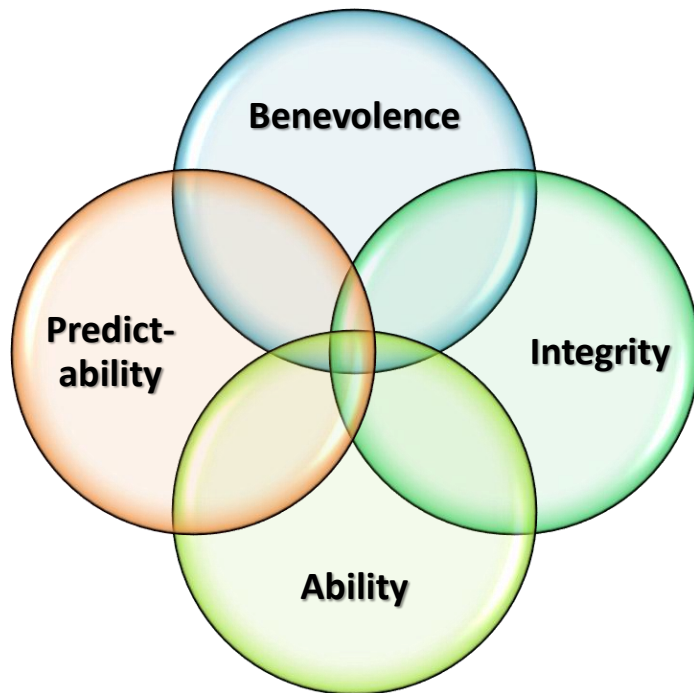
Mastercard/HBS (2022)



UNIVERSITY OF
BIRMINGHAM

Trust & CDR

ABI+ MODEL



TRUST



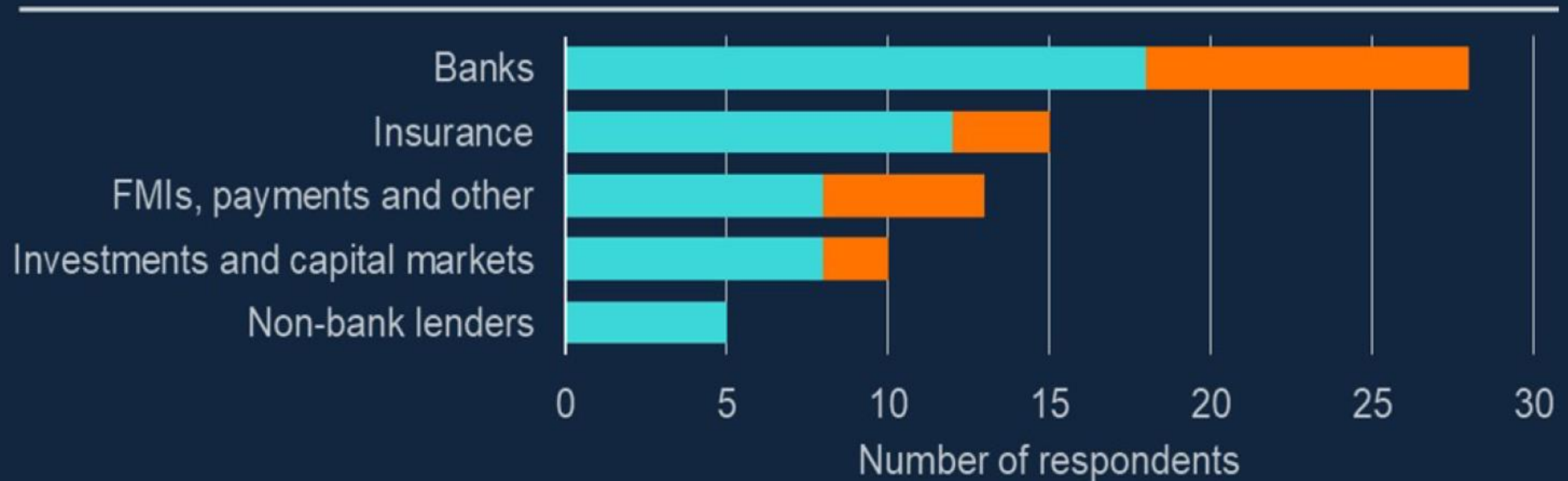
ML/AI GEN AI



Increasing reliability on ML/AI

Chart 1: 72% of firms that responded already use or are developing ML

■ Use ML ■ Do not use ML



GenAI & CDR



**GEN AI – TURING’S
CHILD:**

**LOOK WHAT I
CREATED!!!
= CRITICAL THINKING**

**“HUMAN IN THE LOOP”
RESPONSIBILITY
CDR**



UNIVERSITY OF
BIRMINGHAM

Are humans rational?



stress

“The only problem with common sense
Is that it is not very common”

Voltaire



UNIVERSITY OF
BIRMINGHAM

Rationality & CDR

Ambiguity effect	Focusing effect	Observation selection	
Anchoring	Forer effect	Observer-expectancy effect	
Attentional bias	Barnum	Subject-expectancy effect	
Availability heuristic	Framing effect	Omission bias	
Availability cascade	Recency illusion	Optimism bias	
Backfire	Gambler's law of large numbers	Wishful thinking,	
Bandwagon effect	Hard-easy effect	Valence effect	
Groupthink	Hindsight bias	Positive outcome bias	
Herd behaviour	Hostile media effect	Ostrich effect	
Base rate fallacy	Hot-hand fallacy	Overconfidence	
Belief bias	Hyperbolic discounting	Pareidolia	
Bias blind spot	Dynamic inconsistency.	Pessimism bias.	
Choice-supportive bias	Identifiable victim effect	Planning fallacy	
Clustering	Illusion of control	Post-purchase rationalization	
Confirmation	Illusion of validity	Pro-innovation bias	
Congruence bias	Illusory correlation	Pseudocertainty	
Conjunction	Impact bias	Reactive devaluation	
Conservatism - Bayesian	Information bias	Recency bias	Survivorship bias
Contrast	Insensitivity to sample size	Recency illusion	Texas sharpshooter fallacy
Curse of knowledge	Irrational Just-world	Restraint bias	Time-saving bias
Decoy effect	Less-is-better effect	Rhyme as reason effect	Unit bias
Denomination effect	Loss aversion	Risk compensation	Well travelled road effect
Distinction bias	Sunk cost effects	Peltzman effect	Zero-risk bias
Duration neglect	Ludic fallacy	Selective perception	
Empathy gap	Memory bias	Semmelweis reflex	
Endowment effect	Mere exposure effect	Selection Social comparison	
Essentialism	Mirror-imaging	Social desirability bias	
Exaggerated	Money illusion	Status quo bias	
Experimenter's	Moral credential effect	Stereotyping	
Expectation bias	Negativity bias	Subadditivity effect	
False-consensus	Neglect of probability	Subjective validation	
Functional fixedness	Normalcy bias	Survivorship	



Regulatory landscape

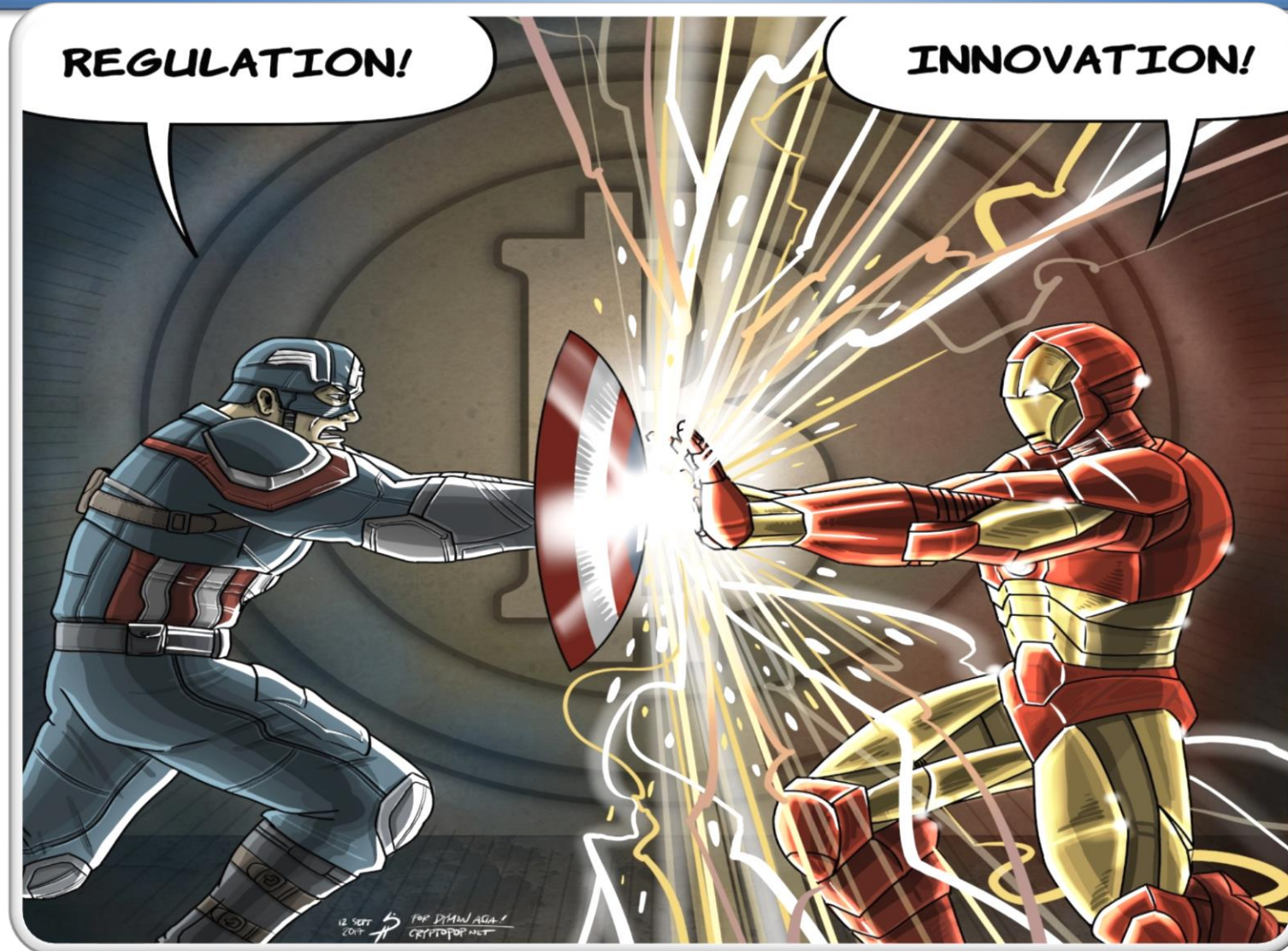


Image source: <https://cryptopop.net/2017/09/19/innovation-vs-regulation/>

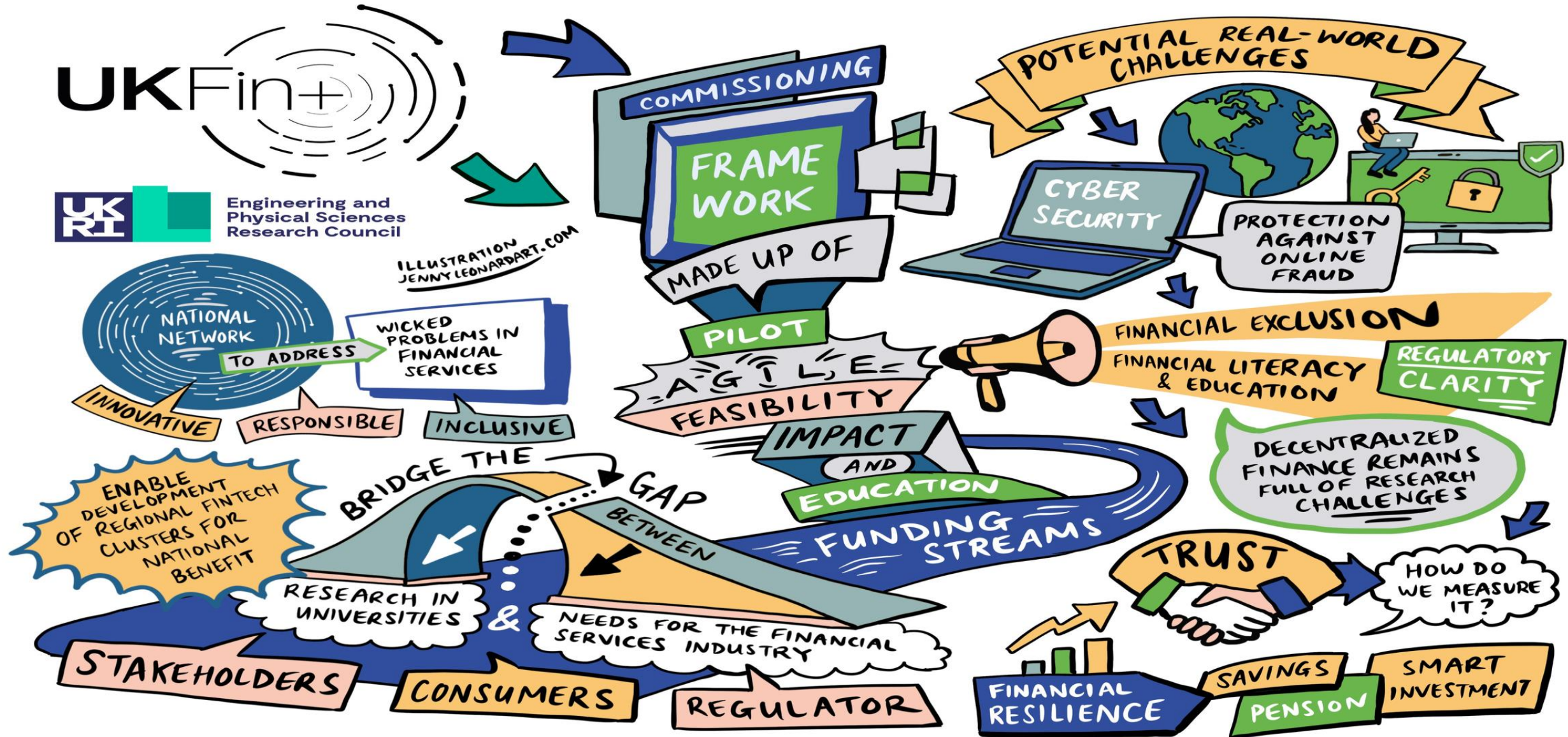


UNIVERSITY OF
BIRMINGHAM

Digital Containment



UKFin+ Network: Academic collaborAction



UKFin+ Network: Agile project/MVP



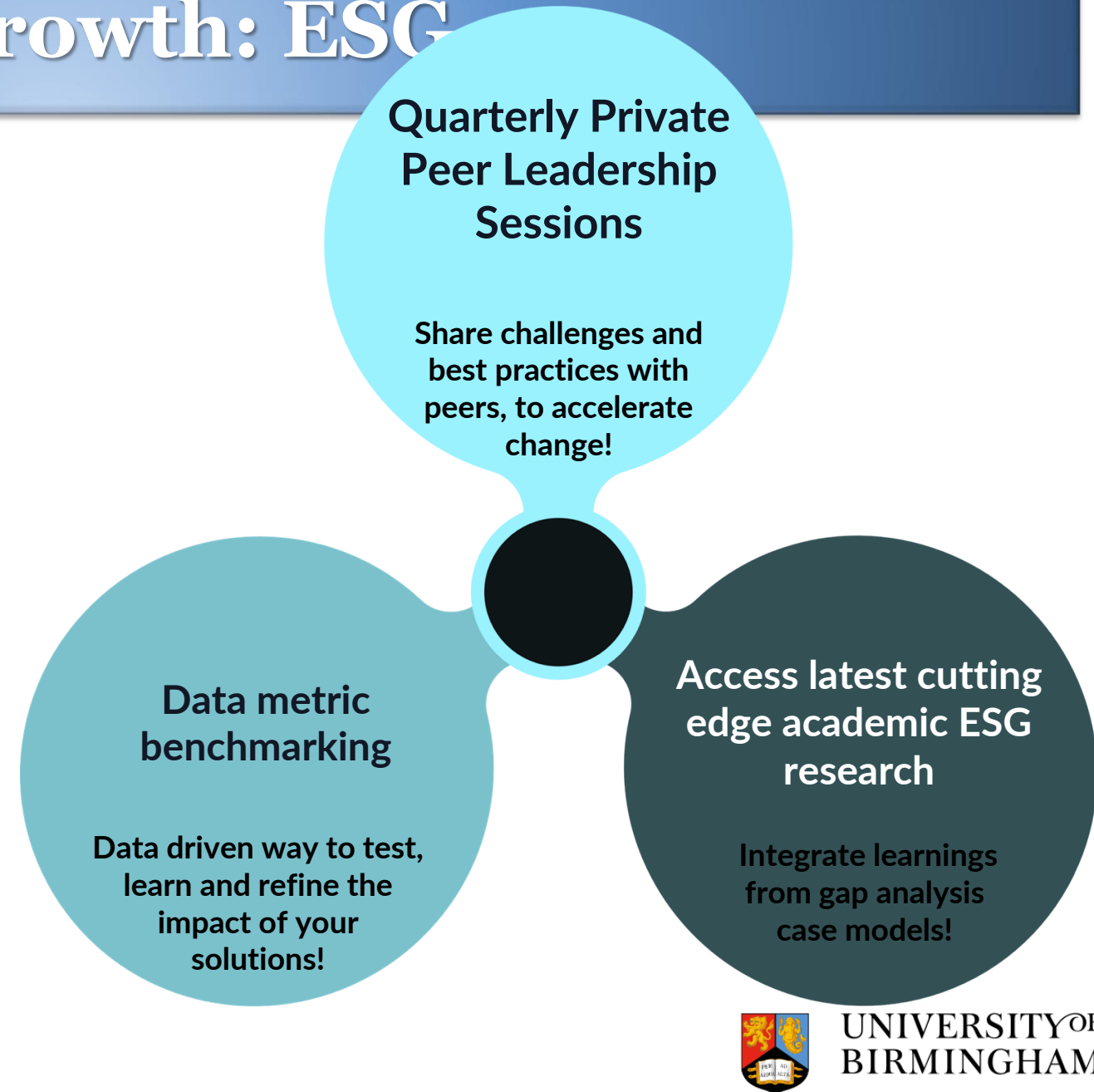
In online black markets (OBMs) many sell and buy cyber vulnerabilities and cybercrime services. Other actors can resell, repackaging, or use such vulnerabilities for hacking or to develop information security solutions. How OBMs work and maintain the fragile balance of ethics, entrepreneurship and innovation, raises a wicked problem for law enforcement, regulators, businesses and researchers. While the primary consideration is for the potential harm, the value created for fintech services and society deserves more attention.



Hub for Responsible Growth: ESG

The Hub & Co-Labs Midlands has been developed to help Midlands firms bridge business gaps, and address industry-wide issues with a starting focus on managing risks and regulations, breaking down silos.

Joining the Hub and Co-Labs enables organisations to collaborate in resolving the current ESG risk and compliance practical challenges facing industry. This includes, co-creating opensource ESG metrics and exploring real time deep fake video detection service. This will increase awareness of deep fake fraud and create a standard in terminology and knowledge of the relevant problem statements and use cases, testing and adoption – specifically for organisations using video identification verification tools. This is underpinned by data-driven evidence, tailored solutions and rigorous academic research.



Hub/Co-Lab model*

Global Collaboration, Local Value

Community

- Ecosystem is global and local, enabling local value + growth
- Events, briefings, content, working groups, Q&A
- Collaboration model, impact-focused business model

Creation

- Creating services, solutions and support – bridging business + tech
- Industry Platforms, Hackathons, Proof of Concepts
- Democratising value, ensuring long term value and adoption

Curation

- Creating a global, transparent, structured curation model
- Benchmarking data that 'breaks the mould' and allows 'good' bias
- Ensuring better outcomes for all, by fairer access to support

EU AI Act: consideration/context

3. Post-Market Monitoring: The importance of continuous oversight.

3.1 Monitoring System Requirements

Establishing a post-market monitoring system.

3.2 Data Collection and Analysis

Gathering and analysing data on AI system performance.

3.3 Post-Market Monitoring Plan

Developing a comprehensive monitoring plan.

3.4 Integration with Existing Legislation

Aligning AI monitoring with other EU harmonization legislation.

The European Union's AI Act represents a pioneering step towards regulating artificial intelligence, with a particular focus on high-risk AI systems. Central to this regulatory framework is the concept of post-market monitoring, a process designed to ensure that AI systems continue to operate safely and in compliance with established standards throughout their lifecycle. This involves the systematic collection, documentation, and analysis of performance data, enabling providers to assess and maintain the integrity of AI systems post-deployment (Shaw, 2024).



CDR, Your Responsibility & Why Now?

Why now?

Access to Money
Access to Talent
Reputational Damage

CDR: A FRAMEWORK TO GUIDE IMPACT DECISION MAKING

- Measurement of everything, calculators galore, will NOT save society and the planet.
- Pragmatic and disruptive innovations, using digital technologies, must solve challenges and create value. It is all we understand.

<https://digitalresponsibilityforum.com/>
<https://corporatedigitalresponsibility.net/>

(Permissioned use from Price, R. 2023)



UNIVERSITY OF
BIRMINGHAM

Starts & Ends with PEOPLE



A word cloud featuring the phrase "Thank You" in numerous languages. The central and largest text is "thank you" in blue. Surrounding it are other languages such as Spanish ("gracias", "gracias"), French ("merci", "merci"), German ("danke", "danke"), Italian ("grazie", "grazie"), Japanese ("ありがとう", "ありがとう"), Korean ("감사합니다", "감사합니다"), and many others. The words are arranged in a circular pattern, with some words appearing multiple times in different colors and sizes.