695th Lord Mayor's Ethical AI Initiative
Delivering Ethics Courses For AI Deployers & Builders

Professor Michael Mainelli
Chairman, Z/Yen Group
27 June 2023

@mrmainelli

michael_mainelli@zyen.com

www.zyen.com

Stability AI: "steampunk 'Lord Mayor' of the 'City of London' standing in front of Tower Bridge wearing purple goggles with his battle puffin

Agenda

- What are we kicking-off today?
- What's AI anyway?
- Our 'third way' approach, standards
- Commercial ethics
- The 695th Lord Mayor's Ethical AI Initiative

"Get a detailed grip on the big picture."
Chao Kli Ning

Today's Kick-off

Aim - ethical AI development and deployment
Proposed qualifications: Deployer and Builder certificates

Problems

- AI's potential impact on society
- Importance of ethical considerations
- Data privacy concerns
- Need for transparency, accountability, and inclusivity
- Concerns about bias, transparency, accountability, and safety

Goals

- Ensuring AI benefits society and minimizes harm
- Establishing ethical guidelines and principles
- Encouraging transparency and accountability
- Promoting collaboration and dialogue among stakeholders

What's AI Anyway?

'artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;
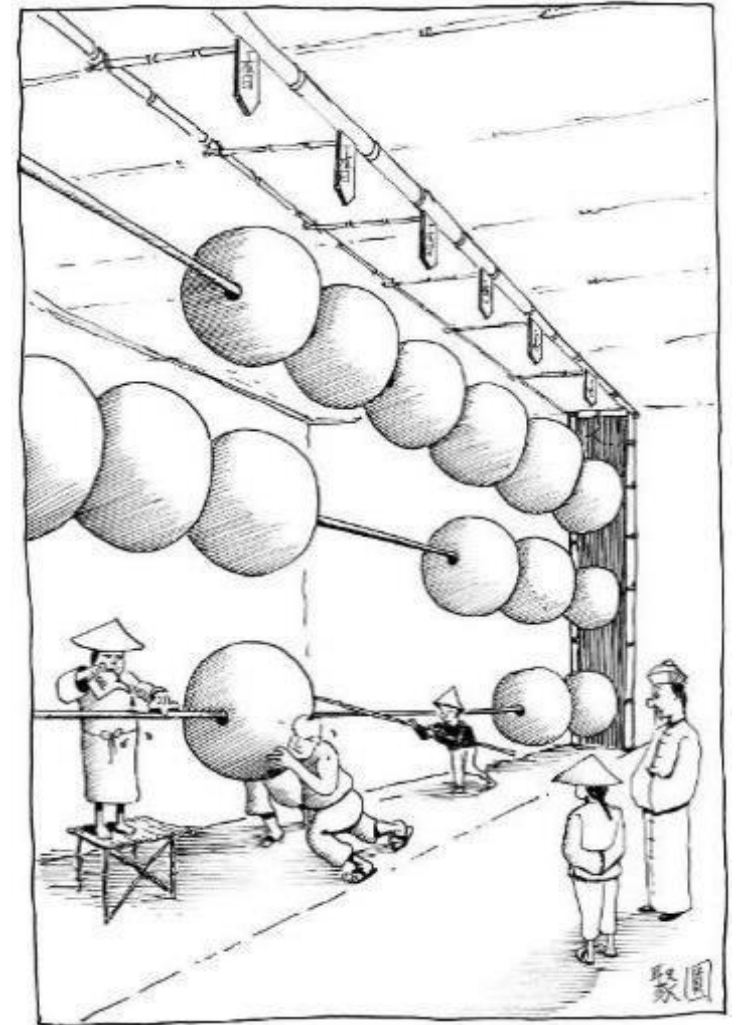
ANNEX I
ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES
referred to in Article 3, point 1
(a)Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
(b)Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
(c)Statistical approaches, Bayesian estimation, search and optimization methods.

Look DAPR

- Dynamic
  - Initiates an action
  - Real-time
- Anomaly & Pattern
  - Identifies unusual behaviour
  - Reinforces successful patterns
- Response
  - Adaptive, moves with the data
  - Integrative, can work without rebuilding the entire IT architecture

"Get a big picture grip on the details."
Chao Kli Ning

More Definition?

"An LLM is designed to produce text that reflects patterns found in its vast training data"

The Economist, "The Ghost In The AI Machine", 24 June 2023, page 79.

AI systems use historic data and algorithms to produce probabilistic outputs:

- probabilistic if-then engines;

- testing, inspection, and certification need to evaluate ranges and probability distributions;

- all historic data is biased – all future data is biased;

- dynamic anomaly & pattern response;

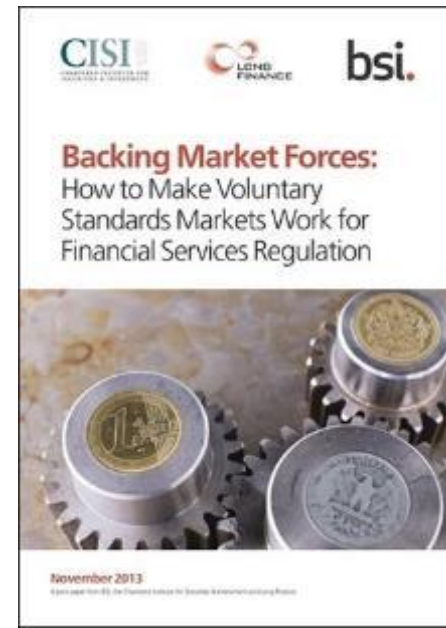- wider system, i.e. retraining & recalibration, are crucial.

Evaluating Machine-Learning (AI)

	Beat the **Human**	Beat the **Machine**
Training data environment **Stable/Gargantuan**	**OK - Static**	**OK - Calibrate**
Training data environment **Volatile/Hard-to-Obtain**	**Autonomy Unlikely**	**Simple Algorithms Only**

All data contains bias, even data manufactured to be without bias.

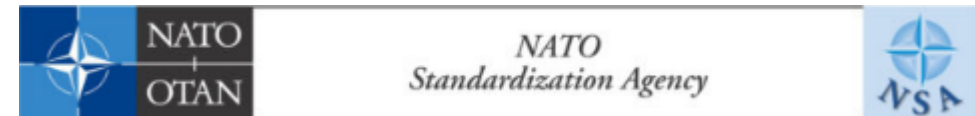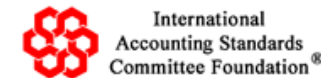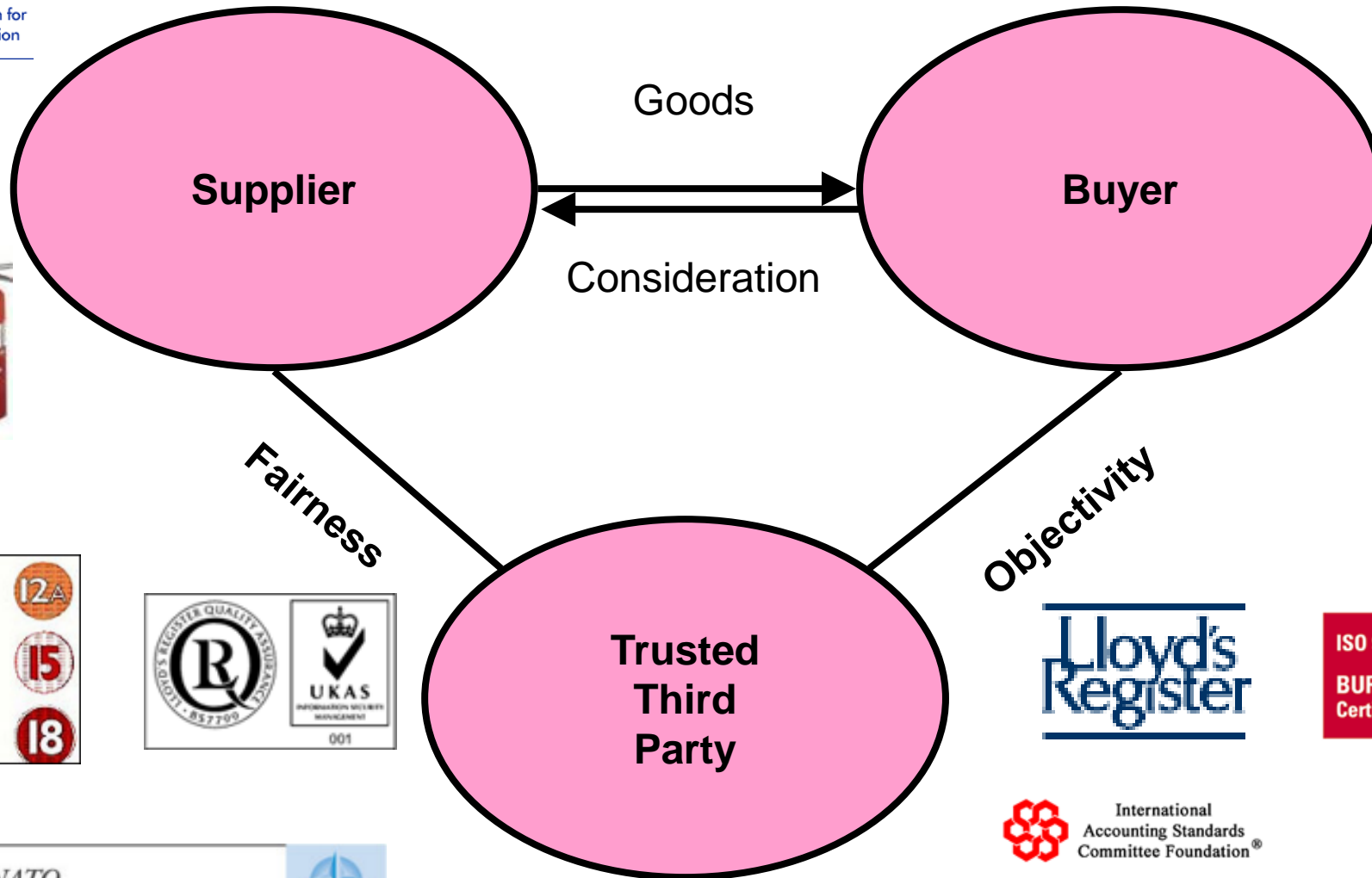# Enslaved East versus <u>Standards Markets</u> versus Wild West

1. Human-centric Approach: The EU AI law emphasizes the importance of ensuring that artificial intelligence (AI) systems are developed and used in a way that respects human rights, ethical values, and fundamental principles.
2. Risk-Based Approach: The law categorizes AI systems into different risk levels based on the potential harm they can cause.
3. Prohibition of Unacceptable Practices: The law prohibits certain AI practices that are considered unacceptable due to their potential to infringe on fundamental rights and freedoms. This includes AI systems that manipulate human behavior, create social scoring systems, and exploit vulnerabilities of specific groups.
4. Transparency and Explainability: The EU AI law emphasizes the importance of transparency and explainability in AI systems. It requires that users are provided with clear and understandable information about the capabilities and limitations of AI systems. High-risk AI systems must also provide explanations for their outcomes to ensure accountability and mitigate biases.
5. Data Governance and Quality: The law promotes responsible data governance by requiring that AI systems are trained on high-quality and unbiased data.
6. Accountability and Oversight: The EU AI law places responsibilities on developers, providers, and users of AI systems. It mandates that high-risk AI systems undergo conformity assessment procedures and be subject to oversight by competent authorities.

CISI   LEND FINANCE   bsi.

Backing Market Forces:
How to Make Voluntary Standards Markets Work for Financial Services Regulation
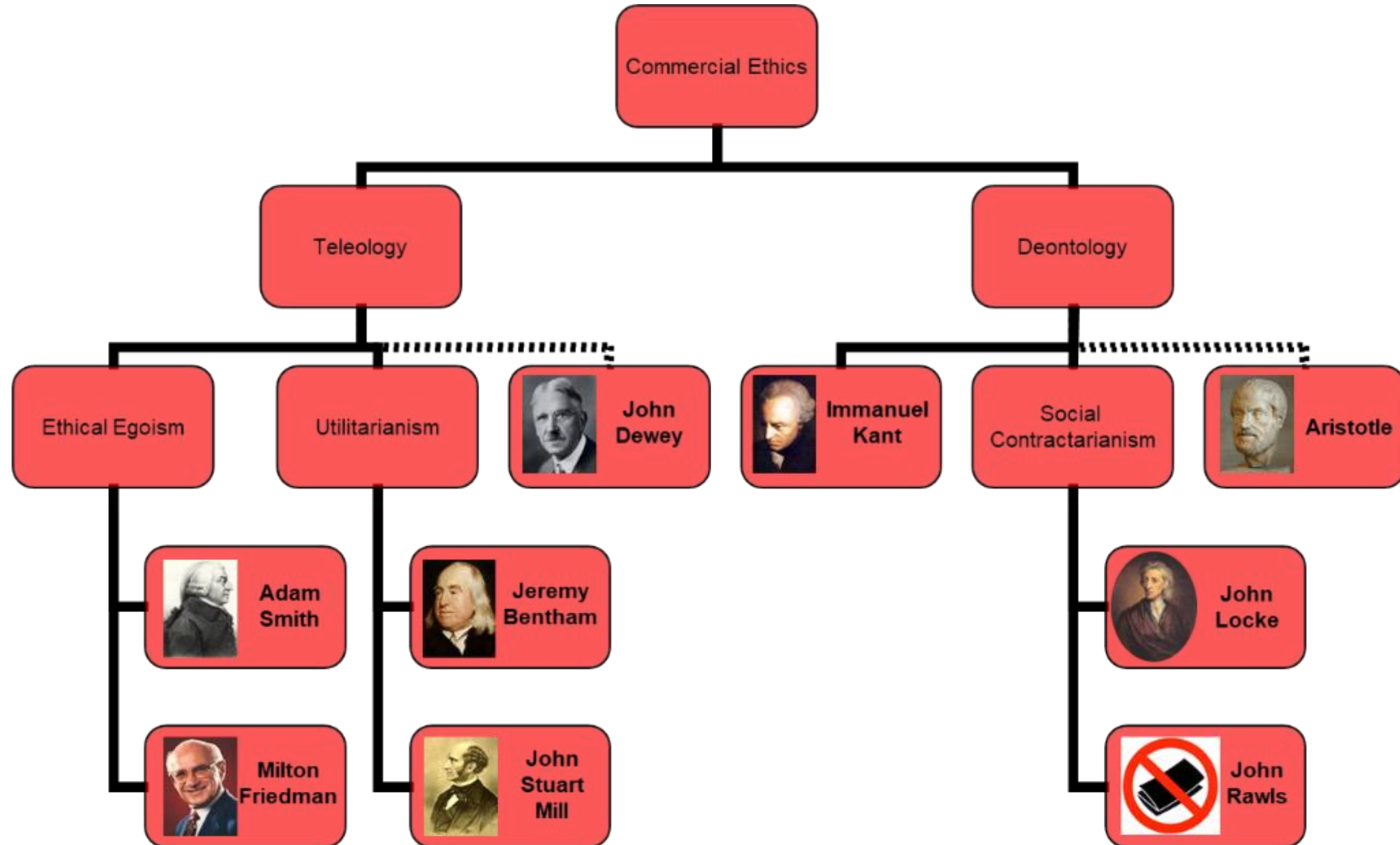
November 2013

Bureaucracy or AI?

Standards Markets – The Third Way - Trusted Third Parties

Course Materials Abundant, And Venerable

- Establish Endeavour – why? background to ethics and schools of ethics (2 hours)
  - UK government guidance - https://www.gov.uk/guidance/data-ethics-and-ai-guidance-landscape
  - UK AI Regulation Approach (White Paper https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach & updated versions thereof)
  - Universal Guidelines for Artificial Intelligence - https://archive.epic.org/international/AIGuidleinesDRAFT20180910.pdf
  - G20 AI Principles - https://oecd.ai/en/wonk/documents/g20-ai-principles
  - OECD AI Principles - https://www.oecd.org/digital/artificial-intelligence/
  - Australia's 8 Ethics Principles - https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles
- Assess & Appraise – what? defining AI and finding AI (3 hours)
  - ISO/IEC TR 24368:2022 standard and the relevant terms
  - Proposal For A Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts - https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
- Lookaheads & Likelihoods – where? what are the ethical issues and risks? (2 hours)
  - "Information technology — Artificial intelligence — Overview of ethical and societal concerns", PD ISO/IEC TR 24368:2022
  - Information technology — Artificial intelligence — Guidance on risk management", BS ISO/IEC 23894:2023
  - Sources for keeping up-to-date
- Options & Outcomes – who? people and roles (2 hours)
  - "Draft BS ISO/IEC 42001 Information Technology - Artificial intelligence - Management system"
  - understanding 'stakeholders'
- Understanding & Undertaking – how? (2 hours)
  - Z/Yen & BCS DIODE methodology - https://www.zyen.com/media/documents/ethical-assessment.pdf
  - Citigroup methodology

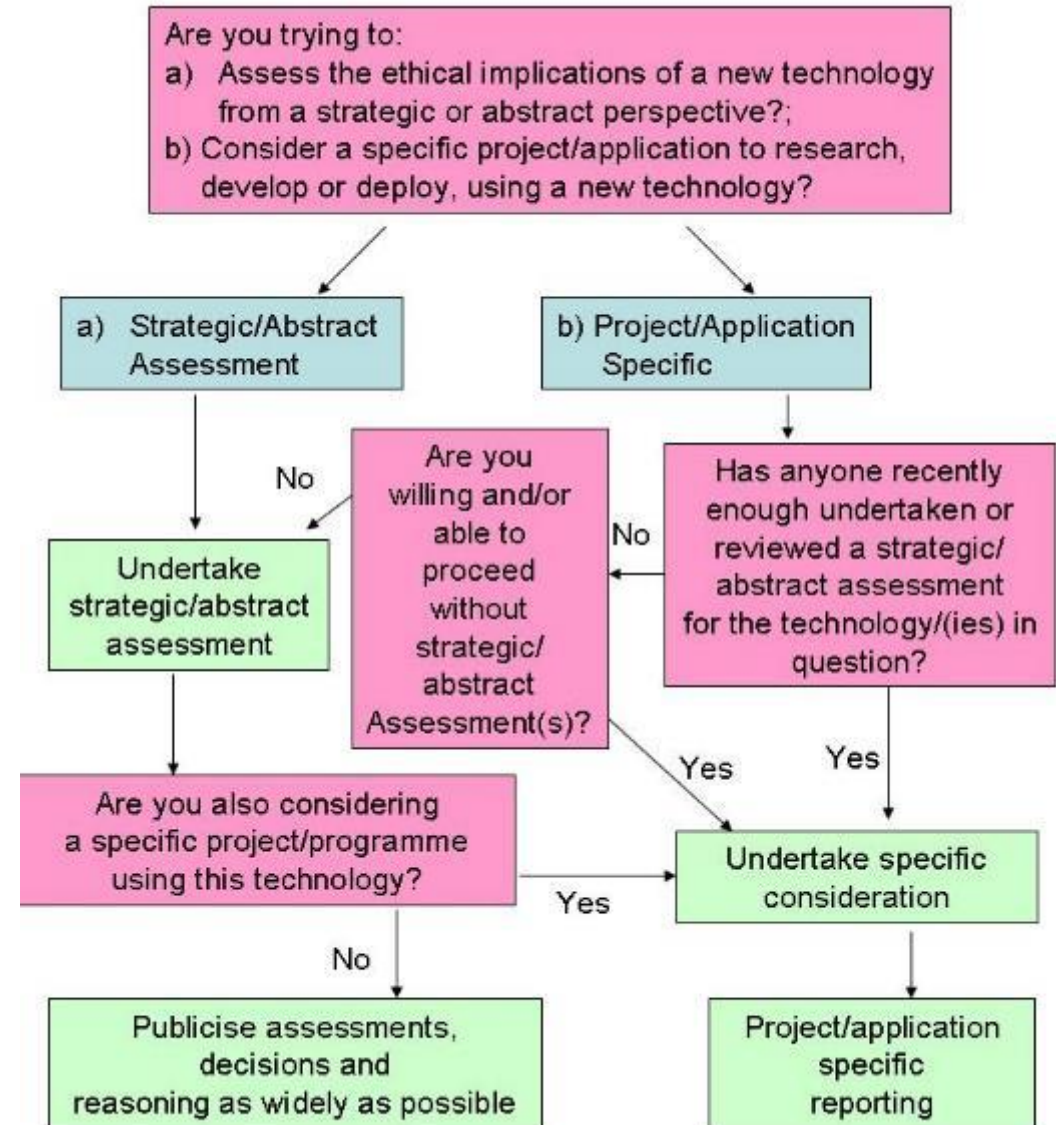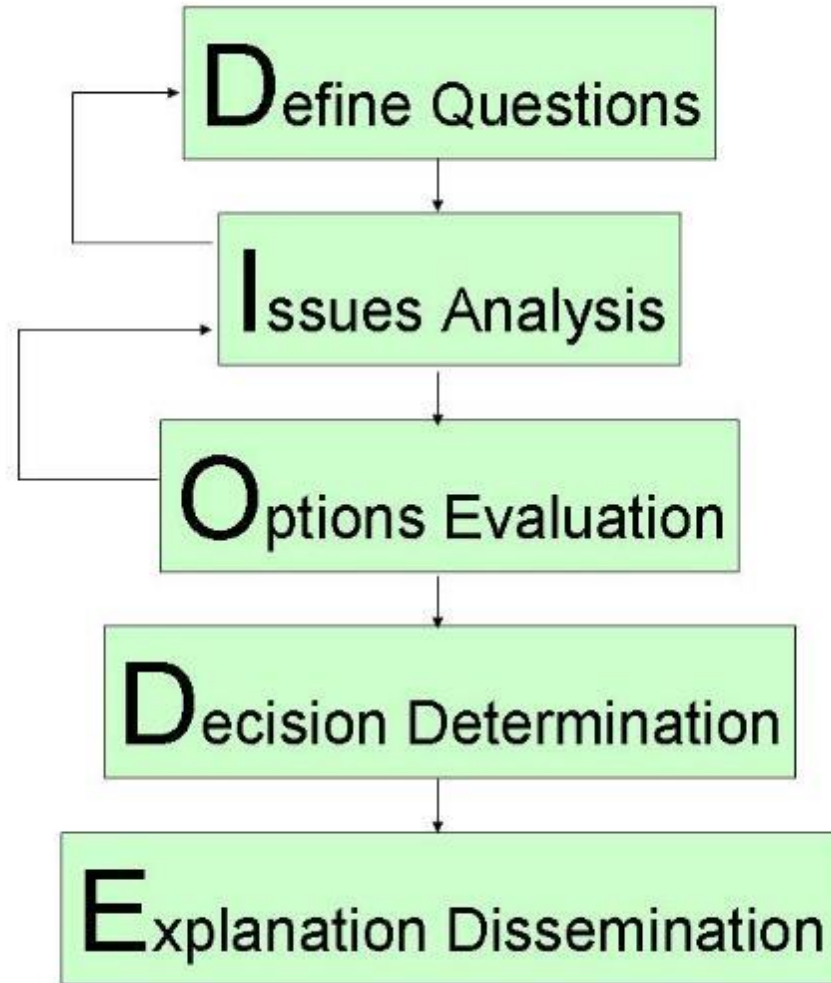# Commercial Ethics In A Nutslide

Principlism

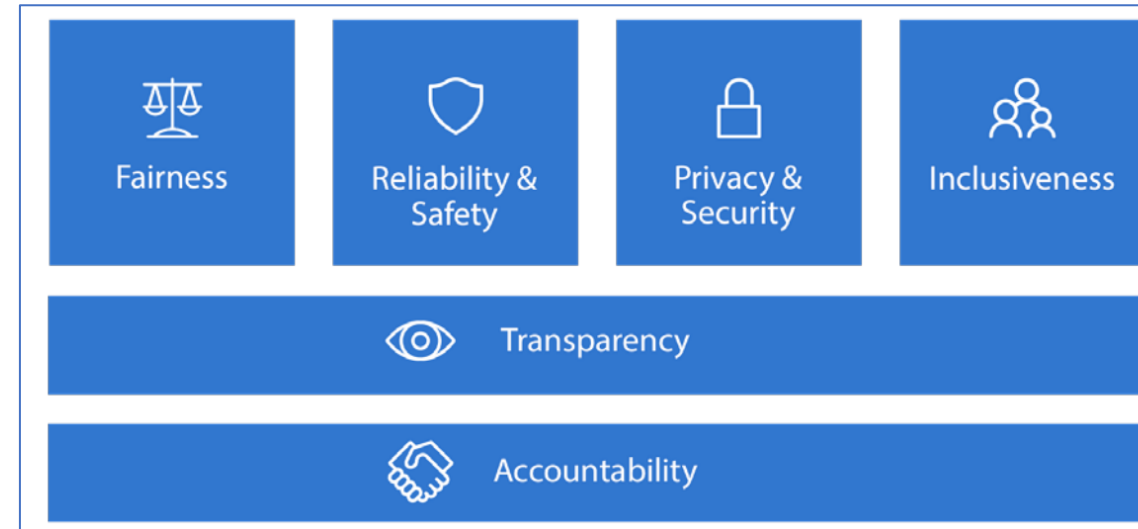|	
Autonomy	Respect for persons. Humans are a means unto themselves and not a means to an end. This encompasses the right to be free to make choices about your body.
Beneficence	Obligation to contribute to person's welfare. Interventions and provisions should provide benefit directly to the patient. This focuses on doing things that are of benefit to another. It requires positive steps to help, and not merely avoiding doing harm.
Nonmaleficence	Obligation not to inflict harm on other persons. Harm is to be avoided or minimized. Underlying tenet of medical professional mission statements (Hippocratic oath).
Justice	For health care, this is the distribution of health (and health care) in a fair and equitable manner. This requires attention to prioritization and rationing. There is no one just way to allocate resources, and most systems utilize several prioritization schemes in concert to attempt to achieve a just distribution.

"Get a detailed grip on the big picture."
Chao Kli Ning

Ethical Decision Making - DIODE

Define Questions

Issues Analysis

Options Evaluation

Decision Determination

Explanation Dissemination

Are you trying to:
a) Assess the ethical implications of a new technology from a strategic or abstract perspective?;
b) Consider a specific project/application to research, develop or deploy, using a new technology?

a) Strategic/Abstract Assessment

b) Project/Application Specific

No

Undertake strategic/abstract assessment

Are you willing and/or able to proceed without strategic/ abstract Assessment(s)?

No

Has anyone recently enough undertaken or reviewed a strategic/ abstract assessment for the technology/(ies) in question?

Are you also considering a specific project/programme using this technology?

Yes

Yes

Yes

Undertake specific consideration

No

Publicise assessments, decisions and reasoning as widely as possible

Project/application specific reporting

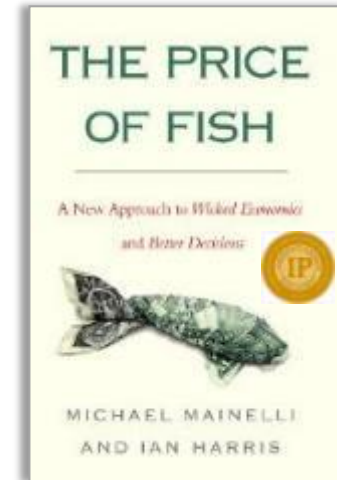# 695th Lord Mayor's Ethical AI Initiative

- 'Students' start in September 2023
- Initiative formal launch in mid-November
- Certificate awards ceremony in Q2 2024
- Second stage to cover additional domains, e.g. law – solicitors & barristers, accountancy, surveying, pharma, ...
- Two types of qualifications: AI Deployers and AI Builders
- CISI, BCS, and others as certificate providers for Deployers
- Exploration of AI Builder course with Open Data Institute
- Exploration of Firmwide certification

When Would We Our Know Commerce Is Working?

"Get a big picture grip on the details."
Chao Kli Ning

Thank you!