Artificial Intelligence: Our Saviour Or Destroyer?

Professor Michael Mainelli
Chairman, Z/Yen Group
19 July 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

Pew Research Center

Experts Doubt Ethical AI Design Will Be Broadly Adopted As the Norm Within the Next Decade

A majority worries that the evolution of artificial intelligence by 2030 will continue to be primarily focused on optimizing profits and social control. They also cite the difficulty of achieving consensus about ethics. Many who expect major progress say it is not likely within the next decade. Still, a portion celebrate coming AI breakthroughs that will improve life

By Lee Rainie, Janna Anderson and Emily A. Vogels

FOR MEDIA OR OTHER INQUIRIES:

Lee Rainie, Director, Internet and Technology Research
Janna Anderson, Director, Elon University's Imagining the Internet Center
Haley Nolan, Communications Associate

202.419.4372
www.pewresearch.org

RECOMMENDED CITATION

Pew Research Center, 2021. "Experts Doubt Ethical AI Design Will Be Broadly Adopted As the Norm in the Next Decade"
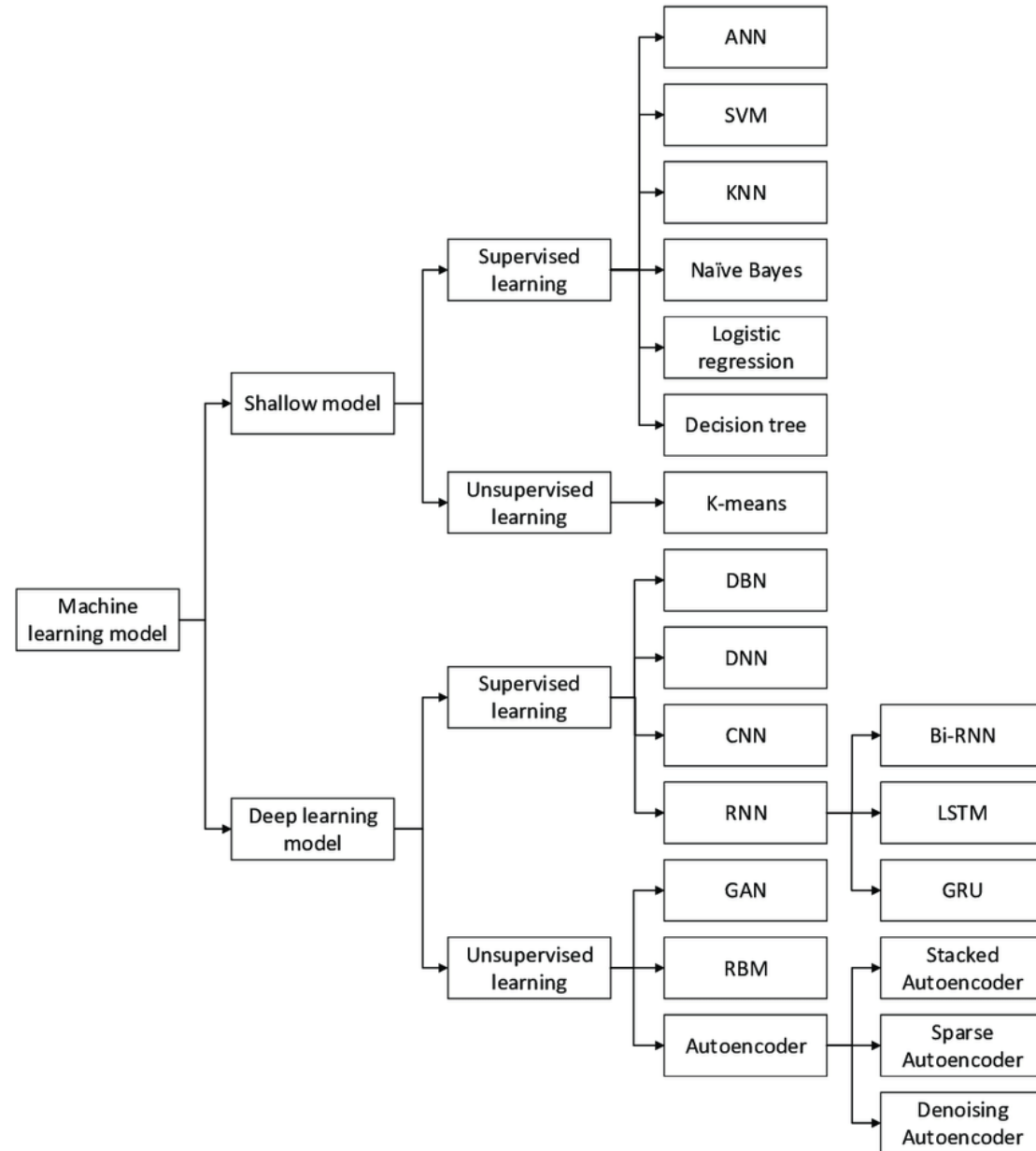
Agenda

- What's AI anyway?
- Saviour or Destroyer?
- Commercial & AI ethics
- A 'third way' approach to control – community standards markets
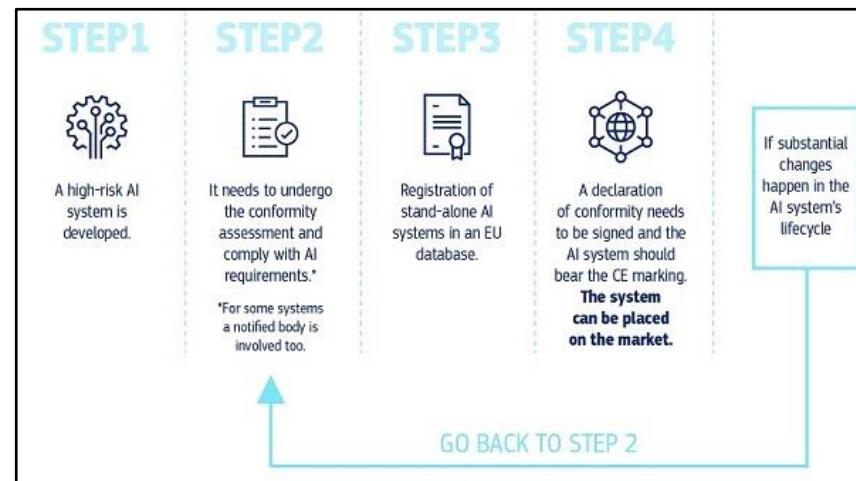- The 695th Lord Mayor's Ethical AI Initiative

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

Machine Learning - Tools & Techniques

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 behaviour, create social scoring systems, and exploit vulnerabilities of specific groups.
4. **Transparency & Explainability**: The EU AI law emphasizes the importance of transparency and explainability in AI systems - users to be provided with clear and understandable information about capabilities and limitations of AI.
5. **Data Governance & Quality**: The law promotes responsible data governance by requiring that AI systems are trained on high-quality and unbiased data.
6. **Accountability & 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.

STEP1	STEP2	STEP3	STEP4	
A high-risk AI system is developed.	It needs to undergo the conformity assessment and comply with AI requirements.* *For some systems a notified body is involved too.	Registration of stand-alone AI systems in an EU database.	A declaration of conformity needs to be signed and the AI system should bear the CE marking. **The system can be placed on the market.**	If substantial changes happen in the AI system's lifecycle

GO BACK TO STEP 2

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.
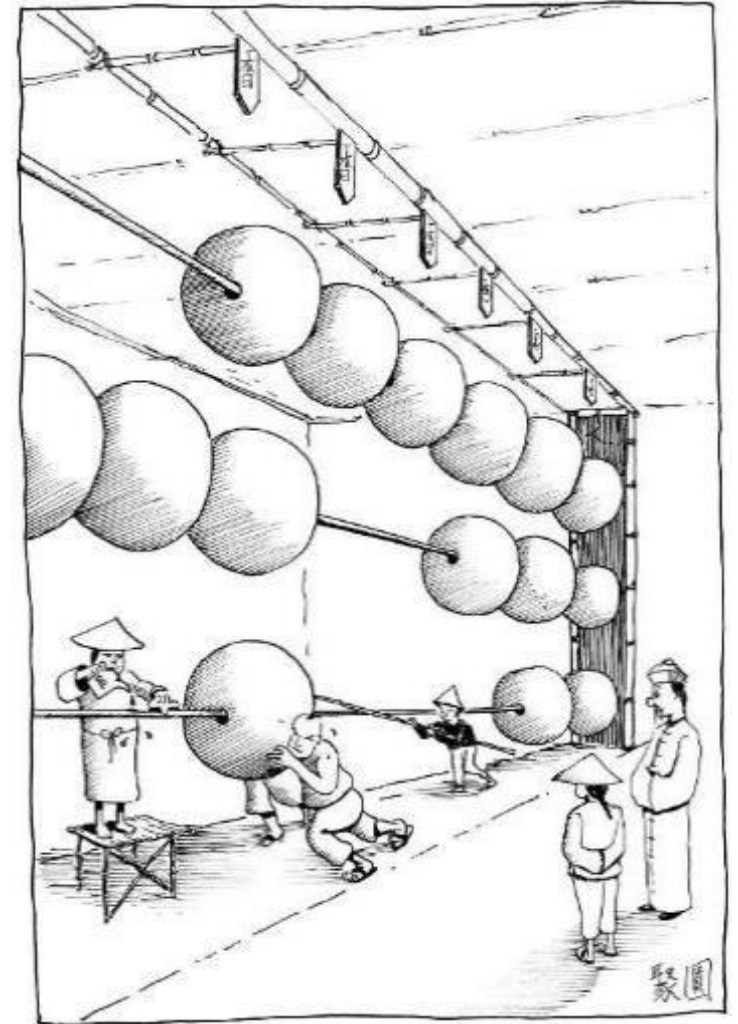
Bureaucracy or AI?

Look DAPR

AI

"an autonomous probabilistic, dynamic anomaly & pattern response system"

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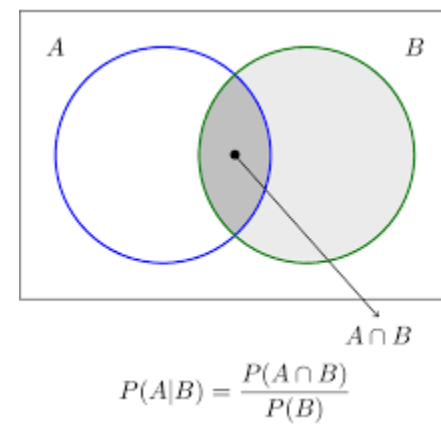
	Beat the **Human**	Beat the **Machine**
Training data environment **Stable/Gargantuan**	**OK - Static**	**OK - Calibrate**
Training data environment **Volatile/Hard-to-Obtain**	**Autonomy Unlikely**	**Simple Algorithms Only**

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

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

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

1940-1950 - Not A New Issue (sic) – 2 December 1950
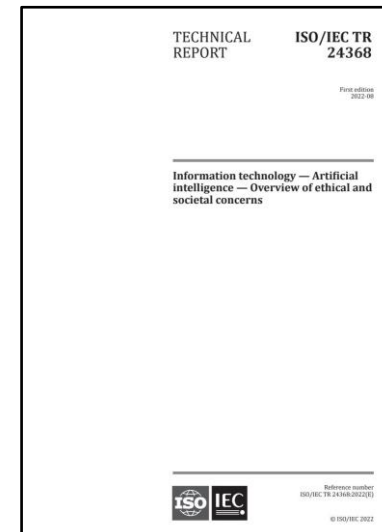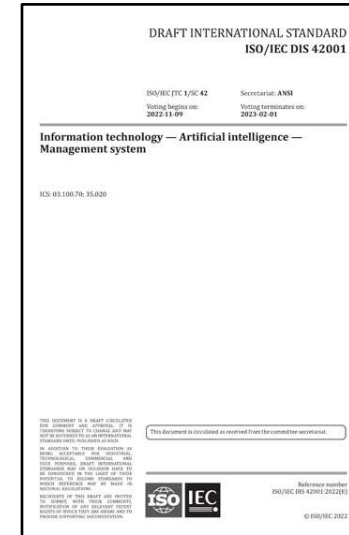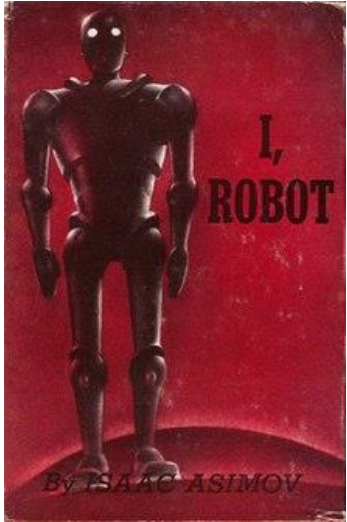
The Three Laws, from the "Handbook of Robotics, 56th Edition, 2058 AD":

First Law
A robot may not injure a human being or, through inaction, allow a human being
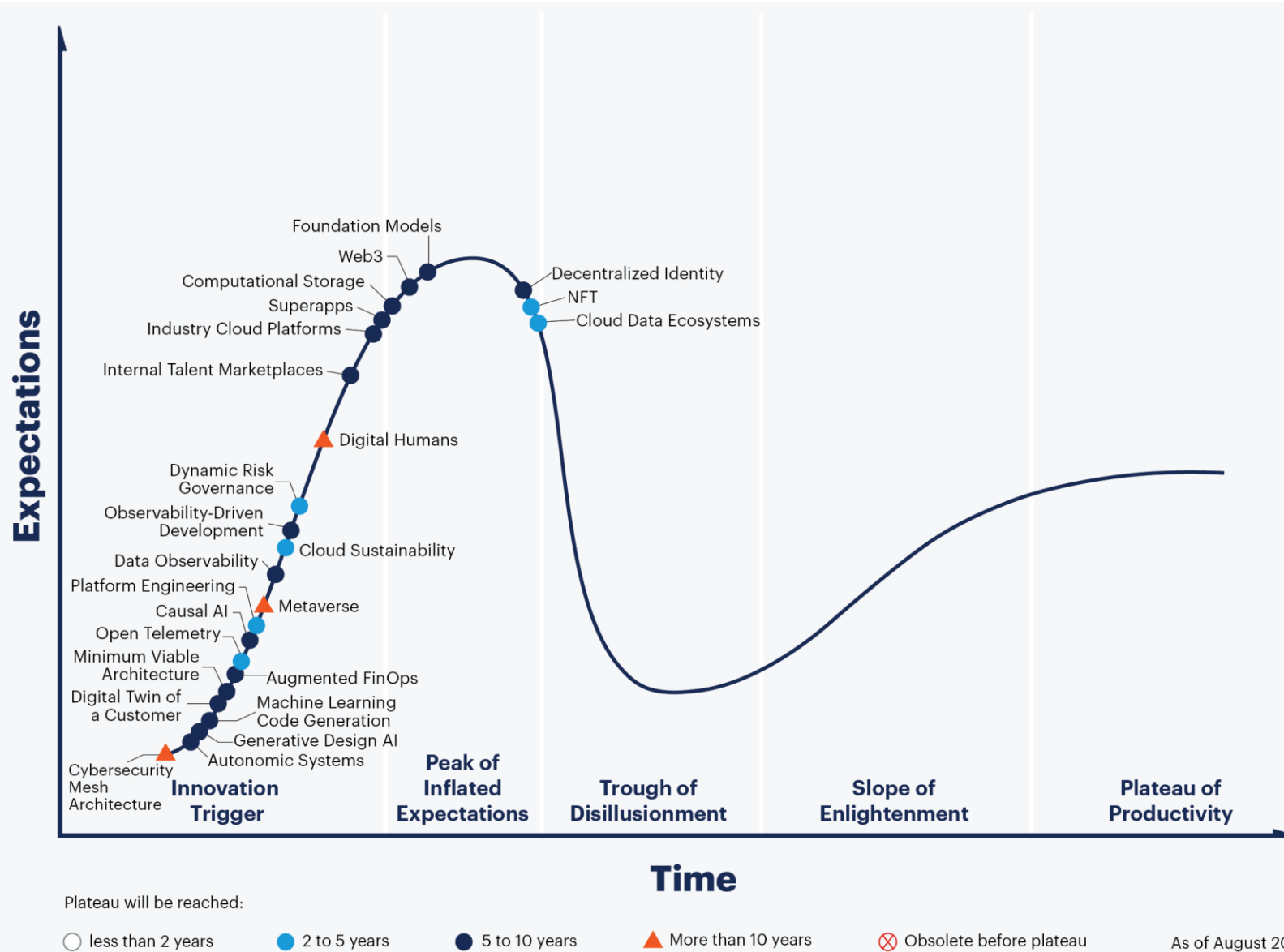to come to harm.

Second Law
A robot must obey the orders given it by human beings except where such orders
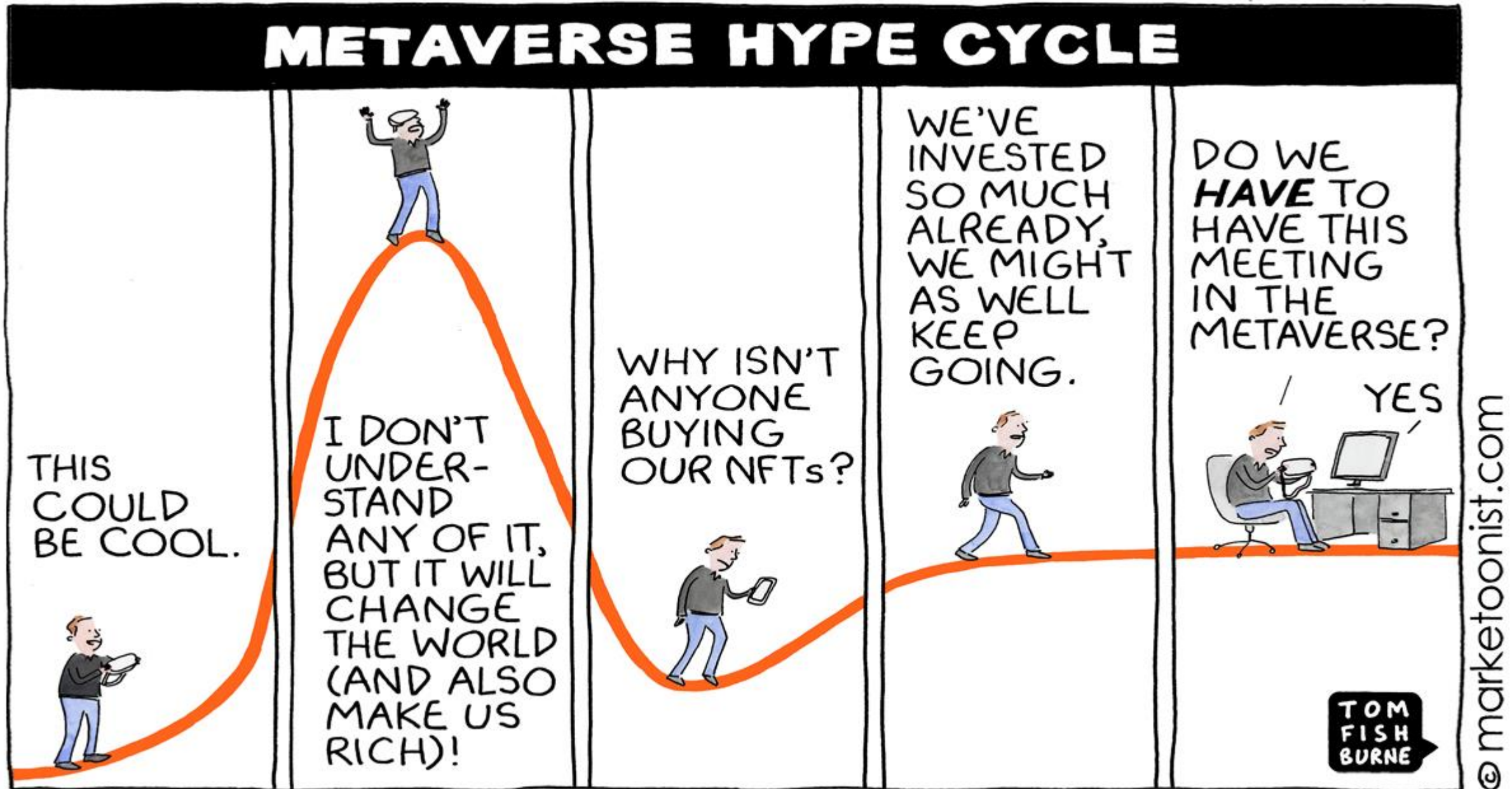would conflict with the First Law.

Third Law
A robot must protect its own existence as long as such protection does not
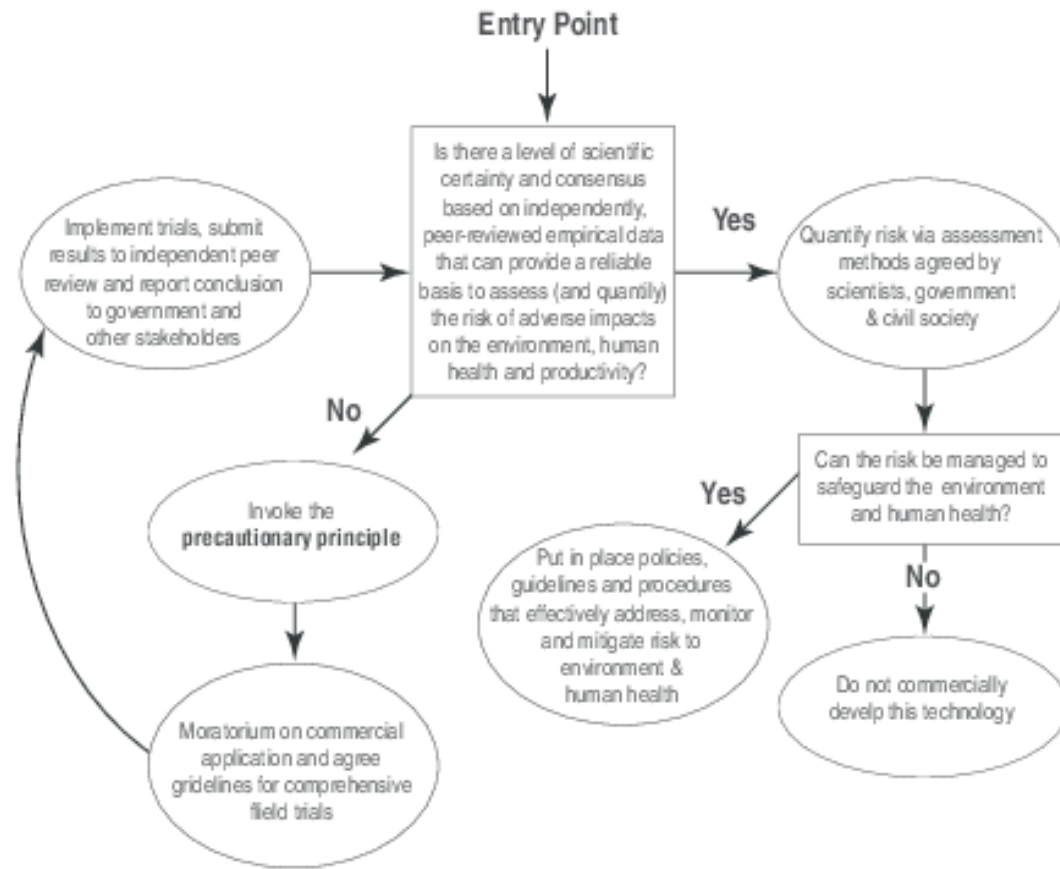conflict with the First or Second Law.

Hype The Hype Cycle

Saviour Or Destroyer Or Muddledom?

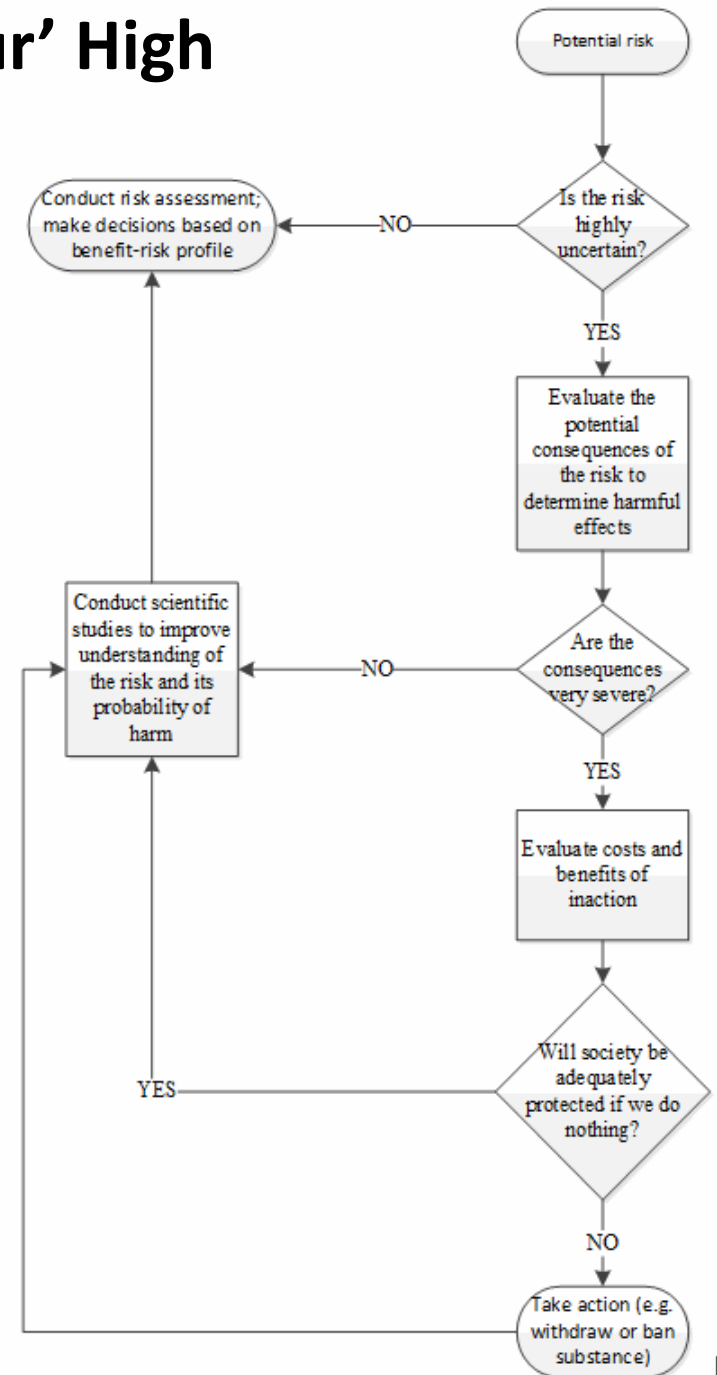|		A) Action Not Taken	B) Action Taken
The Precautionary Principle helps us choose whether an action should, or should not be done, without knowing the risks with certainty.		The Columns A+B Represent a Human CHOICE — Which column should we choose?	
These Two Rows are a **Prediction** or **Guess**. The certainty of risk is unknown and unknowable. We cannot choose the row.	Grave-Harm Caused=False	Benefits of Action not taken. **Life continues**	Benefits Enjoyed. No harm done!
	Grave-Harm Caused=True	Benefits of Action not taken. **Life Continues**	**Extreme Catastrophe** Economic or Environmental Collapse Large Loss of Life **WORST CASE SCENARIO**

https://www.sourcewatch.org/index.php?title=Precautionary_principle

Circular Decisions – Potential 'Saviour' High

Entry Point

Is there a level of scientific certainty and consensus based on independently, peer-reviewed empirical data that can provide a reliable basis to assess (and quantify) the risk of adverse impacts on the environment, human health and productivity?

Yes → Quantify risk via assessment methods agreed by scientists, government & civil society

No → Invoke the precautionary principle

Implement trials, submit results to independent peer review and report conclusion to government and other stakeholders

Can the risk be managed to safeguard the environment and human health?

Yes → Put in place policies, guidelines and procedures that effectively address, monitor and mitigate risk to environment & human health

No → Do not commercially develop this technology

Moratorium on commercial application and agree guidelines for comprehensive field trials

https://www.researchgate.net/figure/The-precautionary-principle_fig1_237817450

Potential risk

Is the risk highly uncertain?
- **NO** → Conduct risk assessment; make decisions based on benefit-risk profile
- **YES** → Evaluate the potential consequences of the risk to determine harmful effects

Are the consequences very severe?
- **NO** → Conduct scientific studies to improve understanding of the risk and its probability of harm
- **YES** → Evaluate costs and benefits of inaction

Will society be adequately protected if we do nothing?
- **YES** → (to Conduct scientific studies)
- **NO** → Take action (e.g. withdraw or ban substance)

https://www.pharmaceuticalonline.com/doc/quality-risk-management-a-brief-history-of-risk-management-in-the-regulation-of-medicinal-products-0001
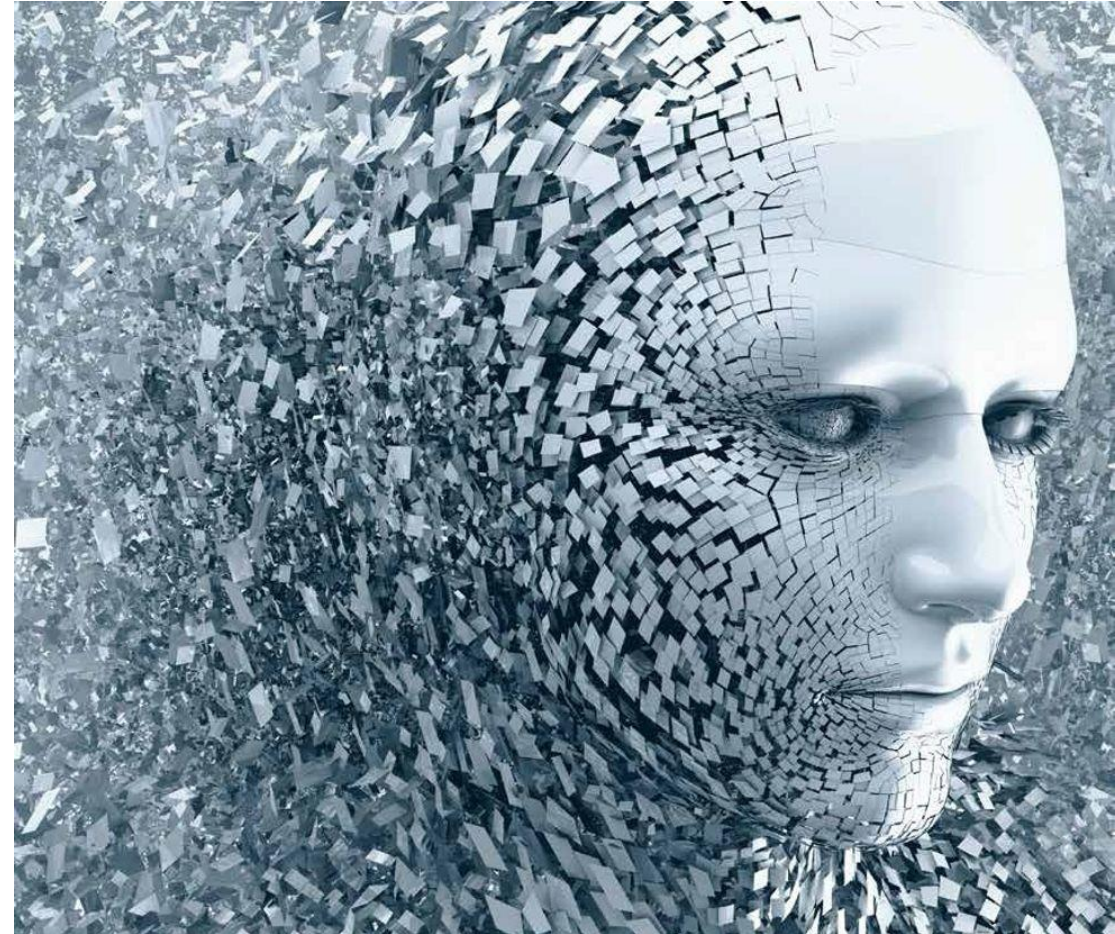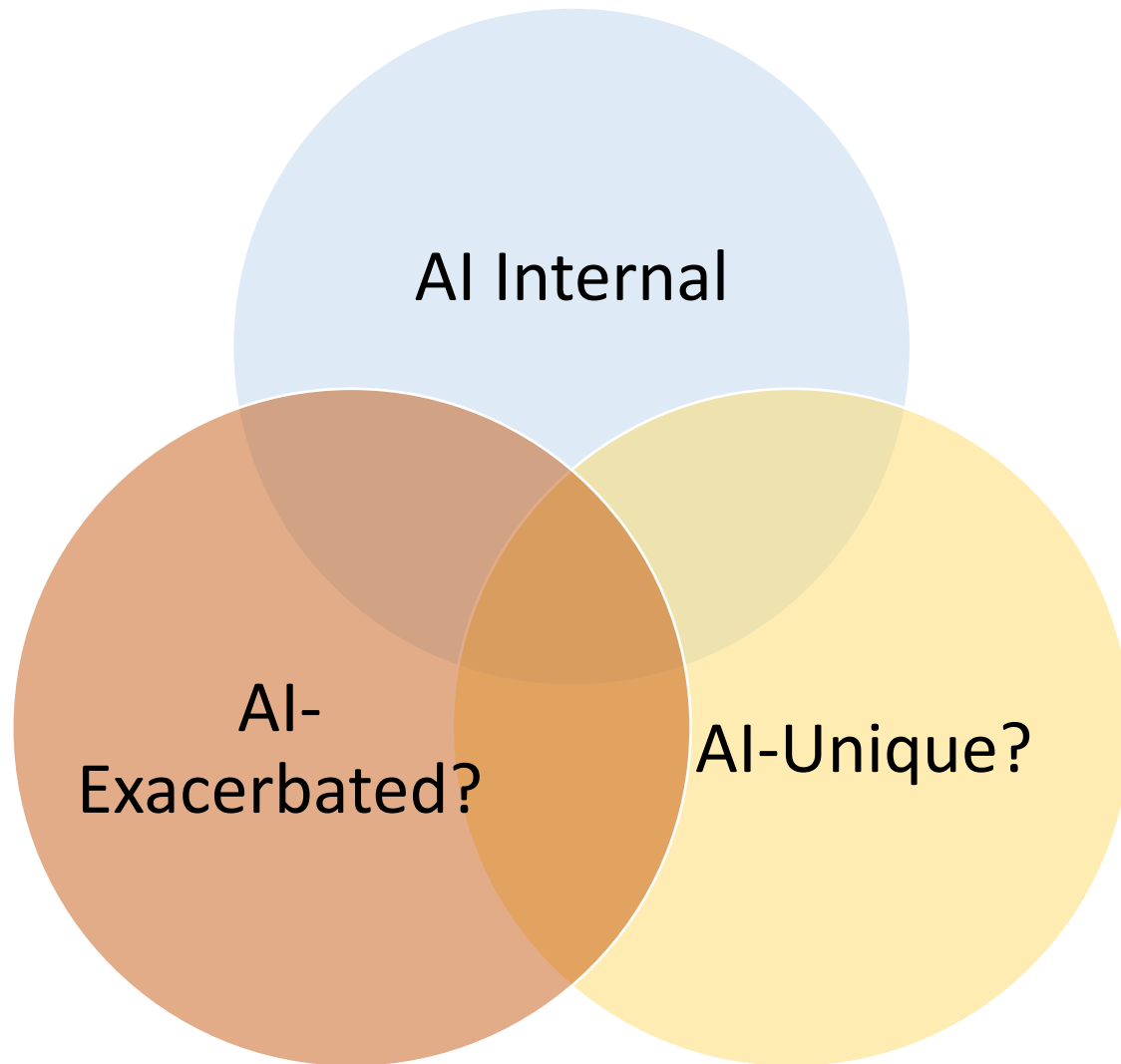
Woke Up To Ethics

New Ethics Needed?
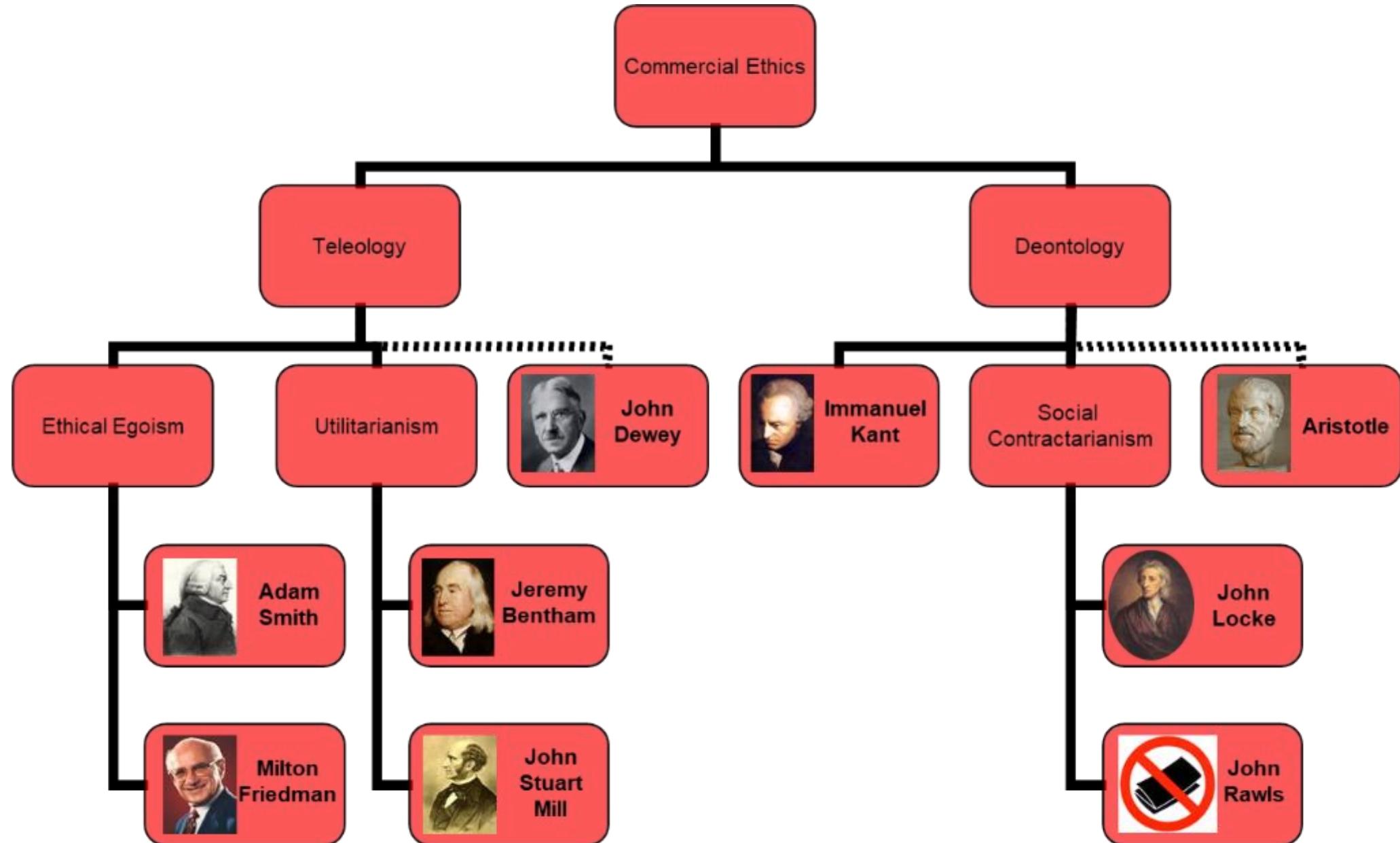
AI Internal

AI-Exacerbated?

AI-Unique?

What is Ethics in Research & Why is it Important?

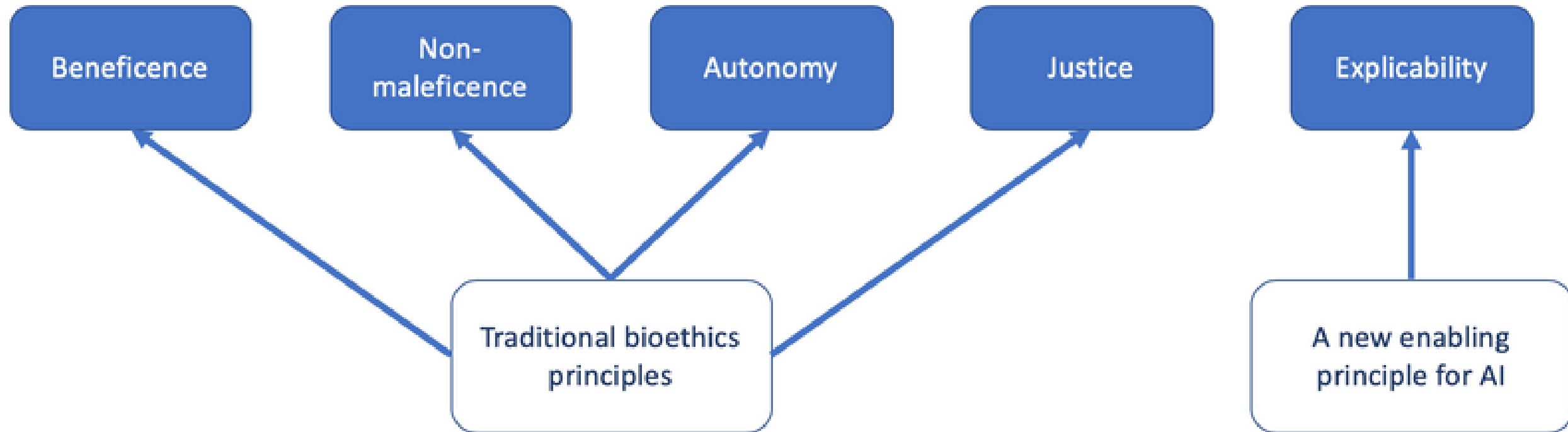# Commercial Ethics In A Nutslide

Principlism

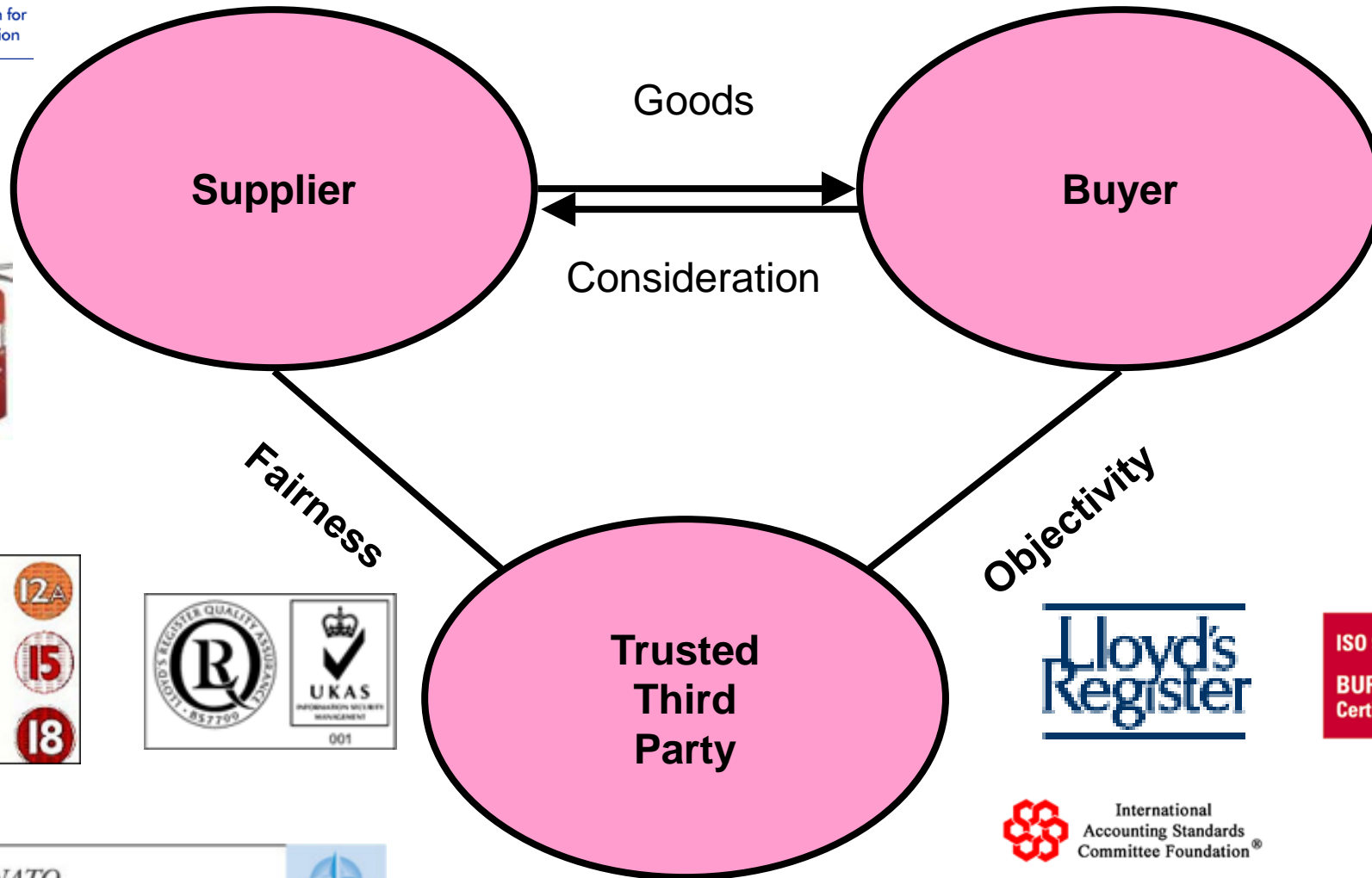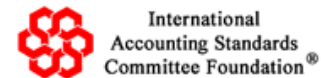|	
Autonomy	Respect for persons. Humans are a means unto themselves and not a means to an end. This encompasses the right to be free to make choices about your body.
Beneficence	Obligation to contribute to person's welfare. Interventions and provisions should provide benefit directly to the patient. 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.

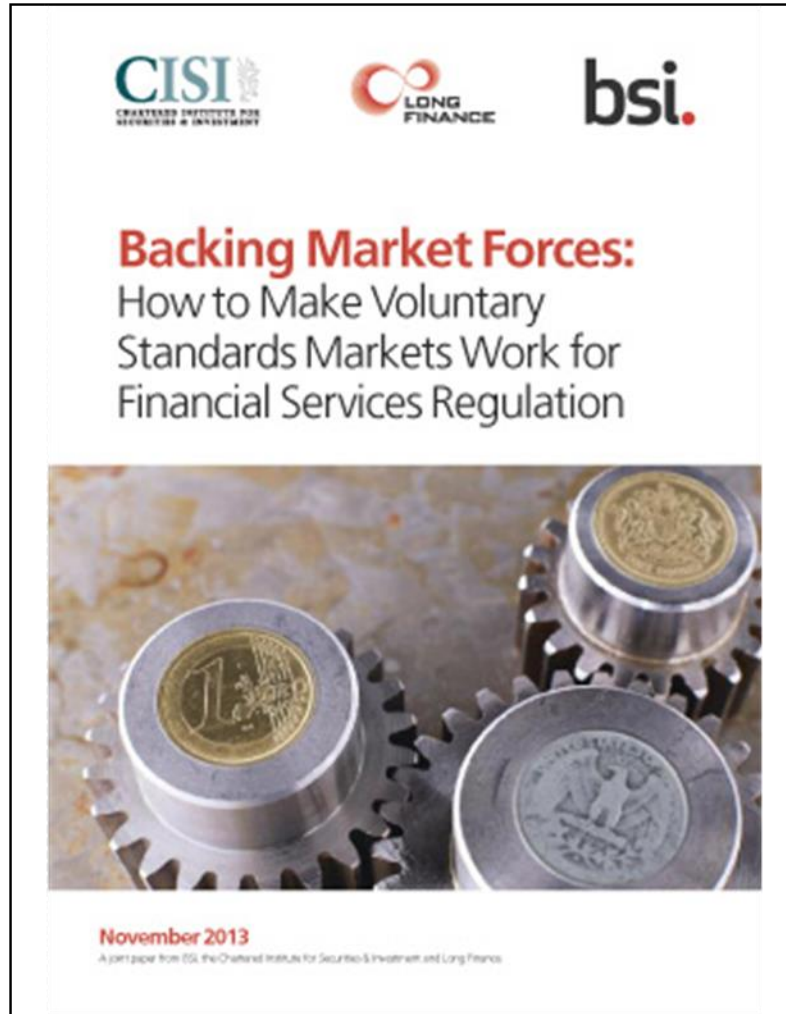**"Get a detailed grip on the big picture."**
Chao Kli Ning

Plus Ones

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

Standards Markets – The Third Way - Trusted Third Parties

Backing Market Forces

- Assessment of the potential value of using VSMs for FS regulation

- VSM could play a much larger role in FS, on their own or combined with regulation

- General awareness of standards, but lower application in FS

- Regulators neutral, but reactive if industry isn't prepared
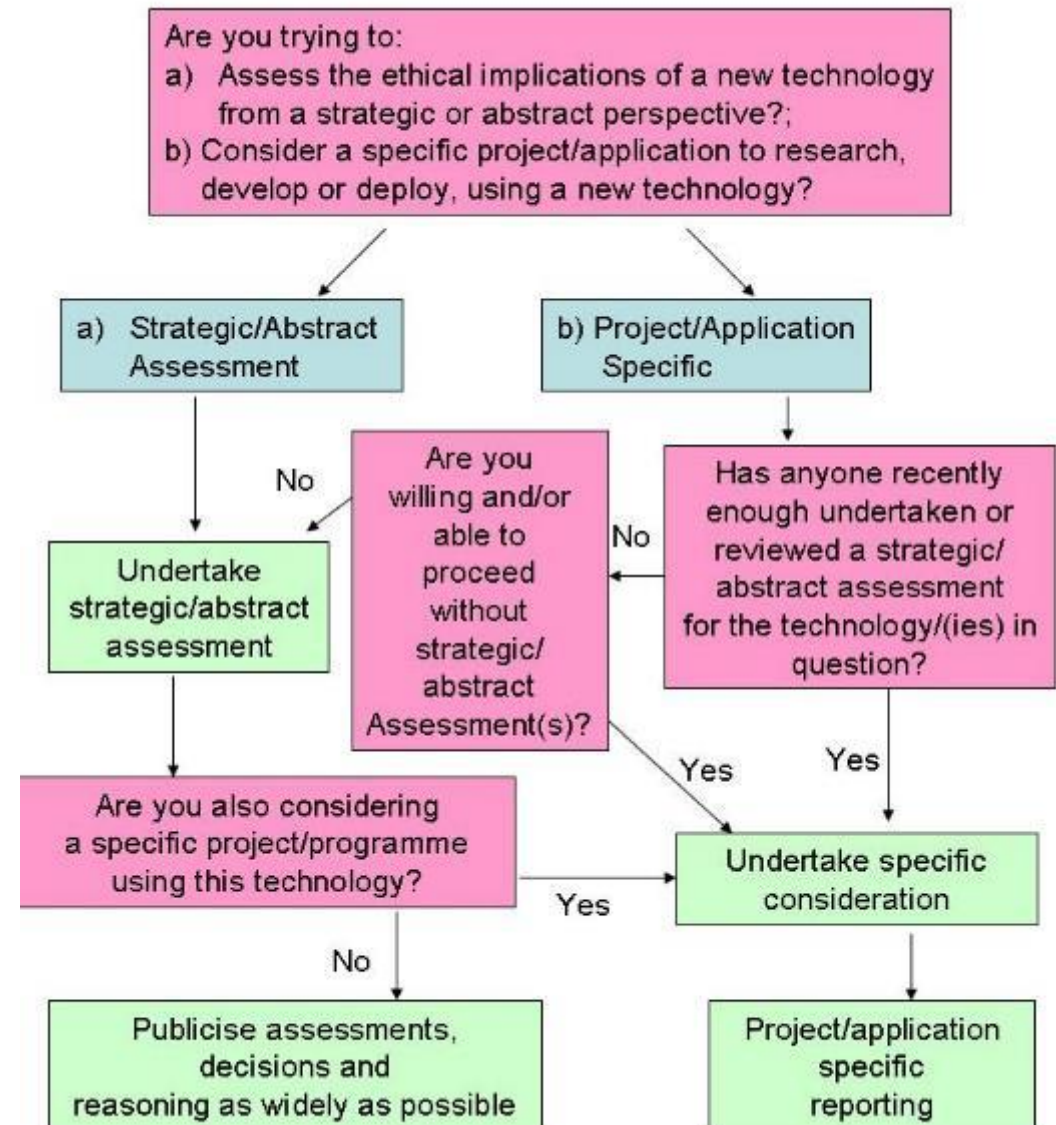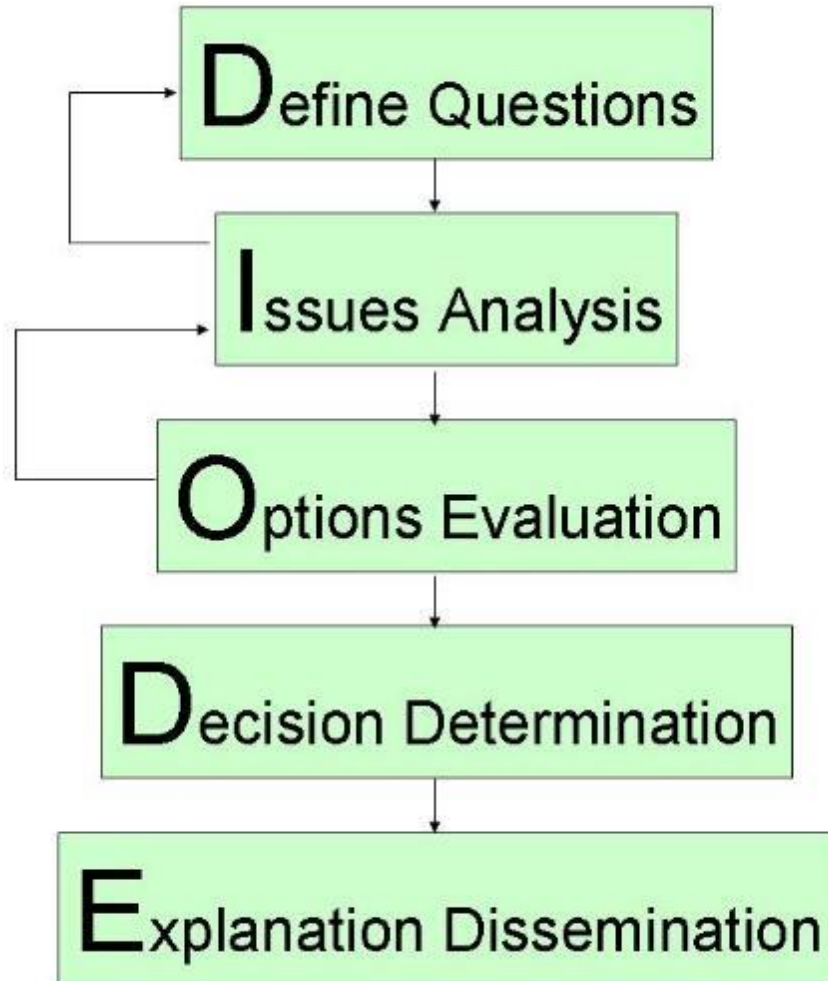
695th Lord Mayor's Ethical AI Initiative

- 'Students' to start in September 2023
- Initiative formal launch in November
- Course completion and certificate award in Q2 2024
- Second stage to cover additional application domains, e.g. law, pharma, …
- Two types of qualifications: AI Deployers and AI Builders
- CISI and BCS as certificate providers
- Proposed course outline for Deployers
- Exploration of AI Builder course development

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

Ethical Decision Making - DIODE

© Z/Yen Group, 2023

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

When Would We Our Know Commerce Is Working?

THE PRICE OF FISH

A New Approach to Wicked Economics and Better Decisions

MICHAEL MAINELLI AND IAN HARRIS

Thank you!

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

LONG FINANCE