AI and Robotics:
The Threats, and A Reconception

Roger Clarke
Xamax Consultancy Pty Ltd, Canberra
Visiting Professor in Technology & Law, UNSW
Visiting Professor in Computer Science, ANU

http://www.rogerclarke.com/EC/AITR.html (Text)
http://www.rogerclarke.com/EC/AITR.pdf (Slides)

AI, Law & Society – 15 May 2024
ANU College of Law

The Original Conception of Artificial Intelligence (AI Old)

- Based on "the conjecture that every aspect of **learning or any other feature of intelligence** can in principle be so precisely described that a machine can be made to simulate it"
- "The hypothesis is that a physical symbol system [of a particular kind] has the necessary and sufficient means for **general intelligent action**"

McCarthy et al. (1955)
Simon (1958, 1969, 1975; 1996, p.23)

The Original Conception of Artificial Intelligence (AI Old)

- Based on "the **conjecture** that every aspect of learning or any other feature of intelligence **can in principle** be so precisely described that a machine can be made to **simulate** it"
- "The **hypothesis** is that a physical symbol system [of a particular kind] has the necessary and sufficient means for **gen**eral intelligent action"

McCarthy et al. (1955)
Simon (1958, 1969, 1975; 1996, p.23)

From Conjecture and Hypothesis To Belief

"Within the very near future - much less than twenty-five years - **we shall have the technical capability of substituting machines for any and all human functions in organisations**.

"**Duplicating problem-solving and information-handling capabilities of the brain is not far off** ... surprising if it were not accomplished within the next decade" (1960)

"By the end of the 2020s [computers **will have**] **intelligence indistinguishable to biological humans**" (2005)

Simon (1960, et seq.)
Kurzweil (2005, p.25)

From Conjecture and Hypothesis To Belief

"**Within** the very near future - **much less than twenty-five years** - we shall have the technical capability of substituting machines for any and all human functions in organisations.

"Duplicating problem-solving and information-handling capabilities of the brain is not far off ... surprising if it were not accomplished **within the next decade**" (**1960**)

"**By the end of the 2020s** [computers will have] intelligence indistinguishable to biological humans" (**2005**)

Simon (1960, et seq.)
Kurzweil (2005, p.25)

From Conjecture and Hypothesis To Belief

"Within the very near future - much less than twenty-five years - we shall have the technical capability of **substituting machines for any and all human functions** in organisations.

"**Duplicating problem-solving and information-handling capabilities of the brain** is not far off ... surprising if it were not accomplished within the next decade" (1960)

"By the end of the 2020s [computers will have] **intelligence indistinguishable to biological humans**" (2005)

Simon (1960, et seq.)
Kurzweil (2005, p.25)

Bifurcation of the Field

- The ‘grand challenge’ aspect:
  ‘Artificial general intelligence' or ‘Strong AI’
  Aspiration to replicate human intelligence
- Human intelligence as Inspiration
  ‘Weak AI’ / ‘Narrow AI’

Separation
But Not Divorce

How to Recognise 'an AI'

Intelligence is exhibited by an artefact if it:

*(1) evidences **perception and cognition** of relevant aspects of its environment*

*(2) has **goals**; and*

*(3) **formulates actions** towards the achievement of those goals*

and?

*(4) **implements those actions***

esp. Albus 1991, Russell & Norvig 2009, McCarthy 2007

Recent Drift in Use of the Term 'AI'

- **AI** is the discipline of research and development of mechanisms and applications of AI systems
- **An AI System** is an engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives

ISO/IEC 22989:2022 Information Technology, Artificial Intelligence concepts and terminology

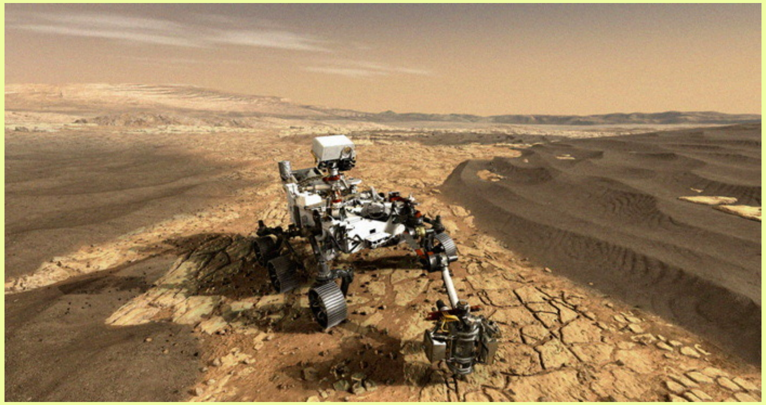
Embodiments of AI

- **Computers**
- **Robots**
  'A Computer that Does' &
  'A Machine that Computes'
- **Humanoid Robots**
  Androids
  Gynoids / Fembots
- **Vehicles**
  Terrestrial
  – Road, Rail, Off-Road
  Airborne
  Water-borne, Submerged
- **Bus-Stops**
  And other everyday Things
- **Cyborgs**

  A Human whose natural capabilities have been enhanced by technological means

  A Hybrid of a human and one or more associated, attached or embedded artefacts

'Terrestrial', Off-Road, Remote



Aug 2021 – https://futurism.com/the-byte/nasas-mars-rover-took-selfie-beautiful

Mechanical Performance of such Challenging Physical Tasks is GOOD

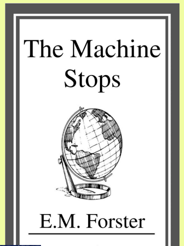
Mechanical Performance of such Challenging Physical Tasks is GOOD

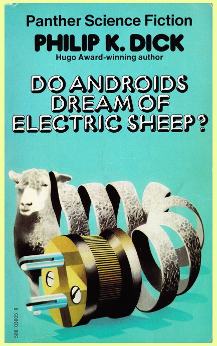
But Intelligence also requires Second-Order Intellect or Insight

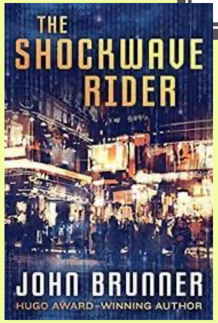
- Values-Driven Formulation of Goals
- Common-Sense Understanding of Context
- Detection of Changes of Relevance
- Ongoing Re-Evaluation of Values
- Ongoing Adaptation of Goals

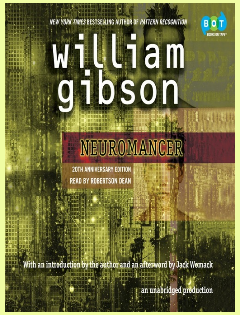
Dreyfus H.L. (1972)
Weizenbaum J. (1976)

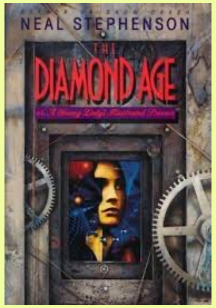
Science Fiction Anticipates Reality




Panther Science Fiction
PHILIP K. DICK
Hugo Award-winning author
DO ANDROIDS DREAM OF ELECTRIC SHEEP?







AI Sceptics are in Good Company

A Distillation of the Threats Inherent in AI

1. **Artefact Autonomy**
   Substantial delegation from humans to non-humans
2. **Inappropriate Assumptions about Data**
   Data selectivity, interpolation, incompatibility, quality
3. **... and about the Inferencing Process**
   Uncontrolled environments, unmodelled systems
4. **Opaqueness of the Inferencing Process**
   Unexplainability, procedural fairness, unaccountability
5. **Irresponsibility**
   Everyone in the chain points at everyone else

https://www.rogerclarke.com/EC/AII.html#Th

Degrees of Autonomy

		Function of the Artefact	Function of the Human
	0	~~NIL~~	Analyse, Decide, Act
Decision Support System	1	Analyse Options	Analyse, Decide, Act
	2	Advise re Options	Analyse, Decide, Act
	3	Recommend Act	Analyse, Approve/Reject Act
Decision System	4	Notify Impending Act	Override/Veto Impending Act
	5	Act and Inform	Interrupt/Suspend/Cancel an Act
	6	Act	~~NIL~~

After Armstrong (2010, p.14),
Sheridan & Verplank (1978, Table 8.2, pp. 8-17-8.19)
as interpreted by Robertson et al. (2019, Table 1)

The Threats Inherent in AI

1. **Artefact Autonomy**
   Substantial delegation from humans to non-humans
2. **Inappropriate Assumptions about Data**
   Data selectivity, interpolation, incompatibility, quality
3. **... and about the Inferencing Process**
   Uncontrolled environments, unmodelled systems
4. **Opaqueness of the Inferencing Process**
   Unexplainability, procedural fairness, unaccountability
5. **Irresponsibility**
   Everyone in the chain points at everyone else

https://www.rogerclarke.com/EC/AII.html#Th

Data & Information Quality Factors

Assessable at time of collection

D1 – Syntactic Validity
D2 – Appropriate (Id)entity Association
D3 – Appropriate Attribute Association
D4 – Appropriate Attribute Signification
D5 – Accuracy
D6 – Precision
D7 – Temporal Applicability

Assessable only at time of use

I1 – Theoretical Relevance
I2 – Practical Relevance
I3 – Currency
I4 – Completeness
I5 – Controls
I6 – Auditability

http://www.rogerclarke.com/EC/BDBR.html#Tab1

The Threats Inherent in AI

1. **Artefact Autonomy**
   Substantial delegation from humans to non-humans
2. **Inappropriate Assumptions about Data**
   Data selectivity, interpolation, incompatibility, quality
3. **... about the Inferencing Process**
   Uncontrolled environments, unmodelled systems
4. **Opaqueness of the Inferencing Process**
   Unexplainability, procedural fairness, unaccountability
5. **Irresponsibility**
   Everyone in the chain / network points at everyone else

https://www.rogerclarke.com/EC/AII.html#Th

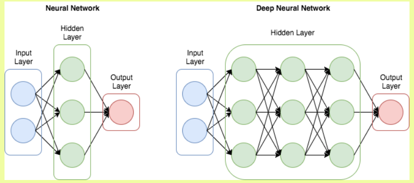
AI / ML/ ANNs

- Machine Learning (ML) is a major branch of AI
- The (currently) dominant technique is **'artificial neural networks' (ANN)**
- ANNs date to 1957, with a surge in the 1980s
- It's been **resurgent since the 2010s because** ...
- ... **Sufficiently powerful processors** (highly-parallel architectures for graphics processing) coincided with **a rash of 'big data' lying around**

AI / ML/ ANNs

- Machine Learning (ML) is a major branch of AI
- The (currently) dominant technique is **'artificial neural networks' (ANN)**
- ANNs date to 1957, with a surge in the 1980s
- It's been **resurgent since the 2010s because** ...
- ... **Sufficiently powerful processors** (highly-parallel architectures for graphics processing) coincided with **a rash of 'big data' lying around**
- **A (Dangerous) Over-Simplification**:
  'Feed an ANN a big, big set of pictures of cats, and it learns to recognise cats'

Neural Nets



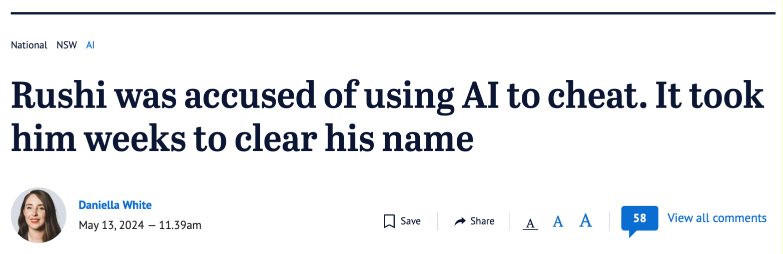
- A set of **connected nodes**, each with an associated **weight**
- Each node performs computations based on incoming data, may adapt the weights, and may pass output to one or more other nodes
- A neural net has to be 'trained'.

  This requires a **learning algorithm**, **operating on a training-set of instances**, to establish initial weights on each connection
- **Later instances are categorised** based on the weights associated with the connections at the time

https://arstechnica.com/science/2019/12/how-neural-networks-work-and-why-theyve-become-a-big-business/

Assumptions Commonly Implicit in AI/ML

- An underlying model of reality
- Near-enough correspondence with reality
- Adequate training-set quality
- Adequate data-item quality
- Adequate data-item correspondence to the phenomenon it purports to represent
- No material training-set bias
- No learning algorithm bias
- Compatibility of data and 'model'
- Logically valid inferences
- Empirically checked inferences



National NSW AI

Rushi was accused of using AI to cheat. It took him weeks to clear his name

Daniella White
May 13, 2024 – 11.39am

Of the more than 300 AI-related instances of suspected cheating identified at Sydney University last year, almost 30 percent were later cleared of wrongdoing.

Major Risk Factors in AI/ML

- **Insufficient, active and careful modelling**
  of real-world problem-solutions, problems, or problem-domains

  cf. lists of input and output variables, (plus intermediating/hidden variables, if ‘deep’)

  cf. implicit variables (‘unsupervised’ ML)
- **No explicit, designed-in real-world relationship**
  And/or inadequate audit of the relationship
- **Loss of the Theory-Empiricism partnership**
  i.e. Empiricism dominates, even replaces, Theory

Socio-Political Impacts and Implications

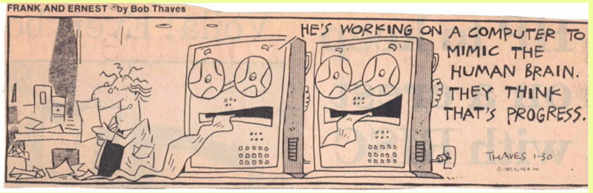
- ***De Facto* Delegation**
  'The computer says no'
- **Unexplainability**
  Accountability Undermined
- **Unfair Decisions, Actions**
  Discriminatory Behaviour
- **Economic, Social Scoring**
  Non-Conformist Victimisation
- **Undefendable Accusations**
  Power, Information Asymmetry
- **'Predestination’**
  e.g. Predictive Policing
- **People-Replacement**
  Effect on Income Distribution
- **Denial of Services, of Movement, of Identity**
  Public Resentment, Violence

‘Artificial’? Or ‘Artefactual’? ‘Intelligence’

‘Artificial’? Or ‘Artefactual’? ‘Intelligence’ What Do We Want From It?

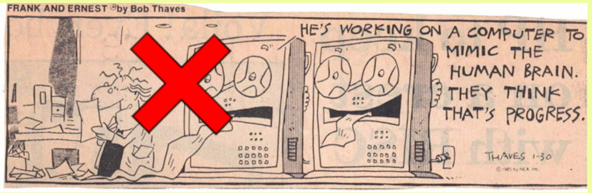
- There are 8 billion people and we’re multiplying (too) fast
- Why would we want yet more Natural Intelligence?

‘Artificial’? Or ‘Artefactual’? ‘Intelligence’ What Do We Want From It?


FRANK AND ERNEST ©by Bob Thaves
HE'S WORKING ON A COMPUTER TO MIMIC THE HUMAN BRAIN. THEY THINK THAT'S PROGRESS.
THAVES 1-30

https://www.frankandernest.com/search/index.php?iid=71150

‘Artificial’? Or ‘Artefactual’? ‘Intelligence’ What Do We Want From It?


FRANK AND ERNEST ©by Bob Thaves
HE'S WORKING ON A COMPUTER TO MIMIC THE HUMAN BRAIN. THEY THINK THAT'S PROGRESS.
THAVES 1-30

- Do things well that humans do poorly, or cannot do at all
- Perform functions within systems that include both humans and artefacts
- Interface effectively, efficiently and adaptably with both humans and other artefacts

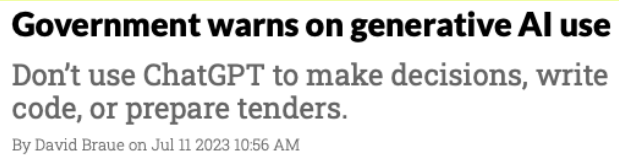
https://www.frankandernest.com/search/index.php?iid=71150

ChatGPT / LLM’s Achilles Heel

- Unsceptical and unbridled enthusiasm was quickly followed by recriminations:
  - Gamma testers conducted serious testing
  - Students submitted mistaken assignments
  - Journals required declarations of ‘no LLM’
  - Lawyers submitted briefs with invented cases
  - ARC Assessors submitted facile reports

ChatGPT / LLM's Achilles Heel

- Unsceptical and unbridled enthusiasm was quickly followed by recriminations:
  - Gamma testers conducted serious testing
  - Students submitted mistaken assignments
  - Journals required declarations of 'no LLM'
  - Lawyers submitted briefs with invented cases
  - ARC Assessors submitted facile reports



Government warns on generative AI use

Don't use ChatGPT to make decisions, write code, or prepare tenders.

By David Braue on Jul 11 2023 10:56 AM

ChatGPT / LLM's Achilles Heel

- Unsceptical and unbridled enthusiasm was quickly followed by recriminations:
  - Gamma testers conducted serious testing
  - Students submitted mistaken assignments
  - Journals required declarations of 'no LLM'
  - Lawyers submitted briefs with invented cases
  - ARC Assessors submitted facile reports
  - Aust Govt places tight limits on its use
- It was designed as a Decision Tool
- It should be designed as a Decision Support Tool

Human Intelligence ⇨ Augmented Intelligence

Augmented Intelligence

- Ashby (1956) on 'intelligence amplification'
- Engelbart (1962) on 'augmenting human intellect'
- Mann (2001) on wearable/body-borne computing, augmented / diminished / mediated reality, sur- / sous- / meta- / equi-veillance, ...
- Araya (2019) on 'augmented intelligence' as "an alternative conceptualization of AI that focuses on its assistive role in advancing human capabilities"
- IEEE Council on Extended Intelligence (2017-19) "it is not AI in isolation, but the social, economic, political, and cultural systems within which these tools are integrated that must be addressed to avoid reductionist outcomes"

http://www.rogerclarke.com/EC/AIYV.html#RTFToC16

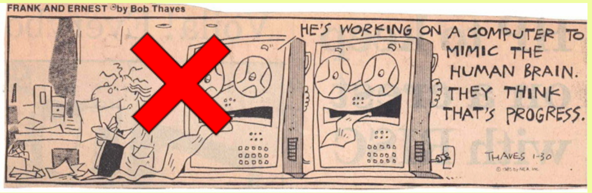
Human Intelligence & ‘Artificial Intelligence’ ??? ⇨ Augmented Intelligence

Human Intelligence & Artefactual ⇨ Augmented Intelligence

Human Intelligence & Artefactual Intellectics ⇨ Augmented Intelligence

Human Intelligence & Complementary Artefactual Intellectics ⇨ Augmented Intelligence

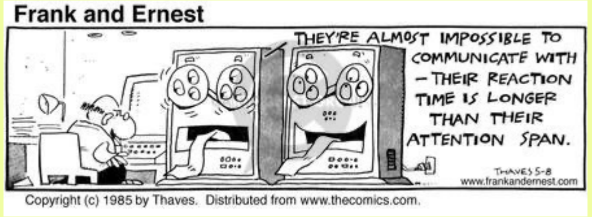
Complementary Artefactual Intellectics
What Do We Want From It?


FRANK AND ERNEST ©by Bob Thaves
HE'S WORKING ON A COMPUTER TO MIMIC THE HUMAN BRAIN. THEY THINK THAT'S PROGRESS.
THAVES 1-30

Do things well that humans
do poorly, or cannot do at all:

- Dull
- Dirty
- Dangerous

Complementary Artefactual Intellectics
What Do We Want From It?


Frank and Ernest
THEY'RE ALMOST IMPOSSIBLE TO COMMUNICATE WITH – THEIR REACTION TIME IS LONGER THAN THEIR ATTENTION SPAN.
THAVES 5-8
www.frankandernest.com

Copyright (c) 1985 by Thaves. Distributed from www.thecomics.com.

Do things well that humans
do poorly, or cannot do at all:

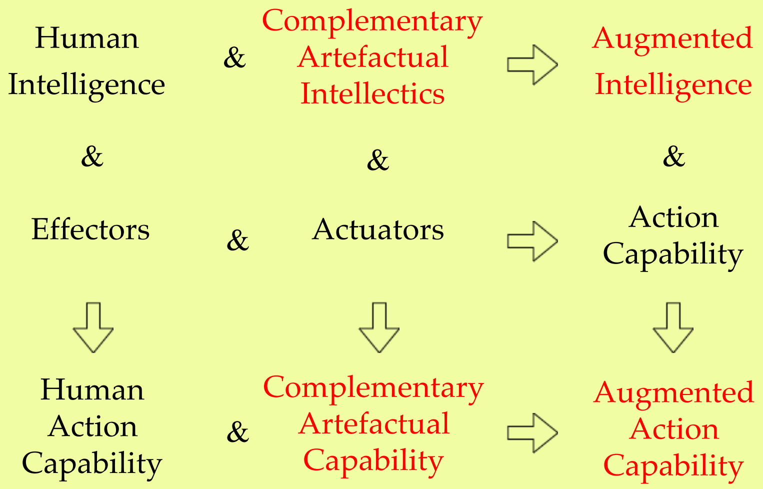
- Dull
- Dirty
- Dangerous
- Precision
- Speed

https://www.frankandernest.com/search/index.php?iid=71488

Effectors & Actuators ⇨ Capability

Human Intelligence & Complementary Artefactual Intellectics ⇨ Augmented Intelligence

Effectors & Actuators ⇨ Capability



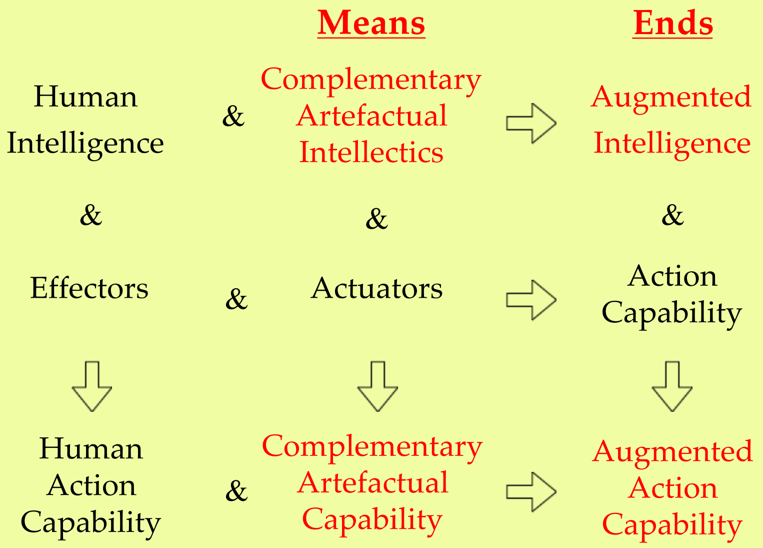
Human Intelligence & Complementary Artefactual Intellectics ⇨ Augmented Intelligence

& & &

Effectors & Actuators ⇨ Action Capability

⇩ ⇩ ⇩

Human Action Capability & Complementary Artefactual Capability ⇨ Augmented Action Capability



Means **Ends**

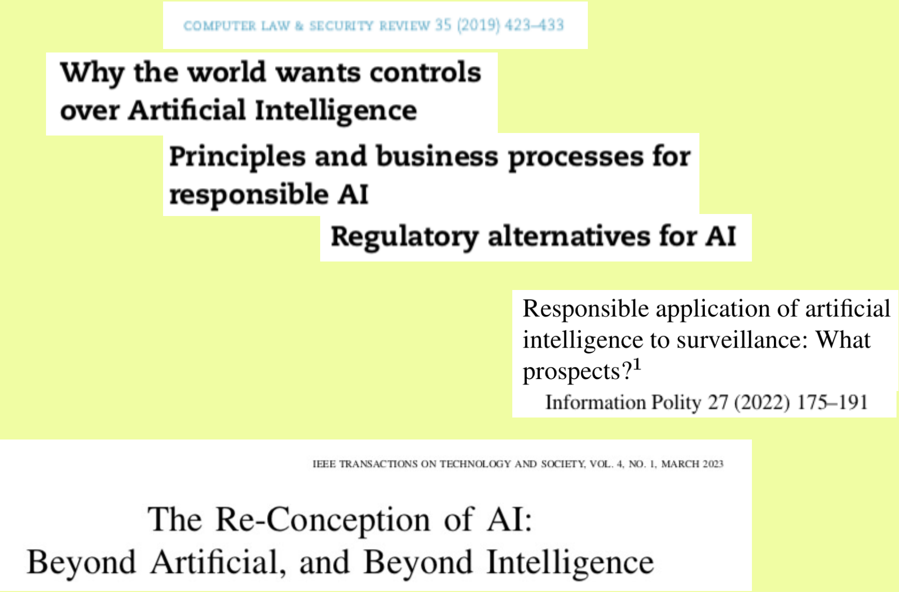
Human Intelligence & Complementary Artefactual Intellectics ⇨ Augmented Intelligence

& & &

Effectors & Actuators ⇨ Action Capability

⇩ ⇩ ⇩

Human Action Capability & Complementary Artefactual Capability ⇨ Augmented Action Capability



COMPUTER LAW & SECURITY REVIEW 35 (2019) 423–433

Why the world wants controls over Artificial Intelligence

Principles and business processes for responsible AI

Regulatory alternatives for AI

Responsible application of artificial intelligence to surveillance: What prospects?[1]

Information Polity 27 (2022) 175–191

IEEE TRANSACTIONS ON TECHNOLOGY AND SOCIETY, VOL. 4, NO. 1, MARCH 2023

The Re-Conception of AI:
Beyond Artificial, and Beyond Intelligence

AI and Robotics:
The Threats, and A Reconception

Roger Clarke
Xamax Consultancy Pty Ltd, Canberra
Visiting Professor in Technology & Law, UNSW
Visiting Professor in Computer Science, ANU

http://www.rogerclarke.com/EC/AITR.html (Text)
http://www.rogerclarke.com/EC/AITR.pdf (Slides)

AI, Law & Society – 15 May 2024
ANU College of Law