The Re-Conception of AI and Robotics as Complementary Artefact Intelligence and Augmented Capability

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

http://www.rogerclarke.com/EC/AITS {.html, .pdf}

ANU AI, ML and Friends Seminar – 13 July 2023

COMPUTER LAW & SECURITY REVIEW 30 (2014) 230–246
Available online at www.sciencedirect.com
ScienceDirect
www.compseconline.com/publications/prodclaw.htm
ELSEVIER
Computer Law & Security Review

Understanding the drone epidemic

What drones inherit from their ancestors

The regulation of civilian drones' impacts on public safety

The regulation of civilian drones' impacts on behavioural privacy

Since 2014, >700 citations

Big Data's Big Unintended Consequences

Marcus R. Wigan, *Oxford Systematics, Swinburne University, and the University of Melbourne*
Roger Clarke, *Xamax Consultancy, University of New South Wales, and Australian National University*

COMPUTER Published by the IEEE Computer Society 0018-9162/13/$31.00 © 2013 IEEE

Big data, big risks

Info Systems J (2016) **26**, 77–90

COMPUTER LAW & SECURITY REVIEW 34 (2018) 467–476
Available online at www.sciencedirect.com
ScienceDirect
www.compseconline.com/publications/prodclaw.htm
ELSEVIER
Computer Law & Security Review

Guidelines for the responsible application of data analytics

31ST BLED ECONFERENCE: DIGITAL TRANSFORMATION: MEETING THE CHALLENGES
JUNE 17 - 20, 2018, BLED, SLOVENIA, CONFERENCE PROCEEDINGS
A. Pucihar, M. Kljajić Borštnar, P. Ravesteijn, J. Seitz & R. Bons

Towards Responsible Data Analytics: A Process Approach

ROGER CLARKE & KERRY TAYLOR

Do Ethical Guidelines have a Role to Play in Relation to Data Analytics and AI/ML?
For AiCE 2020, UniSA, Adelaide, November 2020

Since 2013/16, >500 citations

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

Since 2019, ~200 citations

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 **general 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)

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

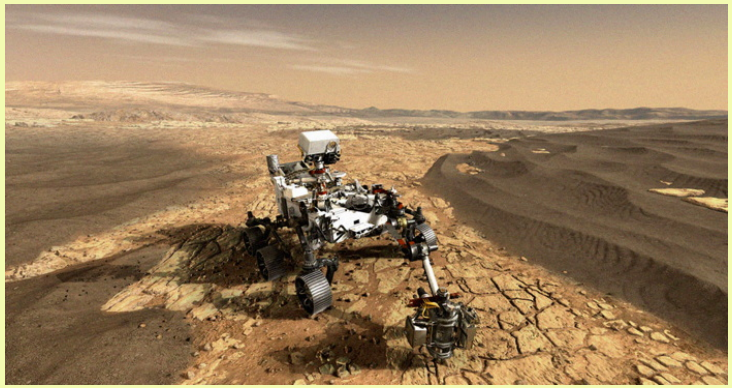
Embodiments of AI

- **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 Difficult Physical Tasks is GOOD

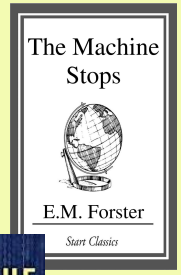
Mechanical Performance of Difficult Physical Tasks is GOOD

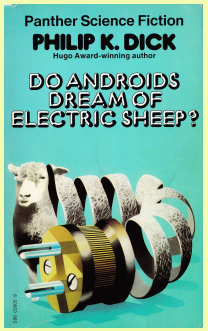
But Intelligence also requires Second-Order Intellect or Insight

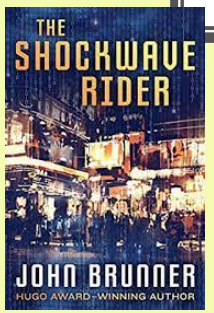
- Values-Driven Formulation of Goals
- Common-Sense Understanding of Context
- Detection of Changes of Relevance
- Ongoing Re-Evaluation of Values
- Ongoing Adaptation of Goals

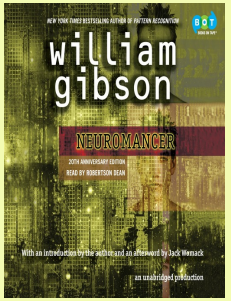
Dreyfus H.L. (1972)
Weizenbaum J. (1976)

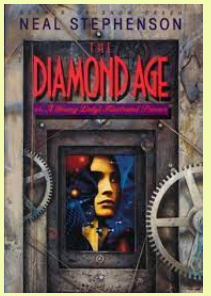
Science Fiction Anticipates Reality











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~~

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

Data 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

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 points at everyone else

Assumptions Often 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

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 may dominate Theory

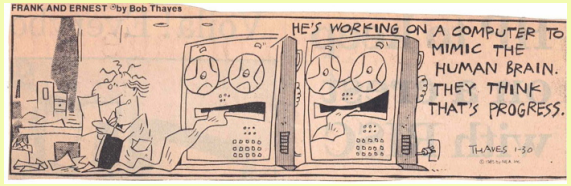
Social Impacts and Implications

- ***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
- **Denial of Services, of Movement, of Identity**
  Public Resentment, Violence
- **'Predestination'**
  Predictive Policing
- **People-Replacement**
  Effect on Income Distribution

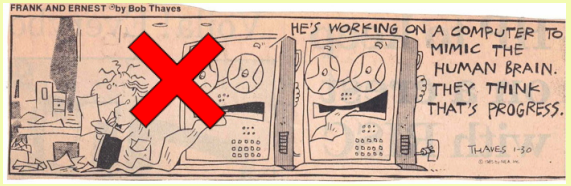
Artificial'? Or 'Artefact'? Intelligence
What Do We Want From It?

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

Artificial'? Or 'Artefact'? Intelligence
What Do We Want From It?



Artificial'? Or 'Artefact'? Intelligence
What Do We Want From It?



- Do things well that humans do poorly, or cannot do at all: Dull, Dirty, Dangerous / Precise, Fast, ...
- Perform functions within systems that include both humans and artefacts
- Interface effectively, efficiently and adaptably, with both humans and other artefacts

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 & **Artefact** 'Intelligence' AI New #1 ⇨ **Augmented** Intelligence AI New #2

'Augmented Intelligence' isn't a new idea

- Ashby (1956) on 'intelligence amplification'
- Engelbart (1962) on 'augmenting human intellect'
- Mann (2001) on 'wearable computing'
- Araya (2019) on 'augmented intelligence' as "an alternative conceptualization of AI that focuses on its assistive role in advancing human capabilities"
- IEEE-DR (2019) on 'symbiotic autonomous systems' (But it treats artefacts as equals with humans, and expressly adopts the mystical transhumanism and posthumanism notions – postulating the emergence of a new species via technology rather than genetics)

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

Human Intelligence & **'Artificial' 'Intelligence'** i.e. AI Old?? ⇨ **Augmented** Intelligence AI New??

Artefact 'Intelligence' has to be Complementary

Effectors & Actuators ⇨ Capability

Human Intelligence & Artefact 'Intelligence' ⇨ Augmented Intelligence

&

Effectors & Actuators ⇨ Capability

Human Intelligence & Artefact 'Intelligence' ⇨ Augmented Intelligence

&

Effectors & Actuators ⇨ Capability

⇩ ⇩ ⇩

Human Capability & **Artefact Capability** ⇨ **Augmented Capability**

Human Intelligence & Artefact 'Intelligence' ⇨ Augmented Intelligence

&

Effectors & Actuators ⇨ Capability

⇩ ⇩ ⇩

Human Capability & **Artefact Capability** ⇨ **Augmented Capability**

Which of these is/are Robotics ?
And 'AI-based autonomous systems'?

		The Means		The End
Human Intelligence	&	Artefact 'Intelligence'	⇨	Augmented Intelligence
&		&		&
Effectors	&	Actuators	⇨	Capability
⇩		⇩		⇩
Human Capability	&	**Artefact Capability**	⇨	**Augmented Capability**

AI, ML and Friends?

Friends
of
Complementary
Artefact
Capabilities
and
Intelligence
Augmentation

AI, ML and Friends?

Friends
of
Complementary
Artefact
Capabilities
and
Intelligence
Augmentation

FOCACIA

The Re-Conception of AI and Robotics as Complementary Artefact Intelligence and Augmented Capability

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

http://www.rogerclarke.com/EC/AITS {.html, .pdf}

ANU AI, ML and Friends Seminar – 13 July 2023