How is Society to Risk-Manage the Threats in AI Applications to Cybersecurity?

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

http://www.rogerclarke.com/EC/SINS21{.html, .pdf}

14th Wksp – Social Implications of National Security
18 Feb 2021

Artificial Intelligence (AI)

- "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**"
- Successions of modest progress, excessive enthusiasm, failure, and 'AI winters' when lack of credibility resulted in limited funding
- Current Foci:
  - Rule-Based Expert Systems, Robotics, Cyborgisation, **AI/ML/ANNs**

http://www.rogerclarke.com/EC/AII.html#AI

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 in the 2010s because ...
- ... Sufficiently powerful processors (highly-parallel architectures for graphics processing) coincided with a rash of 'big data' lying around
- Over-simplification: 'Feed an ANN a big, big set of pictures of cats, and it learns to recognise cats'

Risk Factors in AI/ML/ANNs

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

  Mere lists of input and output variables, and, possibly, intermediating/hidden variables
- **Any relationship to the real world is implicit** rather than being designed-in

  The degree of relationship is seldom audited
- The Theory-Empiricism partnership is lost, with **Empiricism dominating Theory**

Assumptions Implicit in AI/ML/ANNs

- Close model 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

'If you torture data long enough
it will confess to anything'

attr. Ronald Coase (1981)
"How should economists choose?" Warren Nutter Lecture
orig. Darrell Huff (1954) 'How to Lie With Statistics'

AI embodies errors of inference, decision and action arising from the independent operation of artefacts, for which **no rational explanation is available**, which results in inferences, decisions and actions **incapable of investigation, correction and reparation**

A Summary of the Sources of AI's Threats

1. Artefact Autonomy
2. Inappropriate Assumptions ... about Data
3. ... and about the Inferencing Process
4. Opaqueness of the Inferencing Process
5. Irresponsibility

http://rogerclarke.com/EC/AII.html#Th

10 Groups of Principles for Responsible AI

1. Evaluate Positive and Negative Impacts
2. Complement Humans
3. Ensure Human Control
4. Ensure Human Wellbeing and Safety
5. Ensure Consistency with Human Values and Human Rights
6. Deliver Transparency and Auditability
7. Embed Quality Assurance
8. Exhibit Robustness and Resilience
9. Ensure Accountability for Legal and Moral Obligations
10. Enforce, and Accept Enforcement of, Liabilities and Sanctions

http://rogerclarke.com/EC/PRAI.html#Princ

50 Principles for Responsible AI
1. Evaluate Positive and Negative Impacts

1.1 Conceive and design only after ensuring adequate **understanding** of purposes and contexts (E4.3, P5.3, P6.21, P7.1, P15.7, P17.5)

1.2 **Justify objectives** (E3.25)

1.3 Demonstrate the **achievability of postulated benefits** (Pre-condition)

1.4 Conduct **impact assessment** (E7.1, P3.12, P4.1, P4.2, P6.21, P11.8, P17.5)

1.5 Publish **sufficient information to stakeholders** to enable them to conduct impact assessment (E7.3, P3.7, P4.1, P8.3, P8.4, P8.7)

1.6 Conduct **consultation with stakeholders** and enable their participation (E5.2, E7.2, E8.3, P3.7, P8.6, P8.7, P11.8)

1.7 **Reflect stakeholders' justified concerns** (E5.2, E8.3, P3.7, P11.8)

1.8 Justify negative impacts on individuals ('**proportionality**') (E3.21, E7.4, E7.5)

1.9 **Consider less harmful ways** of achieving the same objectives (E3.22)

http://rogerclarke.com/EC/GAIF.html#App1

The Innovation Mantra / Tech Solutionism is Trumping the Precautionary Principle

Enthusiastic marketing means Tech Determinism wins, with risks borne by user-organisations and the public

This is the converse of the 'Precautionary Principle':

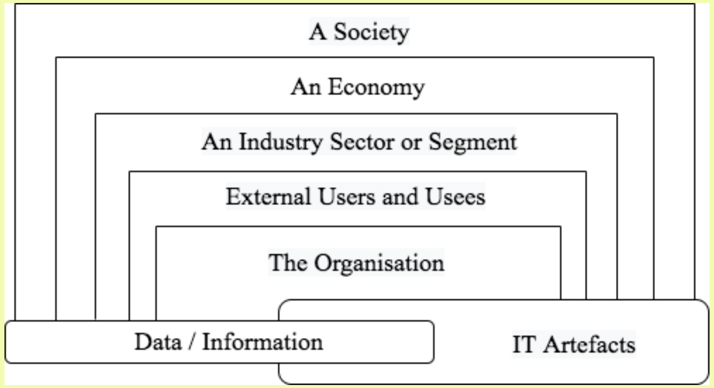
If an action or policy is suspected of causing harm, and scientific consensus that it is not harmful is lacking, then:

Weak Form:
The burden of proof falls on those taking the action

Strong Form:
Actions must be taken to avoid or diminish potential harm

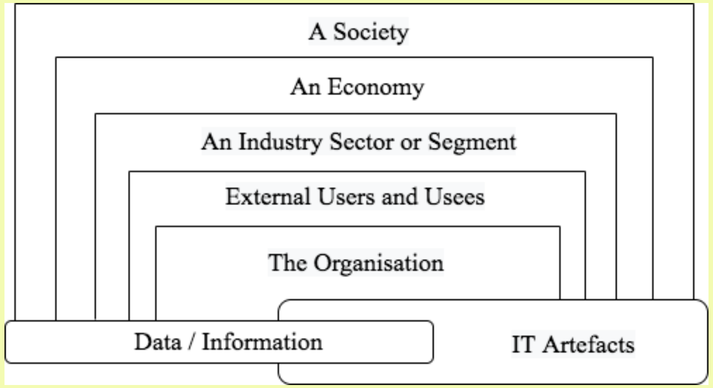
http://www.austlii.edu.au/au/cases/nsw/NSWLEC/2006/133.html

Alternative Scope Definitions for Security Analysis



http://rogerclarke.com/EC/WS-1301.html

Where is 'Cybersecurity'? Where is 'National Security'?



http://rogerclarke.com/EC/WS-1301.html

Cybersecurity – Sample Definitions

- The protection of internet-connected systems such as **hardware, software and data** from cyberthreats (Techtarget – semi-circular, hence unclear)
- The practice of protecting **systems, networks, and programs** from digital attacks (Cisco, Kaspersky – omits environmental and accidental threats)
- Protection of [as above, and] from the disruption or misdirection of the **services** that [IT] provides
- ... protection for the state of the cyber environment and its **users** (Schatz et al. 2017)

What's 'National Security'?

The protection of a nation from attack or other danger by holding adequate armed forces and guarding state secrets

Encompasses economic security, monetary security, energy security, environmental security, military security, political security and security of energy and natural resources

http://definitions.uslegal.com/n/national-security/

"specifically authorized under criteria established by an Executive order to be kept secret in the interest of national defense or foreign policy"

US Freedom of Information Act

Or is this 'National Security'?

- **Public Safety**
  Bombs in aircraft, mayhem in marketplaces,
  major events, e.g. Olympics, election rallies
- **Prominent Person Safety**
  Heads of State; Utterers of unpopular opinions;
  G8, G20, OPEC, APEC, CHOGM, ...
- **Critical Infrastructure Security**
  Military-Industrial (incl. Cryptography) Transport,
  Communications, Energy, Water
  Public Health, Emergency Services
  Law Enforcement, Agriculture, Financial Services

'Terrorism'

The use of violence or the threat of violence,
especially against civilians,
in order to alarm the public,
in the pursuit of political [or politico-religious] goals

'Terrorism' has been conflated with 'National Security'

Why is 'National Security' Exempt from Basic Democratic Principles?

Pre-Conditions

1. Evaluation
2. Consultation
3. Transparency
4. Justification

Design

5. Proportionality
6. Mitigation
7. Controls

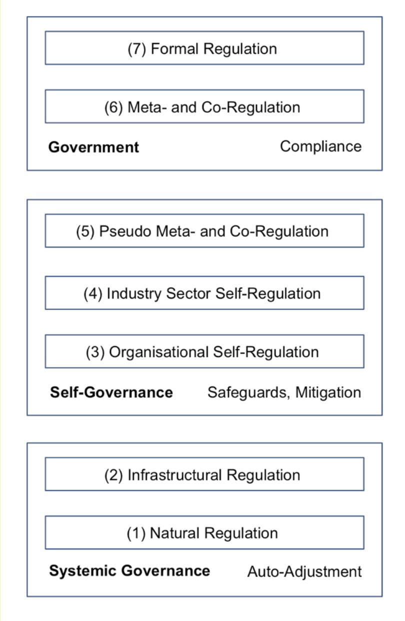
Post-Condition

8. Audit

https://privacy.org.au/policies/meta-principles/

Alternative Regulatory Forms

http://rogerclarke.com/EC/RTF.html#RL


(7) Formal Regulation
(6) Meta- and Co-Regulation
Government
Compliance
(5) Pseudo Meta- and Co-Regulation
(4) Industry Sector Self-Regulation
(3) Organisational Self-Regulation
Self-Governance
Safeguards, Mitigation
(2) Infrastructural Regulation
(1) Natural Regulation
Systemic Governance
Auto-Adjustment

Layer (6) Co-Regulation

- **Legislated Power to approve Codes**, subject to:
  - Compliance with Broad Principles
  - Primacy of Negotiated Codes
  - Fallback of Imposed Codes
- **Code Negotiation** Institution(s), Processes
- **Resources**
- **Enforcement Powers**
- **Assignment** of Enforcement Powers, Resources
- **Obligation** to apply the Powers and Resources

http://www.rogerclarke.com/EC/RAI.html#RAI

A Proposal for Protection against AI-Based Cybersecurity Protections

('Here Be Dragons', Right Now)

Recapitulation

• AI, and AI's Threats	2- 7
• 'Principles for Responsible AI'	8-10
• The Scope of Security Analysis	11-17
• Active Regulation of AI	18-19

http://www.rogerclarke.com/EC/SINS21{.html, .pdf}