# Do Ethical Guidelines have a Role to Play in Relation to Data Analytics and AI/ML?

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

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

**AiCE 2020, UniSA, Adelaide**
**Nov-Dec 2020**




## Scary AI ... Robotics ... Neural Nets ... Autonomous Decision-Making

➔ **An Explosion in Ethical Guidelines**

- **Various Publishers**, incl.
  National and Supra-National Bodies,
  Corporations and Industry Associations,
  Professional Associations,
  Public Interest Advocacy Orgs, Academics
- **Various Scope Definitions**, incl.
  AI, Robotics, Data Analytics, AI/ML




## The Necessary Conditions for Effective Ethical Guidelines

- Comprehensive
- Operationalised not aspirational
- Articulated for specific contexts
- QA before, during and after the fact
- Obligations
- Complaints channels
- Investigational powers and resources
- Meaningful sanctions
- Enforcement powers, resources, commitment




## Data Analytics

- Stat Maths, Ops Res, Mngt Science 1970s
  Data Warehousing, Data Mining 1990s
  Big Data and Data Analytics 2010s
- Volume, Velocity, Variety ... Value ... Veracity

"Massive amounts of data and applied mathematics **replace every other tool** that might be brought to bear.

"**Out with every theory** of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. ...

"Faced with massive data, [the old] approach to science -- hypothesize, model, test -- is becoming obsolete. ...
Petabytes allow us to say: **'Correlation is enough'**"

http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory

## Artificial Intelligence (AI)

- 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**"
- Successions of modest progress, excessive enthusiasm, failure, and 'AI winters' when lack of credibility resulted in limited funding
- The 'Simple Simon' postulate remains a conjecture, even after 75 years

McCarthy et al. (1955)
https://www.aaai.org/ojs/index.php/aimagazine/article/viewFile/1904/1802

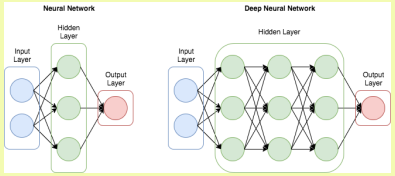



## AI / ML

- Machine Learning 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




## 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 training method / **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/




## Risk Factors in Data Analysis esp. AI/ML

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

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

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

## Assumptions Implicit in AI/ML

- 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, of decision and of action, arising from the more or less independent operation of artefacts, for which **no rational explanations are available**, and which may be **incapable of investigation, correction and reparation**

**Factors**

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


## The 'AI Principles Super-Set' Project

1. Select 30 sets of 'Ethical Guidelines'
2. Extract the exhortations they contain
3. Generate a 'super-set' of Principles
4. Cross-refer in both directions, to enable evaluation and audit
5. Evaluate each source-document against the super-set
6. Evaluate later source-documents against the super-set


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

## 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


## Evaluation of 30 Documents vs 50 Principles
## In Summary: Sparse Coverage

- Only 2 Documents score >50%
  The Draft and (the rather different) Final EC Guidelines
- 27/30 scored in the range 8-34% (i.e. Failed)
- On average, Principles were in only 6 Documents
  The only Principles in 50% of Documents were:
  Ensure people's physical health and safety (80%)
  Fairness / Impartiality (50%)
  Auditability (logging, etc.) (50%)




## Evaluation of 30 Documents vs 50 Principles
## Low Scores of Particular Concern

- Ensure people's wellbeing (Beneficence) 47%
- Ensure that effective remedies exist ... 47%
- **Ensure human control over AI ...** **43%**
- **Conduct impact assessment ...** **33%**
- **Ensure human control over AI autonomy** **23%**
- **Access to humanly-understandable explanations** **23%**
- Reflect stakeholders' concerns in the design 17%
- Test result validity 17%
- **Justify negative impacts (Proportionality)** **13%**
- **Ensure human review before action** **10%**

## The 'AI Principles Super-Set' Project

1. Select 30 sets of 'Ethical Guidelines'
2. Extract the exhortations they contain
3. Generate a 'super-set' of Principles
4. Cross-refer in both directions, to enable evaluation and audit
5. Evaluate each source-document against the super-set
6. Evaluate later source-documents against the super-set
7. Group the source-documents by origin, to see which are more comprehensive



## A Proposition

- Guidelines' Comprehensiveness of Coverage is likely to differ between **Entity-Categories**
- **Self-Regulatory Orientation** – low coverage
  Corporations (4), industry associations (3), professional associations (2), joint associations funded by industry (2)
- **Regulatory Orientation** – higher coverage
  Government organisations (9), NGOs (6), academics (4)



## The Results

**Reg: Very Low Coverage**
**Self-Reg: Very, Very Low Coverage**

| Category of Source | | Count | Sum | Mean | %age | Count | Sum | Mean | %age |
|---|---|---|---|---|---|---|---|---|---|
| Corporation | Co | 4 | 30 | 7.5 | 15.0% | | | | |
| Industry Association | IA | 3 | 19 | 6.3 | 12.7% | | | | |
| Professional Association | PA | 2 | 8 | 4.0 | 8.0% | | | | |
| Joint Association | JA | 2 | 13 | 6.5 | 13.0% | | | | |
| **Total Self-Regulatory Orientation** | | | | | | **11** | **70** | **6.4** | **12.7%** |
| Government Organisation | GO | 9 | 130 | 14.4 | 28.9% | | | | |
| Non-Government Organisation | NGO | 6 | 71 | 11.8 | 23.7% | | | | |
| Academic | Ac | 4 | 42 | 10.5 | 21.0% | | | | |
| **Total Regulatory Orientation** | | | | | | **19** | **243** | **12.8** | **25.6%** |
| | | | | | | **30** | **313** | **10.4** | **20.9%** |



## The Results: Some Specifics

| | Principle | Regulation % | Self-Regulation % |
|---|---|---|---|
| 1 | **Assess Positive and Negative Impacts and Implications** | 31 | 7 |
| 3 | **Ensure Human Control** | | |
| 3.2 | In particular, ensure human control over autonomous behaviour of AI-based technology, artefacts and systems | 32 | 9 |
| 3.4 | Respect each person's autonomy, freedom of choice and right to self-determination | 37 | 0 |
| 3.6 | Avoid deception of humans | 32 | 9 |
| 5 | **Ensure Consistency with Human Values and Human Rights** | | |
| 5.6 | Where interference with human values or human rights is outweighed by other factors, ensure that the interference is no greater than is justified ('harm minimisation') | 26 | 0 |

## Conclusions

- **Self-regulation fails ...**
  **...** on the basic question of whether guidelines that organisations set for themselves and their members are sufficiently comprehensive
- **This is consistent with regulatory theory**:
  If the objective is to manage public risk, then dependence on self-regulation is futile




## 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:

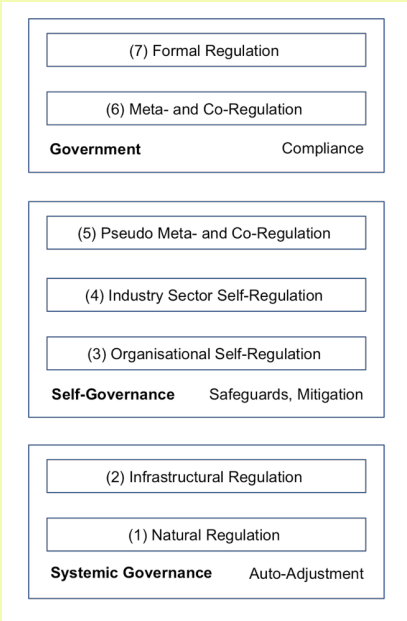
Weak Form:
**The burden of proof falls on those taking the action**

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




## Alternative Regulatory Forms

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







## 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

## Do Ethical Guidelines have a Role to Play in Relation to Data Analytics and AI/ML?

### Agenda

- Motivation: **Effective** Ethical Guidelines
- Data Analytics, AI, AI/ML
- Risk Factors in Data Analytics, AI, AI/ML
- 50 Principles from 30 Documents
- 28/30 Documents Fail against the 50 Ps
- **Self-Reg Documents fail twice as badly**
- Genuine **Co-Regulation** is necessary




## Do Ethical Guidelines have a Role to Play in Relation to Data Analytics and AI/ML?

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

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

**AiCE 2020, UniSA, Adelaide**
**Nov-Dec 2020**




## Positive Exemplars

- **EC (2019) 'Ethics Guidelines for Trustworthy AI'** High-Level Expert Group on Artificial Intelligence, European Commission, April 2019 https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477
- GEFA (2016) 'Position on Robotics and AI' **The Greens / European Free Alliance Digital Working Group**, November 2016 https://juliareda.eu/wp-content/uploads/2017/02/Green-Digital-Working-Group-Position-on-Robotics-and-Artificial-Intelligence-2016-11-22.pdf
- **Clarke R. (2019) 'Principles and Business Processes for Responsible AI'** Computer Law & Security Review 35, 4 (Jul-Aug 2019) 410-422, PrePrint at http://www.rogerclarke.com/EC/AIP.html




## Design and Evaluation Criteria for a Regulatory Regime

**Process**
- Clarity of Aims, Requirements
- Transparency
- Participation
- Reflection of Stakeholder Interests

**Product**
- Comprehensiveness
- Parsimony
- Articulation
- Educative Value

**Outcomes**
- Oversight
- Enforceability
- Enforcement
- Review


http://www.rogerclarke.com/SOS/Drones-PS.html#T2