# Regulatory Alternatives for AI and Robotics

**Roger Clarke**
Xamax Consultancy, Canberra
Visiting Professor, UNSW Law and ANU RSCS

http://rogerclarke.com/DV/RAII.html (Text)
http://rogerclarke.com/DV/RAII.pdf (Slides)

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



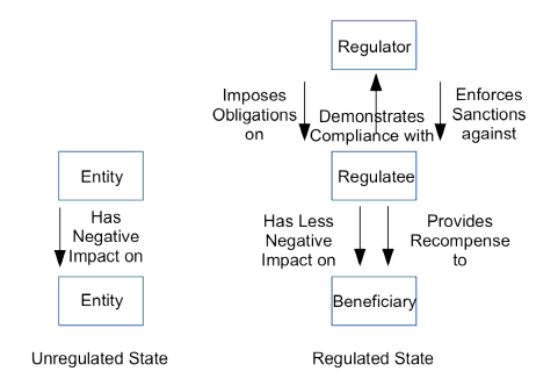

# A Comprehensive Framework for Regulatory Regimes

- Biology
- General Systems Theory
- Industrial Control
- Cybernetics
- Socio-Economic Law and Policy




# A Comprehensive Framework for Regulatory Regimes

- Biology
- General Systems Theory
- Industrial Control
- Cybernetics
- Socio-Economic Law and Policy







# Dimensions of Regulatory Contexts

- **Cultures** (sub-dimensions: ethnic, lingual, religious, philosophical)
- **Systems of Law** (code, common law)
- **Divisions of Law** (private, public; criminal, commercial)
- **Jurisdictional Structures** (international; supra-national; national, incl. single-, dual-level)
- **Industry Sectors** (categories of goods and services, and associated threats)

## 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



## Regulatory Mechanisms

- **Factors that influence the behaviour of entities** towards policy objectives, in particular by **stimulating** appropriate behaviours, and **deterring, preventing, sanctioning** harmful ones
- Mechanisms may be:
  - **Natural Processes**
  - **Instruments / Measures / Artefacts**:
    - with extraneous purposes
    - with a regulatory purpose



**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



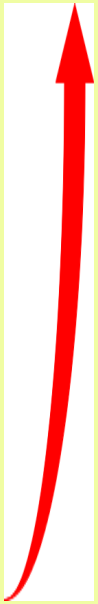
## Accountability depends on Decision-Criterion Transparency

- A-rationality
  - Unexplainability
    - Unreplicability
      - Unauditability
        - Uncorrectability
          - Unaccountability

**Quo vadis natural justice and procedural fairness?**

# Proposition:

# A Hierarchy of Regulatory Mechanisms

Formalism
Compulsion





## Natural Regulation

**Intrinsic to the relevant socio-economic system**

Competitive Behaviour
Reputational Effects
Cost/Benefit Trade-Offs

'Leave well alone'
'If it ain't bust, don't fix it'






## Infrastructural Regulation

The mechanical steam governor:
Reinforce positive aspects
Inhibit negative aspects

Automated ...
Monitoring
Exception condition detection
Adjustment of parameters
Deployment of countermeasures
Suspension of activities

Byproduct, Retro-fitted on,
or Architected in






## Infrastructural Regulation Examples

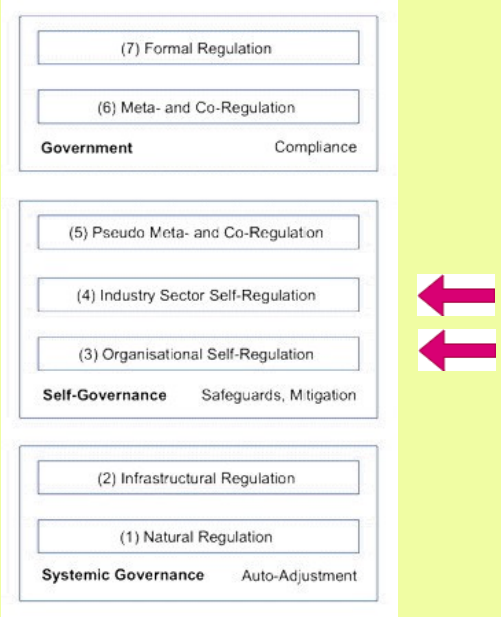
Dam sluice-gates
that automatically adjust
to water-level, water-flows
and precipitation events

Lessig's 'West Coast Code' –
computer and network architecture,
standards and protocols

'The {Extended} Laws of Robotics'
http://rogerclarke.com ...
... /SOS/Asimov.html#LawsExt


## Self-Regulation

- **Ethical Guidelines**
- **Codes of Ethics**
- **Codes of Practice**
- Meta-Brands
- Complaint Schemes
- Customer Charters
- Ethics Statements
- Corp. Soc. {& Env'l} Responsibility (CSR) Statements
- **Risk Assessment & Risk Management**






## 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



## Principles for Responsible AI 10 Themes / 50 Principles

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



## 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
1.2 **Justify objectives**
1.3 Demonstrate the achievability of postulated benefits **('justification')**
1.4 Conduct **impact assessment**
1.5 Publish **sufficient information to stakeholders** to enable them to conduct impact assessment
1.6 Conduct **consultation with stakeholders** and enable their participation
1.7 **Reflect stakeholders' justified concerns**
1.8 Justify negative impacts on individuals ('**proportionality**')
1.9 Consider less harmful ways of achieving the same objectives ('**mitigation**')

# Risk Assessment

**For Organisations**

- ISO 31000/10 – Risk Mngt Process Standards
- ISO 27005 etc. – Information Security Risk Mngt
- NIST SP 800-30 – Risk Mngt Guide for IT Systems
- ISO 8000 – Data Quality Process Standard
- ISACA COBIT, ITIL, PRINCE2, ...



# Risk Assessment

**For Organisations**

- ISO 31000/10 – Risk Mngt Process Standards
- ISO 27005 etc. – Information Security Risk Mngt
- NIST SP 800-30 – Risk Mngt Guide for IT Systems
- ISO 8000 – Data Quality Process Standard
- ISACA COBIT, ITIL, PRINCE2, ...

**For Participants/Users and for 'Usees'**

- Technology Assessment (TA)
- Privacy Impact Assessment (PIA)

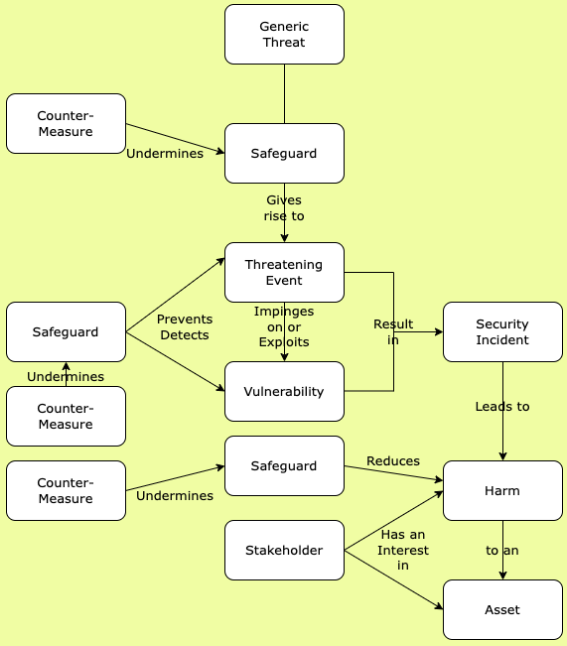


# The Conventional Security Model

# that underpins Risk Assessment

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

http://rogerclarke.com/DV/MSRA-VIE.html#RA






# Generic Risk Management Strategies

**Proactive Strategies**

- Avoidance
- Deterrence
- Prevention
  e.g. Redundancy

**Reactive Strategies**

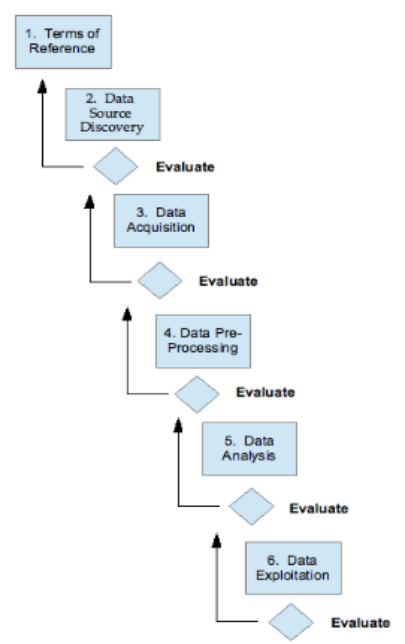
- Detection
- Isolation / Mitigation
- Recovery
- Transference
  e.g. Insurance

**Non-Reactive Strategies**

- Tolerance / Acceptance
  e.g. Self-Insurance
- Abandonment
- Dignified Demise / Graceful Degradation
- Dismal Demise / Graceless Degradation

## A Conventional Business Process for Data Analytics

## incl. AI/ML/ANN





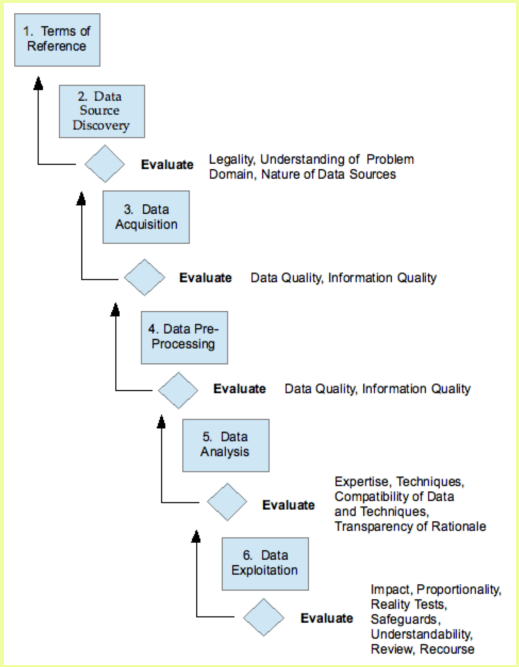



## A Conventional Business Process for Data Analytics

## incl. AI/ML/ANN

## Articulated

http://www.rogerclarke.com/EC/BDBP.html#PD







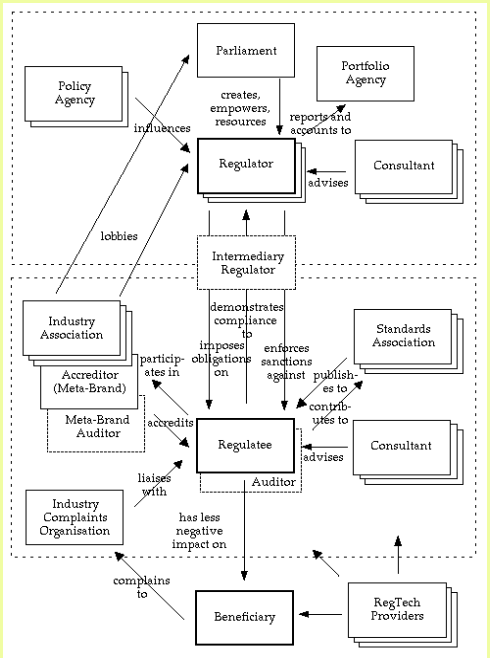

## A View of Self-Regulation

Wolves herd sheep
not for the benefit of the sheep
but for the benefit of the wolves





## Players in Regulatory Schemes



**Meta-Regulation**
Govt-regulated (req'd?) industry self-regul'n

**Co-Regulation**
- Statutory basis
- **Institutionalised negotiation process**
- **All stakeholders in**
- Actual enforcement




# Modalities of Law

| | | |
|---|---|---|
| 1. | Prohibition | You must not |
| 2. | Conditional Prohibition | You must not unless |
| 3. | Silence | It's up to you |
| 4. | Conditional Permission | You may, provided that |
| 5. | Permission | You may |
| 6. | Mandation | You must |



# Modality Expressions

- ... is exempt ...
- Despite ...
- {However} ... not ... if ...
- ... shall not ... unless ...
- ... other than ...
- ... not being ...
- ... do not apply to ...
- ... not in relation to / in respect of ...
- ... but does not include ...
- ... established ... otherwise than ...
- ... {an entity} other than ...
- ... is not covered by ...
- Except so far as the contrary intention appears, ...
- A permitted general situation exists in relation to ...
- ... a permitted purpose ...
- ... such steps as are reasonable in the circumstances …
- ... if it is practicable to specify ...
- ... in such form as is appropriate ...
- ... does not apply if ...
- ... unless the information is reasonably necessary for, or directly related to, one or more of the entity's functions or activities ...
- ... unless it is unreasonable or impracticable
- ... usually ...
- ... is likely to ...
- Despite ...
- ... reasonably believes ...



# The Political Economy of AI & Robotics

- **Adversarial – Constructive Tension**
  - Exercise of Organisational Power e.g. imposed Terms, surv'ance PITs
  - Exercise of (Weaker) People-Power e.g. lies, complaints, lawsuits, PETs

## The Political Economy of AI & Robotics

- **Adversarial – Constructive Tension**
  - Exercise of Organisational Power
    e.g. imposed Terms, surv'ance PITs
  - Exercise of (Weaker) People Power
    e.g. lies, complaints, lawsuits, PETs
- **Collaborative – Co-Regulation**
  - A Stakeholder-Negotiated Design, Regulator-Enforced
  - Designed-in Balance



## 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



## Regulatory Alternatives for AI and Robotics

**Roger Clarke**
Xamax Consultancy, Canberra
Visiting Professor, UNSW Law and ANU RSCS

http://rogerclarke.com/DV/RAII.html (Text)
http://rogerclarke.com/DV/RAII.pdf (Slides)

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