



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

**Computer Law
&
Security Review**



Regulatory alternatives for AI

Roger Clarke^{a,b,c,*}

^a Xamax Consultancy Pty Ltd, Canberra, Australia

^b Australian National University, Canberra, Australia

^c University of N.S.W., Sydney, Australia

ARTICLE INFO

Keywords:

Natural controls
Precautionary principle
Infrastructural regulation
Co-regulation
Data analytics
AI/ML
Robotics
Cyborgisation

ABSTRACT

Artificial Intelligence (AI) is enjoying another of its periodic surges in popularity. To the extent that the current promises are fulfilled, AI may deliver considerable benefits. Whether or not it does so, however, AI harbours substantial threats. The first article in this series examined those threats. The second article presented a set of Principles and a business process whereby organisations can approach AI in a responsible manner. Given how impactful AI is expected to be, and the very low likelihood that all organisations will act responsibly, it is essential that an appropriate regulatory regime be applied to AI.

This article reviews key regulatory concepts, and considers each of the various forms that regulatory schemes can take. Given the technical and political complexities and the intensity of the threats, co-regulation is proposed as the most appropriate approach. This involves the establishment of a legislated framework with several key features. The parliament needs to declare the requirements, the enforcement processes and sanctions, and allocate the powers and responsibilities to appropriate regulatory agencies. In addition, it delegates the development and maintenance of the detailed obligations to an independent body, comprising representatives of all stakeholder groups, including the various categories of the affected public.

© 2019 Roger Clarke. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Current manifestations of Artificial Intelligence (AI) are again attracting considerable attention. AI is claimed to offer great promise in such areas as labour-savings, decision-making and action that are more rapid, or reliable, or better-quality, and the discovery of new and valuable information that would otherwise have remained hidden. On the other hand, AI embodies many and serious threats. In the first article in this series, the following interpretation was made of the nature of public concerns about AI:

AI gives rise to errors of inference, of decision and of action, which arise 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

The root-causes of this cluster of concerns were identified as artefact autonomy, assumptions about data and about the inferencing process, the opaqueness of the inferencing process, and the failure to sheet home responsibility to legal entities. A reasoned analysis is needed of appropriate ways in which societies can manage the risks and still extract AI

* Corresponding author at: Xamax Consultancy Pty Ltd, Canberra, Australia.

E-mail address: Roger.Clarke@xamax.com.au

technology's achievable benefits. This article canvasses the possibilities.

A review of key regulatory concepts is first presented, including criteria for the design and evaluation of a regulatory regime. A discussion of natural controls enables definition of a threshold test for regulatory intervention. The ineffectiveness of existing laws is highlighted. The various forms that regulatory schemes can take are then outlined, and the relevance of each to AI's particular features is considered. The article concludes with a proposal for a co-regulatory framework for the management of public risk arising from AI.

2. Regulatory concepts

This section briefly summarises key concepts within regulatory theory, in order to establish a basis on which the analysis of alternative approaches to the regulation of AI can proceed. There are many definitions of the notion of 'regulation'. See, for example, [Black \(2002\)](#) and [Brownsword and Goodwin \(2012\)](#). The following, instrumentally useful definition of regulation is adopted in this work:

Regulation is the exercise of control over the behaviours of entities

This definition is phrased in such a manner as to encompass not only purpose-designed instruments of policy but also accidental and incidental control mechanisms. The adopted expression avoids any terms relating to means, and excludes the ends to which the regulation is addressed. The objectives of regulatory schemes are commonly contested, and they change over time. In addition, the effectiveness with which the means achieve the ends is not a definitional feature, but rather an attribute of regulatory regimes.

The previous article in the series discussed the range of stakeholders in AI initiatives, and argued that AI-based artefacts and systems are so impactful that organisations need to consider the interests of all stakeholders. Regulation operates at a layer above individual artefacts, systems and technologies, and aims to achieve control over them, generally through the entities that develop and apply them.

Rather than the stakeholder notion, which is a useful analytical tool within organisations, a different approach to characterising the relevant entities is more appropriate to the present purpose. The working definition of regulation provided above refers to the entities whose behaviour is subject to control. In all, three primary categories of entity are usefully distinguished, as follows:

- 'Regulatees' are entities that are subject to a regulatory scheme. Regulatees may include corporations, unincorporated business enterprises, government agencies, cooperatives, incorporated and unincorporated associations, and individuals.
- 'Regulators' are entities that exercise powers in order to achieve control over the behaviours of regulatees. A regulator may be a tightly-controlled government agency or a relatively independent commission, but it may alterna-

tively be an incorporated body such as a stock exchange, or perhaps even an association that administers an industry Code of Conduct.

- 'Beneficiaries' are entities that are advantaged by the regulatory arrangements, whether intentionally, accidentally or incidentally. The categories of entities may include any of the forms listed under 'regulatees'. However, the notion may be regarded as extending to social values such as trust in social and economic institutions, and environmental qualities. The foreground sub-category is the entities whose interests are threatened by AI, and whose disadvantages are to be mitigated by the regulatory arrangements. Encompassing more than only the intended beneficiaries is useful, however, because most schemes have accidental winners (e.g. incumbent regulatees may gain strategic benefits over new entrants), and they naturally become opponents of changes to a scheme that might diminish or remove their advantages.

Detailed analysis of any particular regulatory context requires a much more granular analysis. Other entity-categories of particular importance are representatives of and intermediaries for regulatees (such as lawyers, insurers, financiers and consultants), and advocates for the interests of beneficiaries. Those entities support flows of market signals, which are crucial to an effective regulatory scheme. A more comprehensive model is in [Clarke \(2018a\)](#).

The design of any particular regulatory regime reflects the aims of players in the processes of regulation, de-regulation and re-regulation. The coherence and completeness varies greatly among schemes, depending on the degree of conflict among interests and the power of the various parties involved in or affected by the design process. Subsequent amendments to regulatory regimes may extend, retract or simplify the requirements, but often increase their complexity.

Guidance in relation to the design of regulatory regimes, and in relation to the evaluation of existing schemes, is provided by the criteria listed in [Table 1](#). This was developed by drawing on a wide range of literature, with [Gunningham et al. \(1998\)](#), [Hepburn \(2006\)](#) and [ANAO \(2007\)](#) being particularly useful.

A large body of theory exists relating to regulatory mechanisms ([Braithwaite, 1982](#); [Braithwaite and Drahos, 2000](#); [Drahos, 2017](#)). During the second half of the 20th century, an appropriate form for a regulatory scheme was seen as involving a regulatory body that had available to it a comprehensive, graded range of measures, in the form of an 'enforcement pyramid' or 'compliance pyramid' ([Ayres and Braithwaite, 1992](#), p. 35). That model envisages a broad base of encouragement, including education and guidance, which underpins mediation and arbitration, with sanctions and enforcement mechanisms such as directions and restrictions available for use when necessary, and suspension and cancellation powers to deal with serious or repeated breaches.

In recent decades, however, further forms of regulation have emerged, many of them reflecting the power of regulatees to resist and subvert the exercise of power over their behaviour. The notion of 'governance' has been supplanting the

Table 1 – Criteria for the design and evaluation of a regulatory regime.
 Extended version of Table 2 in [Clarke and Bennett Moses \(2014\)](#).

Process

- **Clarity of aims and requirements**
Purposes and obligations are understandable by regulatees and beneficiaries
- **Transparency**
Development and review processes are open, and requirements are published
- **Participation**
All stakeholders are involved in development and review processes
- **Reflection of Stakeholder Interests**
The needs of beneficiaries are addressed, and the legitimate interests of regulatees are reflected

Product

- **Comprehensiveness**
All relevant aspects are encompassed within a coherent framework
- **Parsimony**
The regime is no more onerous or expensive than is justified
- **Articulation**
The requirements are sufficiently specific and operationalised to enable effective and efficient implementation by regulatees
- **Educative value**
Requirements are expressed in explanatory and instructive form, rather than in abstract, discursive prose

Outcomes

- **Oversight**
Regulated behaviours are subject to monitoring
- **Enforceability**
Regulated behaviours are subject to enforcement actions, by beneficiaries directly, and by enforcement agencies
- **Enforcement**
Enforcement agencies have appropriate powers and resources, and apply them in order to achieve compliance
- **Transparency**
Actions taken by regulators, and responses by regulatees, are published, thereby influencing the behaviour of all regulatees
- **Review**
The scheme is reviewed and adapted to ensure that the outcomes correspond to the aims.

notion of ‘government’, with Parliaments and Governments in many countries withdrawing from the formal regulation of industries ([Scott, 2004](#); [Jordan et al., 2005](#)). Much recent literature has focussed on deregulation, through such mechanisms as ‘regulatory impact assessments’ designed to justify the ratcheting down of measures that constrain corporate freedom, and euphemisms such as ‘better regulation’ to disguise the easing of corporations’ ‘compliance burden’. Meanwhile, government agencies resist the application of regulatory frameworks to themselves, resulting in a great deal of waste and corruption going unchecked.

It might seem attractive to organisations to face few legal obligations and hence to be subject to limited compliance risk exposure. On the other hand, the absence or weakness of regulation encourages behaviour that infringes reasonable public expectations. Cavalier organisational behaviour may be driven by executives, groups and even lone individuals who perceive opportunities. This can give rise to substantial direct and indirect threats to the reputation of every organisation in the sector. It is therefore in each organisation’s own self-interest for a modicum of regulation to exist, in order to provide a protective shield against media exposés, and to avoid stimulating a public backlash and regulatory activism.

The range of alternative forms that regulatory schemes can take is examined in a later section. First, however, it is important to consider the extent to which natural controls may

cause regulatory intervention to be unnecessary and even harmful, and hence to identify the circumstances in which intervention may be justifiable.

3. Natural controls and the justification of intervention

AI technologies and AI-based artefacts and systems may be subject to limitations as a result of processes that are intrinsic to the relevant socio-economic system ([Clarke, 1995, 2014a, 2014b](#)). AI may even stimulate natural processes whose effect is to limit adoption, or to curb or mitigate negative impacts.

A common example of a natural control is doubt about the technology’s ability to deliver on its proponents’ promises, resulting in inventions being starved of investment. Where innovative projects succeed in gaining early financing rounds, it may transpire that the *development and/or operational costs* are too high, or the number of instances it would apply to and/or the benefits to be gained from each application may be too small to justify the investment needed to develop artefacts or implement and deploy systems.

In some circumstances, the realisation of the potential benefits of a technology may suffer from *dependence on infrastructure that is unavailable or inadequate*. For example, computing could have exploded in the third quarter of the 19th century, rather than 100 years later, had metallurgy of the day

been able to support Babbage's 'difference' and 'analytical' engines and sufficient investment had been secured.

Another form of natural control is the exercise of *countervailing power* by entities that perceive negative impacts on their interests. Common examples are the market power of competitors, suppliers, customers and employees, and the institutional power of regulators, financiers and insurers. It has long been feasible for opponents to stir up public opprobrium through the media, and further opportunities are now provided by social media. A related factor is *reputational effects*, whereby early implementations may excite opposition because of a perception that the approach is harmful to important social values, resulting in a seriously negative public image. A case study is provided by Boeing's stall-prevention feature for the 737 MAX, which was in practice unable to be overridden by pilots, resulting in two crashes and over 300 deaths in 2018–19, leading to suspension of operational use of the aircraft (FAA, 2019).

The economic aspects of natural controls require closer attention. The postulates that an individual who "intends only his own gain" is led by "an invisible hand" to promote the public interest (Smith, 1776), and that economic systems are therefore inherently self-regulating, have subsequently been bolstered by transaction cost economics (Williamson, 1979). Limits to inherent self-regulation have been noted, however, such as 'the tragedy of the (unmanaged) commons' notion (Hardin, 1968, 1994; Ostrom, 1999). Whereas neo-conservative economists commonly recognise 'market failure' as the sole justification for interventions, Stiglitz (2008) adds 'market irrationality' (e.g. circuit-breakers to stop bandwagon effects in stock markets) and 'distributive justice' (e.g. safety nets and anti-discrimination measures).

In the case of AI, evidence of market failure was noted in the previous article in this series. Despite various technologies being operationally deployed, no meaningful organisational, industry or professional self-regulation exists. Such codes and guidelines as exist cover a fraction of the need, and are in any case unenforceable. Meanwhile market irrationality is evident in the form of naive acceptance by user organisations of AI promoters' claims; and distributive justice is being negatively impacted by unfair and effectively unappealable decisions in such areas as credit-granting and social welfare administration.

A further important insight that can be gained from a study of natural controls is that regulatory measures can be designed to reinforce natural processes. For example, approaches that are applicable in a wide variety of contexts include adjusting the cost/benefit/risk balance perceived by the players, by subsidising costs, levying revenues and/or assigning risk. For example, applying strict liability to operators of drones and driverless cars could be expected to encourage much more careful risk assessment and risk management.

An appreciation of pre-existing and enhanced natural controls is a vital precursor to any analysis of regulation, because the starting-point needs to be:

What is there about the natural order of things that is inadequate, and how will intervention improve the situation?

For example, the first of 6 principles proposed by the Australian Productivity Commission was "Governments should

not act to address 'problems' through regulation unless a case for action has been clearly established. This should include evaluating and explaining why existing measures are not sufficient to deal with the issue" (PC, 2006, p.v). That *threshold test* is important, in order to ensure a sufficient understanding of the natural controls that exist in the particular context.

In practice, parliaments seldom act in advance of new technologies being deployed. Reasons for this include lack of understanding of the technology and its impacts, prioritisation of economic over social issues and hence a preference for stimulating new business rather than throttling it at birth, and more effective lobbying by innovative corporations than by consumers and advocates for social values.

An argument exists to the effect that, the more impactful the technology, the stronger the case for anticipatory action by parliaments. Examples of technologies that are frequently mentioned in this context are nuclear energy and various forms of large-scale extractive and manufacturing industries whose inadequate regulation has resulted in massive pollution. A '*precautionary principle*' has been enunciated (Wingspread, 1998). Its strong form exists in some jurisdictions' environmental laws, along the lines of:

When human activities may lead to morally unacceptable harm that is scientifically plausible but uncertain, actions shall be taken to avoid or diminish that potential harm. (TvH, 2006).

Beyond environmental matters in a number of specific jurisdictions, however, the precautionary principle is merely an ethical norm to the effect that:

If an action or policy is suspected of causing harm, and scientific consensus that it is not harmful is lacking, then the burden of proof falls on those taking the action

The first article in this series argued that AI's threats are readily identifiable and substantial. Even if that contention is not accepted, however, the scale of impact that AI's proponents project as being inevitable is so great that the precautionary principle applies, at the very least in the weaker of its two forms. A strong case for regulatory intervention therefore exists, unless it can be shown that appropriate regulatory measures are already in place. The following section accordingly presents a brief survey of existing regulatory arrangements.

4. Existing laws

This section first considers general provisions of law that may provide suitable protections, or at least contribute to a regulatory framework. It then reviews initiatives that are giving rise to AI-specific laws.

4.1. Generic laws

Applications of new technologies are generally subject to existing laws (Bennett Moses, 2013). These include the various forms of commercial law, particularly contractual obligations including express and implied terms, consumer rights laws, and copyright and patent laws. In some contexts – including

robotics, cyborg artefacts, and AI software embedded in devices – product liability laws are likely to apply. Further laws that assign risk to innovators may apply, such as the tort of negligence, as may laws of general applicability such as human rights law, anti-discrimination law and data protection law. The obligations that corporations law assigns to company directors are relevant. Further sources of regulatory impact are likely to be the laws relating to the various industry sectors within which AI is applied, such as road transport law, workplace and employment law, and health law.

However, particularly in common law jurisdictions, there is likely to be a great deal of uncertainty about the way in which laws will be applied by tribunals and courts if any particular dispute reaches them. This acts to some extent as a deterrent against innovation, and can considerably increase the costs incurred by proponents, and delay deployment. From the viewpoint of people who perceive themselves to be negatively affected by the innovation, on the other hand, legal channels for combatting those threats may be inaccessible, expensive, slow and even entirely ineffectual.

Particularly with ‘breakthrough’, revolutionary and disruptive technologies, existing laws are likely to be ill-fitted to the new context, because those laws were “designed around a socio-technical context of the relatively distant past” (Bennett Moses, 2011, p. 765), and without knowledge of the new form. In some cases, existing law may hinder new technologies in ways that are unhelpful to both the innovators and those affected by them. In other cases, existing law may have been framed in such a manner that it does not apply to the new form, or judicial calisthenics has to be performed in order to make it appear to apply. For a case study of judicial calisthenics in the software copyright arena, see Clarke (1988).

4.2. AI-specific laws

Spatially-constrained industrial robotics, in production-lines and warehouses, is well-established. Various publications have discussed general questions of robot regulation (e.g. Leenes and Lucivero, 2014; Scherer, 2016; HTR, 2018a, 2018b), but few identify AI-specific laws. Even such vital aspects as worker safety and employer liability appear to depend not on technology-specific laws, but on generic laws, which may or may not have been adapted to reflect the characteristics of the new technologies.

In HTR (2017), South Korea is identified as having enacted the first national law relating to robotics generally: the Intelligent Robots Development Distribution Promotion Act of 2008. It is almost entirely facilitative and stimulative, and barely even aspirational in relation to regulation of robotics. There is mention of a ‘Charter’, “including the provisions prescribed by Presidential Decrees, such as ethics by which the developers, manufacturers, and users of intelligent robots shall abide” – but no such Charter appears to exist. A mock-up of a possible form for such a Charter is provided by Akiko (2012). HTR (2018c) offers a regulatory specification in relation to research and technology generally, including robotics and AI.

In relation to autonomous motor vehicles, a number of jurisdictions have enacted laws. See Palmerini et al. (2014, pp.36–73), Holder et al. (2016), DMV-CA (2018), Vellinga (2017), which reviews laws in the USA at federal level, California, the

United Kingdom, and the Netherlands, and Maschmedt and Searle (2018), which reviews laws in three States of Australia. Such initiatives have generally had a strong focus on economic motivations, the stimulation and facilitation of innovation, exemptions from some existing regulation, and limited new regulation or even guidance. One approach to regulation is to leverage off natural processes. For example, Schellekens (2015) argued that a requirement of obligatory insurance was a sufficient means for regulating liability for harm arising from self-driving cars. In the air, legislatures and regulators have moved very slowly in relation to the regulation of drones (Clarke and Bennett Moses, 2014; Clarke, 2016).

Automated decision-making about people has been subject to French data protection law for many years. In mid-2018 this became a feature of European law generally, through the General Data Protection Regulation (GDPR) Art. 22, although doubts have been expressed about that Article’s effectiveness (Wachter et al., 2017).

On the one hand, it might be that AI-based technologies are less disruptive than they are claimed to be, and that laws need little adjustment. On the other, a mythology of ‘technology neutrality’ pervades law-making. Desirable as it might be for laws to encompass both existing and future artefacts and processes, genuinely disruptive technologies have features that render existing laws ambiguous and ineffective.

Not only is AI not subject to adequate natural controls, but such laws as currently apply appear to be inadequate to cope with the substantial threats it embodies. The following section accordingly outlines the various forms of regulatory intervention that could be applied.

5. The hierarchy of regulatory forms

This section reflects the regulatory concepts outlined earlier, and presents alternatives within a hierarchy based on the degree of formality of the regulatory intervention. An earlier section considered *Natural Regulation*. In Fig. 1, this is depicted as the bottom-most layer (1) of the hierarchy.

Regulatory theorists commonly refer to ‘instruments’ and ‘measures’ that can be used to achieve interventions into natural processes. In principle, their purpose is the curbing of harmful behaviours and excesses; but in some cases the purpose is to give the appearance of doing so, in order to hold off stronger or more effective interventions. Fig. 1 depicts the intentionally-designed regulatory ‘instruments’ and ‘measures’ as layers (2)–(6), built on top of natural regulation.

The second-lowest layer in the hierarchy, referred to as (2) *Infrastructural Regulation*, is a correlate of artefacts like the mechanical steam governor. Features of the infrastructure on which the regulatees depend can reinforce positive aspects of the relevant socio-economic system and inhibit negative aspects. Those features may be byproducts of the artefact’s design, or they may be retro-fitted onto it, or architected into it. For example, early steam-engines did not embody adequate controls, and the first governor was a retro-fitted feature; but, in subsequent iterations, controls became intrinsic to steam-engine design.

Information technology (IT) assists what were previously purely mechanical controls, such as where dam sluice-gate



Fig. 1 – The hierarchy of regulatory forms.

settings are automatically adjusted in response to measures of water-level, catchment-area precipitation events or increases in feeder-stream water-flows. IT, and AI-augmented IT, provide many opportunities. One popular expression for infrastructural regulation in the context of IT is ‘West Coast Code’ (Lessig, 1999; Hosein et al., 2003). A range of constraints exists within computer and network architecture – including standards and protocols – and within infrastructure – including hardware and software.

In the context of AI, a relevant form that ‘West Coast Code’ could take is the embedment in robots of something resembling ‘laws of robotics’. This notion first appeared in an Asimov short story, ‘Runaround’, published in 1942; but many commentators on robotics cling to it. For example, Devlin (2016) quotes a professor of robotics as perceiving that the British Standard Institute’s guidance on ethical design of robots (BS, 2016) represents “the first step towards embedding ethical values into robotics and AI”. On the other hand, a study of Asimov’s robot fiction showed that he had comprehensively demonstrated the futility of the idea (Clarke, 1993). No means exists to encode into artefacts human values, nor to embed within them means to reflect differing values among various stakeholders, nor to mediate conflicts among values and objectives (Weizenbaum, 1976; Dreyfus, 1992).

Switching attention to the uppermost layer of the regulatory hierarchy, (6) *Formal Regulation* exercises the power of a parliament through statutes. In common law countries at least, statutes are supplemented by case law that clarifies the application of the legislation. Formal regulation demands

compliance with requirements that are expressed in more or less specific terms, and is complemented by sanctions and enforcement powers. Lessig underlined the distinction between infrastructural and legal measures by referring to formal regulation as ‘East Coast code’.

Regulatory requirements are not necessarily unqualified, and are not necessarily expressed in a negative form. Clarke and Greenleaf (2018) distinguish a number of modalities, identified in Table 2. Combinations of prohibition of some categories of behaviour, and mandation of other forms, are commonly complemented by qualified forms of approval and disapproval, subject to more or less clear pre- and post-conditions being fulfilled.

A narrow interpretation of law is that it is rules imposed by a politically recognised authority. An expansive interpretation of law, on the other hand, recognises a broader set of phenomena, including delegated legislation (such as Regulations); treaties that bind the State; decisions by courts and tribunals that influence subsequent decisions; law made by private entities, but endorsed and enforced by the State, particularly through contracts and enforceable undertakings; and quasi-legal instruments such as memoranda of understanding (MOUs) and formal Guidance Notes (Clarke and Greenleaf, 2018).

Formal regulation often involves a specialist government agency or parliamentary appointee exercising powers and resources in order to enforce laws and provide guidance to both regulatees and beneficiaries. The legal authority and resources assigned to regulators may be limited, in which case the regime is appropriately described as pseudo-regulation, and the entity as a mere oversight agency rather than a regulator.

Various forms of AI may lie within the scope of an existing agency or appointees, as is commonly the case with self-driving cars, remotely-operated aircraft, and medical implants that include data-processing capabilities. In relation to AI in general, however, six decades after the AI era was launched, the major jurisdictions around the world have not established new regulatory agencies or suitably empowered existing ones. The EU has yet to move beyond a discussion document issued by the Data Protection Supervisor (EDPS, 2016), and a preliminary statement (EC, 2018). The UK Data Protection Commissioner has only reached the stage of issuing a discussion paper (ICO, 2017). The current US Administration’s policy is entirely stimulatory in nature, and mentions regulation solely as a barrier to economic objectives (WH, 2018). Principles have been proposed by a diverse array of organisations, but most are of the nature of aspirations rather than obligations. Examples include the European Greens Alliance (GEFA, 2016), the UNI

Table 2 – Modalities of law.

1.	Prohibition	You must not
2.	Conditional prohibition	You must not unless
3.	Silence	It is up to you
4.	Conditional permission	You may, provided that
5.	Permission	You may
6.	Mandation	You must

Global Union (UGU, 2017), the Japanese government (Hirano, 2017), a House of Lords Committee (HOL, 2018), and the French Parliament (Villani, 2018).

Regulation of the formal kind imposes considerable constraints and costs. The intermediate layers (3)–(5) seek to reduce the considerable constraints and imposts inherent in formal regulation. The lowest of these layers, (3) *Organisational Self-Regulation*, includes internal codes of conduct and ‘customer charters’, and self-restraint associated with expressions such as ‘business ethics’ and ‘corporate social responsibility’ (Parker, 2002). It was noted in the second article of this series that directors of corporations are required by law to pursue the interests of the corporation ahead of all other interests. It is therefore unsurprising, and even to be expected, that organisational self-regulation is almost always ineffectual from the viewpoint of the supposed beneficiaries, and often not even effective at protecting the organisation itself from bad publicity. Recent offerings by major corporations include IBM (2018), Google (Pichai, 2018) and MS (2019). For an indication of the scepticism with which such documents are met, see Newcomer (2018).

The mid-point of the hierarchy is (4) *Industry Sector Self-Regulation*. Corporations club together for various reasons, some of which can be to the detriment of other parties, such as collusion on bidding and pricing. The activities of industry associations are, however, capable of delivering benefits for others, as well as for their members. In particular, collaborative approaches to infrastructure can improve services, reduce costs for the sector’s customers, and even embed infrastructural regulatory mechanisms.

It could also be argued that, if norms are promulgated by the more responsible corporations in an industry sector, then misbehaviour by the industry’s ‘cowboys’ would be highlighted. Codes of conduct, or of practice, or of ethics, and Memoranda of Understanding (MoUs) within an industry are claimed to have, and may even have, some regulatory effect. In practice, however, the impact of Industry Codes on corporate behaviour is seldom significant. Few such Codes are sufficiently stringent to protect the interests of other parties, and the absence of enforcement undermines the endeavour. The more marginal kinds of suppliers ignore them, and responsible corporations feel the pinch of competition and reduce their commitment to them. As a result, such Codes generally act as camouflage, obscuring the absence of safeguards and thereby holding off actual regulatory measures. In the AI field, examples of industry coalitions eagerly pre-countering the threat of regulation include FLI (2017), ITIC (2017), and PoAI (2018).

A particular mechanism used in some fields is accreditation schemes, sometimes referred to as ‘good housekeeping’ ‘ticks-of-approval’. These are best understood as meta-brands. The conditions for receiving the tick, and retaining it, are seldom materially protective of the interests of the nominal beneficiaries (Clarke, 2001; Moores and Dhillon, 2003). For a case study in active deception, by the data protection mark TrustE/TrustArc, see Connolly (2008), Connolly et al. (2014).

By their nature, and under the influence of trade practices/anti-monopoly/anti-cartel laws, industry self-regulatory mechanisms are generally non-binding and unenforceable. Further, they are subject to gaming by regulatees, in order to reduce their effectiveness and/or onerousness,

or to give rise to collateral advantages, such as lock-out of competitors or lock-in of customers. As a result, the two self-regulatory layers are rarely at all effective. Braithwaite (2017) notes that “self-regulation has a formidable history of industry abuse of privilege” (p. 124), and the conclusion of Gunningham and Sinclair (2017) is that ‘voluntarism’ is generally an effective regulatory element only when it exists in combination with ‘command-and-control’ components.

A role can also be played by professional associations, because these may balance public needs against self-interest somewhat better than industry associations. Their impact is, however, far less pronounced than that of industry associations. Moreover, the initiatives to date of the two largest international bodies are underwhelming. ACM (2017) uses weak forms such as “should” and “are encouraged to”, and IEEE (2017) offers lengthy prose but unduly vague and qualified principles. Neither has to date provided the guidance needed by professionals, managers and executives.

Industry standards are of relevance in some segments of organisational practices. HTR (2017) lists industry standards issued by the International Standards Organisation (ISO) in the AI arena. A considerable proportion of industry standards focus on inter-operability, and some others describe business processes intended to achieve quality assurance. Public safety is an area of strength, particularly in the field commonly referred to as ‘safety-critical systems’ (e.g. Martins and Gorschek, 2016). Hence some of the physical threats embodied in AI-based systems might be avoided, mitigated and managed through the development and application of industry standards; but threats to economic and social interests are seldom addressed. Even in the business process area, progress is remarkably late, and slow. In 2016, the IEEE Standards Association announced a program to produce a Standard P7000, a ‘Model Process for Addressing Ethical Concerns During System Design’. A further 3 years on, nothing has been published.

During the last four decades, as parliaments have struggled to understand and cope with new technologies, several further regulatory forms have emerged. In one sense they are intermediate between (often heavy-handed) formal regulation and (mostly ineffective and excusatory) self-regulation. In a manner consistent with Gunningham and Sinclair (2017), they blend ‘voluntarism’ with ‘command-and-control’ components.

In Grabowsky (2017), the notion of ‘enforced self-regulation’ is traced to Braithwaite (1982), and the use of the term (5a) ‘Meta-Regulation’, in its sense of ‘government-regulated industry self-regulation’, to Gupta and Lad (1983). See also Parker (2007). An example of ‘meta-regulation’ is the exemption of “media organisations” from Australian data protection law (Privacy Act (Cth) s.7B(4)(b)), provided that “the organisation is publicly committed to observe standards that (i) deal with privacy in the context of the activities of a media organisation ...; and (ii) have been published in writing by the organisation or a person or body representing a class of media organisations”. There is no provision for any controls, and the ‘standards’ are, unsurprisingly, vacuous and unenforced. Positive exemplars of meta-regulation are very difficult to find.

In parallel, the notion of (5b) ‘Co-Regulation’ emerged (Ayres and Braithwaite, 1992; Clarke, 1999). Broadly, co-regulatory

approaches involve legislation that establishes a regulatory framework but carefully delegates the details. Key elements are authority, obligations, general principles that the regulatory scheme is to satisfy, sanctions and enforcement mechanisms. The detailed obligations are developed through consultative processes among advocates for stakeholders. The result is an enforceable Code, which articulates, and must be consistent with, the general principles expressed in the relevant legislation. The participants necessarily include at least the regulatory agency, the regulatees and the intended beneficiaries of the regulation, and the process must reflect the needs of all parties, rather than being distorted by institutional and market power. Meaningful sanctions, and enforcement of them, are intrinsic elements of a scheme of this nature.

Unfortunately, instances of effective co-regulation are also not easy to find. One reason is that the development process typically excludes or stifles the interests of the less powerful stakeholders. Another is that the terms are often not meaningfully enforced, and may even be unenforceable (Balleisen and Eisner, 2009). In Australia, for example, so-called ‘Enforceable Codes’ exist that are administered by the Australian Communications and Media Authority (ACMA) in respect of TV and radio broadcasting, and telecommunications. Similarly, the Australian Prudential Regulation Authority (APRA) administers such Codes in respect of banking services. The arrangements succeed both in facilitating business and government activities and in offering a veneer of regulation; but they fail to exercise control over behaviour that the public regards as inappropriate, and hence they have little public credibility. In contrast with meta-regulation, however, co-regulation does at least have the scope to deliver effective schemes, provided that all of the key characteristics are designed into the scheme.

This section has outlined the various forms that regulatory intervention can take. In practice, many regulatory regimes adopt a primary form but also incorporate elements from other layers: “in the majority of circumstances, the use of multiple rather than single policy instruments, and a broader range of regulatory actors, will produce better regulation [by means of] the implementation of complementary combinations of instruments and participants ...” (Gunningham and Sinclair, 2017, p. 133).

The following section briefly reviews the characteristics of the various regulatory forms, and assesses each form’s suitability as a means of achieving control over AI technologies and AI-based artefacts and systems.

6. Regulatory indicators

Each of the regulatory forms identified above may have at least some role to play in any particular context. This section considers key factors that variously favour application of each form or militate against its usefulness in achieving control over AI, and in ensuring the implementation of appropriate safeguards against the harms identified in Section 4 of the first article in this series: artefact autonomy, inappropriate assumptions about data and about the inferencing process, the opaqueness of the inferencing process, and irresponsibility.

The analysis takes into account the criteria that were presented in Table 1 for the design and evaluation of regulatory regimes. Particular emphasis is placed on the transparency of process, the reflection of stakeholders’ interests, the articulation of regulatory mechanisms, and enforcement action.

In many circumstances, *natural controls* can be effective, or at least make significant contributions. For example, the periodic spasms of public fear engendered by plane crashes may be sufficient to stall the advance of autonomous flight, and unjust actions by government agencies may stimulate public reactions that force the abandonment of automated decision-making. Natural controls are less likely to be adequate, however, where the technology’s operation is not apparent to the public, the social or socio-technical system is complex or obscure, or one or a few powerful players dominate the field and can arrange it to suit their own needs. The features of the AI industry militate against natural controls being sufficient.

In the IT arena, it is common for *infrastructural regulation* to play a role. AI offers potential for further improvements, including through the current RegTech movement (Clarke, 2018a). Such mechanisms are likely to be at their least effective, however, in circumstances that involve substantial value-conflicts, variability in context, contingencies, and rapid change. The features of the AI industry militate against infrastructural features being prioritised and implemented. For example, even moderately expensive drones lack communication channel redundancy and collision-detection features. Infrastructural regulation may be at its most effective in biomedical engineering, where the precautionary principle is already embedded, and conception and design are followed by careful and graduated trialling and testing (Robertson et al., 2019).

Organisational self-regulation can only have much effect where the intended beneficiaries have considerable power, or the risk of being subjected to expensive and inconvenient formal regulation causes regulatees to establish protections, and to actually apply them. *Industry self-regulation* can only be effective where, on the one hand, a strong industry structure exists, with strong incentives for all industry participants to be members; but, on the other hand, other players are sufficiently powerful to ensure that the scheme delivers advantages to the intended beneficiaries. Unless miscreants feel pain, such schemes lack effectiveness and credibility. AI comprises multiple technologies, which are embodied in many artefacts, which are embedded into many systems, which are subject to many applications. Some of them feature at least some degree of autonomy. All of them are complex and obscure, and unlikely to be understood even by executives, marketers and policy-makers, let alone by the affected public. AI is dynamic. The entities active in AI are in many cases small, unstable, rapidly-changing, and short-lived. There is no strong industry structure. It appears highly unlikely that organisational and industry self-regulation will be able to deliver effective protection against the substantial threats embodied in AI.

Meta-regulation could only be effective if it imposed a comprehensive set of requirements on organisational and industry self-regulatory mechanisms. It appears unlikely that many positive exemplars will emerge, and the technological and

industry complexities make it particularly unlikely that AI will be a suitable field for its application.

Formal regulation can bring the full force of law to bear. However, the processes of design, drafting and debate are subject to political forces, and these are in turn influenced by powerful voices, which generally work behind the scene and whose effects are therefore obscured and difficult to detect, far less to counter. As a result, many formal regulatory schemes fail against many of the criteria proposed in [Table 1](#) above. At the other extreme, some formal regulatory arrangements are unduly onerous and expensive, and most are inflexible. All are slow and challenging to adapt to changing circumstances, because of the involvement of powerful voices and political processes. The complexities are such that there are minimal chances of coherent discussions about AI taking place in parliaments. Attempts at formal regulation are therefore highly likely to either founder or to deliver statutes that are at the same time onerous and ineffective. Exceptions may arise, however, such as the prohibition of fully autonomous passenger flights.

Co-regulation, on the other hand, offers real prospects of delivering value to beneficiaries. A trigger is necessary, such as a zealous, powerful and persuasive Minister, or a coalition of interests within or adjacent to a particular sector. A highly-representative forum must come together, and negotiate a workable design. Relevant government(s), government agencies and parliament(s) must have and sustain commitment, and must not succumb to vested interests. A regulator must be empowered and resourced, and supported against the inevitable vicissitudes such schemes encounter. The following section further articulates this proposition.

7. A co-regulatory framework for AI

In [Sections 3](#) and [4](#) of this article, it was shown that the pre-conditions for regulatory intervention exist, by virtue of AI's substantial impacts, and the inadequacy of natural controls and existing laws. Further, significant harm could be inflicted on individuals and society, and – through reputational harm to technology and to organisations applying it – to the economy as well. The precautionary principle is therefore applicable, and regulation is necessary in advance of rather than following development and deployment.

The previous section concluded that the most effective regulatory form is co-regulation, working in combination with some natural controls and infrastructural features, together with some formal law, for example prohibition of fully autonomous passenger aircraft. To the extent that organisations apply the multi-stakeholder risk management process and the 50 Principles proposed in the second article in this series, organisational and industry self-regulation might make material contributions.

This section provides an outline of a regulatory scheme that could be applied at the level of AI as a whole. The diversity among the various forms of AI is such, however, that there would be considerable advantages in separating the initiative into several technology-specific streams.

A critical question in the design of one or more regulatory schemes for AI is specifically what is to be regulated.

The generic notion of AI is diffuse, and not in itself a suitable target. The first article in the series argued that *Complexity Intelligence* and *Intellectics* would be more suitable focal-points.

However, regulatory requirements are generally imposed on a category of entities, in respect of a category of activities. An appropriate approach to regulating AI would therefore be to apply the distinctions made in [Table 2](#) of the second article in the series, and impose requirements on entities involved in respectively the research, invention, innovation, dissemination and application of AI technology, artefacts, systems and installed systems.

At the heart of such a scheme is a comprehensive legislated framework which incorporates at least the elements expressed in [Table 3](#).

There is scope within such a co-regulatory scheme for various entities to make contributions.

Corporations at all points in the AI supply-chain can address issues, through intellectual engagement by executives, resource commitment, acculturation of staff, adaptation of business processes, controls and audit mechanisms. These activities can ensure the establishment, operation and adaptation of standards-compliant internal complaints-handling processes, and communications with other corporations in the supply chain and with other stakeholders, through Code negotiation institutions and processes and otherwise.

Industry associations can act as focal points for activities within their own sectors. This might include specific guidance for organisations within the particular industry; second-level complaints processes behind those of the member-corporations; infrastructure that implements protective technologies; and awareness-raising and educational measures.

Individuals need to be empowered, and encouraged, to take appropriate actions on their own behalf. This is only feasible if awareness-raising and educational measures are undertaken, (relatively informal) complaints processes are instituted at corporate and industry association levels, and (somewhat less informal) complaints, compliance-enforcement and damages-awarding processes are established through regulatory agencies, tribunals and the courts. In some cultures, particularly that of the United States, self-reliance may be writ especially large, while in others, it may play a smaller role, with a correspondingly larger, more powerful and better-funded regulatory agency.

The question remains as to specifically what technologies and applications within the broad AI field should be within-scope. In the first article in the series, four exemplar technologies were identified. Of these, robotics, particularly in public spaces and in motion, appears to be a prime contender for early regulation. Similarly, action is needed in relation to machine-learning techniques such as neural networks. The scope would be usefully defined as encompassing other AI-derived techniques such as rule-based expert systems, and consideration needs to be given as to why data analytics as a whole should not be subject to the same regulatory regime. A scheme of this nature could be readily developed for the medical implants field, and then adapted for other forms of cyborgisation.

The second paper in this series presented 50 Principles for Responsible AI, organised within 10 Themes. These lend

Table 3 – A comprehensive co-regulatory framework.

1.	A delegated authority	Power delegated to an independent Commission or a Minister to approve one or more Codes, and successive versions and replacements of them, subject to: a. a set of requirements with which Codes must comply b. an articulated set of principles Codes must embody c. primacy for negotiated Codes; and d. a reserve ability to impose Codes if, or to the extent that, negotiated Codes are not achieved
2.	Code institution(s)	One or more Code negotiation and maintenance institutions and processes, whose functions are: a. to operationalise the (necessarily abstract) requirements into Codes b. to do so by means of consultative processes c. to achieve active involvement and agreement from all stakeholders including, and especially, the intended beneficiaries of the regulatory scheme d. to reflect the criteria for effective regulation (such as those in Table 1); and e. to articulate principles for responsible AI, in operational form
3.	Code development resources	Resources to support that or those negotiation and maintenance institutions and processes
4.	Enforcement mechanisms	Enforcement powers and resources, and assignment of them to one or more existing and/or new regulatory agencies. Agency functions must include oversight of consultative processes, supervision of compliance, conduct of own-motion investigations and complaint investigations, imposition of penalties on miscreants, prosecution of offenders, research into technological and environmental changes, provision of an information clearing house, and provision of a focal point for adaptation of the law, and of Codes
5.	Enforcement obligations	Obligations on the regulatory agency/ies to apply the enforcement powers and resources

themselves as a template for expressing the requirements with which Codes need to comply, or as a checklist for evaluating drafts developed in some other manner. In the case of data analytics generally, and neural networks in particular, these Principles are complemented by a set of Guidelines for the Responsible Application of Data Analytics (Clarke, 2018b).

8. Conclusions

AI cannot deliver on its promises unless the substantial public risks that the technologies, artefacts, systems and applications entail are subjected to appropriate forms of public risk management. A range of alternative regulatory approaches is feasible. The co-regulatory approach has been argued to be the most appropriate to apply to the problem. A degree of articulation of the proposal has been presented.

Public hand-wringing about the risks inherent in AI is of no value unless it stimulates constructive action that addresses those risks. Meanwhile, proponents of promising technologies face the likelihood of strong public and institutional backlash against their innovations. It is therefore in the interests of all stakeholders in AI for credible public processes to be conducted, resulting in credible regulatory regimes that address, and are seen to address, the public risks, and that are no more onerous on the AI industry than is justified. The framework proposed in this article provides a blueprint for such processes and regimes.

Acknowledgement

This version has benefited from valuable feedback by Prof. Graham Greenleaf of UNSW Law, Sydney.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.clsr.2019.04.008](https://doi.org/10.1016/j.clsr.2019.04.008).

REFERENCES

- ACM. Statement on algorithmic transparency and accountability. Association for Computing Machinery; 2017, at https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf.
- ANAO. Administering regulation: better practice guide. Australian National Audit Office; 2007, at http://www.anao.gov.au/~media/Uploads/Documents/administering_regulation_.pdf.
- Akiko. South Korean robot ethics charter; 2012, Akiko's Blog, at <https://akikok012um1.wordpress.com/south-korean-robot-ethics-charter-2012/>.
- Ayres I, Braithwaite J. Responsive regulation: transcending the deregulation debate. Oxford Univ. Press; 1992.
- Balleisen EJ, Eisner M. The promise and pitfalls of co-regulation: how governments can draw on private governance for public purpose, Ch 6. In: Moss D, Cisternino J, editors. 'New Perspectives on Regulation' The Tobin Project; 2009. p. 127–49. at http://elearning.muhammadjirien.org/index.php/catalog/download/filename/New_Perspectives_Full_Text.pdf#page=127.
- Bennett Moses L. Agents of change: how the law 'copes' with technological change. Griffith Law Rev 2011;20(4):764–94, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2000428.
- Bennett Moses L. How to think about law, regulation and technology – problems with “technology” as a regulatory target. Law, Innov & Technol 2013;5(1):1–20.

- Black J. (2002) 'Critical reflections on regulation' 27 *Aust J Legal Philos* 1.
- Braithwaite J. *Enforced self-regulation: a new strategy for corporate crime control*. *Mich Law Rev* 1982;80(7):1466–507.
- Braithwaite J. (2017) 'Types of responsiveness' Chapter 7 in Drahos, pp. 117–132, at <http://press-files.anu.edu.au/downloads/press/n2304/pdf/ch07.pdf>.
- Braithwaite B, Drahos P. *Global business regulation*. Cambridge University Press; 2000.
- Brownsword R, Goodwin M. *Law in Context: law and the technologies of the twenty-first century: text and materials*. Cambridge University Press; 2012.
- BS. *Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems*. British Standards Institute; 2016.
- Clarke R. *Judicial understanding of information technology: the case of the Wombat ROMs*. *Comput J* 1988;31(1):25–33. PrePrint at <http://www.rogerclarke.com/SOS/WombatROMs-1988.html>.
- Clarke R. 'Asimov's laws of robotics: implications for information technology' in two parts. *IEEE Comput* 1993;26(12):53–61. (December) and 27,1 (January 1994) 57–66, at <http://www.rogerclarke.com/SOS/Asimov.html>.
- Clarke R. (1995) 'A normative regulatory framework for computer matching' *J Comput InfLaw* XIII, 4 (Summer) 585–633, PrePrint at <http://www.rogerclarke.com/DV/MatchFrame.html#IntrCtls>.
- Clarke R. (1999) 'Internet privacy concerns confirm the case for intervention' *Commun. ACM* 42, 2 (Feb) 60–67, PrePrint at <http://www.rogerclarke.com/DV/CACM99.html>.
- Clarke R. *Meta-brands*. *Priv Law Policy Report* 2001;7(11). PrePrint at <http://www.rogerclarke.com/DV/MetaBrands.html>.
- Clarke R. *What drones inherit from their ancestors*. *Comput Law Secur Rev* 2014a;30(3):247–62. PrePrint at <http://www.rogerclarke.com/SOS/Drones-I.html>.
- Clarke R. *The regulation of the impact of civilian drones on behavioural privacy*. *Comput Law Secur Rev* 2014b;30(3):286–305. PrePrint at <http://www.rogerclarke.com/SOS/Drones-BP.html#RN>.
- Clarke R. *Appropriate regulatory responses to the drone epidemic*. *Comput Law Secur Rev* 2016;32(1):152–5. (Jan-Feb) PrePrint at <http://www.rogerclarke.com/SOS/Drones-PAR.html>.
- Clarke R. *The opportunities afforded By RegTech: a framework for regulatory information systems*. Xamax Consultancy Pty Ltd; 2018a Working Paper, at <http://www.rogerclarke.com/EC/RTF.html>.
- Clarke R. *Guidelines for the responsible application of data analytics*. *Comput Law Secur Rev* 2018b;34(3):467–76, <https://doi.org/10.1016/j.clsr.2017.11.002>. PrePrint at <http://www.rogerclarke.com/EC/GDA.html>.
- Clarke R, Bennett Moses L. 'The regulation of civilian Drones'. *Comput Law Secur Rev* 2014;30(3):263–85. PrePrint at <http://www.rogerclarke.com/SOS/Drones-PS.html>.
- Clarke R, Greenleaf GW. 'Dataveillance regulation: a research framework'. *J Law Inf Sci* 2018;25(1). PrePrint at <http://www.rogerclarke.com/DV/DVR.html>.
- Connolly C. (2008) 'Trustmark Schemes Struggle to Protect Privacy' *Galexia*, at http://www.galexia.com/public/research/assets/trustmarks_struggle_20080926/trustmarks_struggle_public.pdf.
- Connolly C, Greenleaf G, Waters N. 'Privacy self-regulation in crisis? TRUSTe's 'deceptive' practices' *Priv Laws Bus Int Rep* 2014;132:13–17. at <http://www.austlii.edu.au/au/journals/UNSWLRS/2015/8.pdf>.
- Devlin H. 'Do no harm, don't discriminate: official guidance issued on robot ethics'. *The Guardian* 2016. at <https://www.theguardian.com/technology/2016/sep/18/official-guidance-robot-ethics-british-standards-institute>.
- DMV-CA. *Autonomous vehicles in California*. California Department of Motor Vehicles; 2018, at <https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/bkgd>.
- Drahos P, editor. *Regulatory theory: foundations and applications*. ANU Press; 2017, at <http://press.anu.edu.au/publications/regulatory-theory/download>.
- Dreyfus HL. *What computers still can't do: a critique of artificial reason*. MIT Press; 1992.
- EC (2018) 'Statement on artificial intelligence, Robotics and 'Autonomous' Systems' European Group on Ethics in Science and New Technologies' European Commission, at http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf.
- EDPS. *Artificial intelligence, robotics, privacy and data protection*. European Data Protection Supervisor; 2016, at https://edps.europa.eu/sites/edp/files/publication/16-10-19_marrakesh_ai_paper_en.pdf.
- FAA. *Emergency order of prohibition*. Federal Aviation Administration; 2019, at https://www.faa.gov/news/updates/media/Emergency_Order.pdf.
- FLI. *Asilomar AI principles*. Future of Life Institute; 2017, at <https://futureoflife.org/ai-principles/?cn-reloaded=1>.
- GEFA (2016) 'Position on robotics and AI' The Greens/European Free Alliance Digital Working Group, at <https://juliareda.eu/wp-content/uploads/2017/02/Green-Digital-Working-Group-Position-on-Robotics-and-Artificial-Intelligence-2016-11-22.pdf>.
- Grabowsky P. (2017) 'Meta-regulation' Chapter 9 in Drahos, pp. 149–161, at <http://press-files.anu.edu.au/downloads/press/n2304/pdf/ch09.pdf>.
- Gunningham N. & Sinclair D. (2017) 'Smart Regulation', Chapter 8 in Drahos, pp. 133–148, at <http://press-files.anu.edu.au/downloads/press/n2304/pdf/ch08.pdf>.
- Gunningham N., Grabosky P, & Sinclair D. (1998) 'Smart regulation: designing environmental policy' Oxford University Press.
- Gupta A, Lad L. *Industry self-regulation: an economic, organizational, and political analysis*. *Acad Manag Rev* 1983;8(3):416–25.
- Hardin G. *The tragedy of the commons*. *Science* 1968;162:1243–8. at <http://cescos.fau.edu/gawliklab/papers/HardinG1968.pdf>.
- Hardin (1994) 'Postscript: the tragedy of the unmanaged commons' *Trends Ecol Evol* 9, 5 (May) 199
- Hepburn G. (2006) 'Alternatives to traditional regulation' OECD Regulatory Policy Division, undated, at <http://www.oecd.org/gov/regulatory-policy/42245468.pdf>.
- Hirano. *AI R&D guidelines*. Proceedings of the OECD conference on AI developments and applications, 2017. <http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-hirano.pdf>.
- HOL. *AI in the UK: ready, willing and able? Select committee on artificial intelligence*. House of Lords; 2018, at <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>.
- Holder C, Khurana V, Harrison F, Jacobs L. 'Robotics and law: key legal and regulatory implications of the robotics age (Part I of II)'. *Comput Law Secur Rev* 2016;32(3):383–402.
- Hosein G, Tsavios P, Whitley E. 'Regulating architecture and architectures of regulation: contributions from information systems'. *Int Rev Law Comput Technol* 2003;17(1):85–98.
- HTR. *Robots: no regulatory race against the machine yet*. The Regulatory Institute; 2017, at <http://www.howtoregulate.org/robots-regulators-active/#more-230>.

- HTR. Report on artificial intelligence: part I – the existing regulatory landscape. 2018a. at http://www.howtoregulate.org/artificial_intelligence/.
- HTR. Report on artificial intelligence: part II – outline of future regulation of AI. 2018b. at <http://www.howtoregulate.org/ai/part2/#more-327>.
- HTR. Research and technology risks: part IV – a prototype regulation. The Regulatory Institute; 2018c, at <http://www.howtoregulate.org/prototype-regulation-research-technology/#more-298>.
- IBM. Everyday ethics for artificial intelligence; 2018, at <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>.
- ICO (2017) 'Big data, artificial intelligence, machine learning and data protection' UK Information Commissioner's Office, Discussion Paper v.2.2, at <https://ico.org.uk/for-organisations/guide-to-data-protection/big-data/>.
- IEEE. Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems (A/IS). IEEE; 2017 Version 2, at http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.
- ITIC. AI policy principles. Information Technology Industry Council; 2017 undated, at <https://www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf>.
- Jordan A, Wurzel RKW, Zito A. The rise of 'new' policy instruments in comparative perspective: has governance eclipsed government? *Polit Stud* 2005;53(3):477–96.
- Leenes R, Lucivero F. 'Laws on robots, laws by robots, laws in robots: regulating robot behaviour by design'. *Law, Innov & Technol* 2014;6(2):193–220.
- Lessig L. Code and other laws of cyberspace. Basic Books; 1999.
- Martins LEG, Gorschek T. 'Requirements engineering for safety-critical systems: a systematic literature review'. *Inf Softw Technol J* 2016;75:71–89.
- Maschmedt A, Searle R. Driverless vehicle trial legislation – state-by-state. King & Wood Malleson; 2018, at <https://www.kwm.com/en/au/knowledge/insights/driverless-vehicle-trial-legislation-nsw-vic-sa-20180227>.
- Moore TT, Dhillon G. 'Do privacy seals in e-commerce really work?'. *Commun ACM* 2003;46(12):265–71.
- MS. Microsoft AI principles. Microsoft; 2019 undated, at <https://www.microsoft.com/en-us/ai/our-approach-to-ai>.
- Newcomer E. What Google's AI principles left out: we're in a golden age for hollow corporate statements sold as high-minded ethical treatises. Bloomberg; 2018, at <https://www.bloomberg.com/news/articles/2018-06-08/what-google-s-ai-principles-left-out>.
- Ostrom E. Coping with tragedies of the commons. *Annu Rev Polit Sci* 1999;2:493–535. at <https://www.annualreviews.org/doi/full/10.1146/annurev.polisci.2.1.493>.
- Palmerini E. et al. (2014). 'Guidelines on regulating robotics delivery' EU Robolaw Project, at http://www.robolaw.eu/RoboLaw_files/documents/robolaw_d6.2_guidelinesregulatingrobotics_20140922.pdf.
- Parker C. The open corporation: effective self-regulation and democracy. Cambridge University Press; 2002.
- Parker C. Meta-regulation: legal accountability for corporate social responsibility?. In: McBarnet D, Voiculescu A, Campbell T, editors. *The New Corporate Accountability: Corporate Social Responsibility and the Law*; 2007.
- PC. 'Rethinking regulation' report of the taskforce on reducing regulatory burdens on business. Productivity Commission; 2006, at <http://www.pc.gov.au/research/supporting/regulation-taskforce/report/regulation-taskforce2.pdf>.
- Pichai S. AI at Google: our principles. Google Blog; 2018, at <https://www.blog.google/technology/ai/ai-principles/>.
- PoAI. Our work (thematic pillars). Partnership on AI; 2018, at <https://www.partnershiponai.org/about/#pillar-1>.
- Robertson LJ, Abbas R, Alici G, Munoz A, Michael K. Engineering-based design methodology for embedding ethics in autonomous robots. *Proc. IEEE* 2019;107(3):582–99. at <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8620254>.
- Scott C. Regulation in the age of governance: the rise of the post-regulatory state. In: Jordana J J, Levi-Faur D, editors. *The politics of regulation*. Edward Elgar; 2004.
- Schellekens M. 'Self-driving cars and the chilling effect of liability law'. *Comput Law Secur Rev* 2015;31(4):506–17.
- Scherer MU. Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. *Harvard J Law Technol* 2016;29(2):354–400.
- Smith A. In: Strahan W, Cadell T, editors. *The wealth of nations*. London; 1776.
- Stiglitz J. 'Government failure vs. market failure' principles of regulation. Initiative for Policy Dialogue; 2008 Working Paper #144, at http://policydialogue.org/publications/working_papers/government_failure_vs_market_failure/.
- TvH. Telstra Corporation Limited v Hornsby Shire Council, NSWLEC 133. esp. paras. 113–83, at <http://www.austlii.edu.au/au/cases/nsw/NSWLEC/2006/133.htm>
- UGU. Top 10 principles for ethical AI. UNI Global Union; 2017, at http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf.
- Vellinga NE. 'From the testing to the deployment of self-driving cars: legal challenges to policymakers on the road ahead'. *Comput Law Secur Rev* 2017;33(6):847–63.
- Villani C. (2018) 'For a meaningful artificial intelligence: towards a French and European strategy' Part 5 – what are the ethics of AI?. Mission for the French Prime Minister, pp. 113–130, at https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.
- Wachter S, Mittelstadt B, Floridi L. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *Int Data Priv Law* 2017;7(2):76–99 <https://academic.oup.com/idpl/article/7/2/76/3860948>.
- Weizenbaum J. *Computer power and human reason*. Penguin: W.H. Freeman & Co.; 1976. p. 1984.
- WH. 'Summary of the 2018 white house summit on artificial intelligence for American industry' Office of Science and Technology Policy. White House; 2018, at <https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf>.
- Williamson OE. 'Transaction-cost economics: the governance of contractual relations'. *J Law Econ* 1979;22(2):233–61.
- Wingspread (1998). Wingspread statement on the precautionary principle, at <http://sehn.org/wingspread-conference-on-the-precautionary-principle/>.