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

Version of 14 October 2020

For AiCE 2020, UniSA, Adelaide, November 2020

Roger Clarke **

Abstract

During the last few years, large numbers of documents have been published that proclaim themselves to be ethical guidelines for the responsible use of powerful new forms of inferencing. A project was undertaken to examine 30 such documents and infer from them a super-set of principles. The comprehensiveness of each individual document was then assessed by checking its coverage of that super-set. This paper applies the resulting data to address the question of whether any of the 30 guidelines appears tenable as an element in the protection of people against harm arising from data analytics. It was found that almost all such guidelines, and especially those published by corporations and industry associations, fall dramatically short of the needs of those affected by irresponsible application of data analytics.

Keywords: data science, neural networks, algorithmic bias, co-regulation

1 Introduction

A great many sets of ethical guidelines have been promoted during the period 2015-2020, which purport to protect people against the harms that may arise from the application of data analytics to big data. Particular concern exists about the forms of data analytics associated with the Machine Learning branch of Artificial Intelligence (AI/ML).

For ethical guidelines to provide a satisfactory basis for addressing the risks, they would need to fulfil a considerable number of conditions. The guidelines would need to be comprehensive, sufficiently detailed, clear, and articulated in a manner that enables their interpretation in the intended contexts.

Further, guidelines in themselves achieve nothing. They need to be actually applied, and quality assurance mechanisms need to be in place prior to, during and after the fact. Critical elements that must be encompassed include formal obligations on organisations to comply with the guidelines, complaints processes, investigational powers and resources, access to redress, sanctions against misbehaviour, and powers, resources and commitment to impose those sanctions. A comprehensive set of criteria for the design and evaluation of a regulatory regime for AI/ML was proposed in Clarke (2019a). A guidelines-based strategy might be implemented through one or more of the available regulatory approaches. It is useful to distinguish six regulatory approaches, grouped under the headings of formal law, self-governance and systemic governance (Clarke 2019c).

This paper considers one key, underlying question. It examines the extent to which individual sets of ethical guidelines are sufficiently comprehensive in their coverage. It commences by briefly reviewing the areas of big data, data analytics and AI/ML, and defining terms. It then summarises the risk factors that need to be managed if the technology's potential benefits are to be achieved without undue harms arising.

The main body of the paper draws on a study of 30 sets of ethical guidelines, from which a super-set of 50 Principles for Responsible AI was compiled. The particular contribution of the present paper is an assessment of the extent to which the super-set of 50 derived principles is embodied in each of the 30 sets, and hence whether any of these sets is tenable as a means of protecting against harm arising from applications of AI/ML. The sets of guidelines that were produced by corporations, industry associations and professional bodies, on the one hand, are contrasted against those put forward by academics, government agencies and non-government organisations (NGOs) on the other.

2. Data Analytics and AI/ML

In recent years, several fields have drifted towards one another. The moderate enthusiasm engendered by 'data warehousing' and 'data mining' in the 1990s has been replaced by unbridled euphoria about 'big data' and 'data analytics'. The characteristics of big data were originally depicted as 'volume, velocity and variety' (Laney 2001), then 'value' was added, and finally 'veracity' put in an appearance in Schroeck et al. (2012). A strong form of the big data claim is that "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. ... [F]aced with massive data, [the old] approach to science -- hypothesize, model, test -- is becoming obsolete. ... Petabytes allow us to say: 'Correlation is enough'" (Anderson 2008).

The term 'data analytics' has been used in technical disciplines for many years to encompass the techniques applied to data collections in order to draw inferences. Previous decades of work in statistical sciences, operations research, management science and data mining have delivered a very substantial array of analytical tools, and more are being developed. Chen et al. (2012) distinguishes two phases to date, characterising generation 1.0 as "data management and warehousing, reporting, dashboards, ad hoc query, search-based [business intelligence (BI)], [online analytical processing (OLAP)], interactive visualization, scorecards, predictive modeling, and data mining" (p. 1166) and generation 2.0, evident since the early 2000s, as being associated with web and social media analytics, including sentiment analysis, and associated-rule and graph mining, much of which is dependent on semantic web notions and text analysis tools (pp. 1167-68). The authors anticipated a generation 3.0, to cope with mobile and sensor-generated data. The term 'fast data' has since emerged, to refer to near-real-time analysis of data-streams (e.g. Pathirage & Plale 2015).

The term Artificial Intelligence (AI) was coined in a project proposal in 1955, 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" (McCarthy et al. 1955, emphasis added). What is usefully dubbed the 'Simple Simon' hypothesis was continually re-asserted throughout the second half of the 20th century, e.g. Simon (1960, 1996). The field has been characterised by successions of modest progress, excessive enthusiasm, failure, and 'AI winters' during which lack of credibility resulted in limited funding. A sceptical view is that, 65 years after the AI project was initiated, its primary tenet remains a mere conjecture, and hence AI's winters can be reasonably expected to continue to follow its summers.

Many attempts have been made to distill out the sense in which the juxtaposition of the two words is to be understood. Conventionally (Albus 1991, Russell & Norvig 2009, McCarthy 2007):

> *Intelligence is exhibited by an artefact if it:*
>
> *(1) evidences perception and cognition of relevant aspects of its environment;*
>
> *(2) has goals; and*
>
> *(3) formulates actions towards the achievement of those goals.*

The most common technique in the 'machine learning' (ML) branch of AI is 'artificial neural networks' (ANN). Although ANN has been in use since the 1950s, limited progress was made until sufficiently powerful processors became available, beginning in the late 1980s and particularly during the last decade. For the categories of researchers creating and applying AI/ML, this happily coincided with the rash of 'big data' lying around waiting to be fed in.

Neural nets involve a set of nodes (each of which is analogous to the biological concept of a neuron), with connections or arcs among them, referred to as 'edges'. Each connection has a 'weight' associated with it. Each node performs computations based on incoming data, and may as a result adapt its internal state, in particular the weight associated with each arc, and may pass output to one or more other nodes. A neural net has to be 'trained'. This is done by selecting a training method (or 'learning algorithm'), selecting a 'training-set' from the available data, and feeding the training-set to the network in order to load up the initial set of weights on the connections between nodes.

Enthusiasts see great prospects in neural network techniques, e.g. "There has been a number of stunning new results with deep-learning methods ... The kind of jump we are seeing in the accuracy of these systems is very rare indeed" (Markoff 2012). They claim that noisy and error-ridden data presents no problems, provided that there's enough of it. They also claim that the techniques have a wide range of application areas. Sceptics, on the other hand, perceive that the techniques' proponents overlook serious weaknesses (Marcus 2018).

3. Risk Factors in Data Analysis and AI/ML

Unlike previous techniques for developing software, neural networking approaches need not begin with active and careful modelling of a real-world problem-solution, problem or even problem-domain. Rather than comprising a set of entities and relationships that mirrors what the analyst has determined to be the key elements and processes of a real-world system, a neural network model may be merely lists of input variables and output variables (and, in the case of 'deep' networks, one or more levels of intermediating variables). To the extent that a model exists, in the sense of a representation of the real world, it is implicit rather than express. The weights imputed for each connection reflect the characteristics firstly of the training-set that was fed in, and secondly of the particular learning algorithm that was imposed on the training-set.

AI/ML in effect treats empiricism as entirely dominating theory. This inverts what has hitherto been the norm in the sciences, where theory and observation are inherently intertwined, with each dependent on the other. The quite fundamental question of the legitimacy or otherwise of pure empiricism combines with questions about the selectivity, accuracy and compatibility of the data to give rise to deep uncertainty about AI/ML's affinity with the real world circumstances to which it is applied.

Inferences drawn using neural networking inevitably reflect errors and biasses inherent in the implicit model, in the selection of real-world phenomena for which data was created, in the selection of the training-set, and in the particular learning algorithms used to develop the application. Means are necessary to assess the quality of the implicit model, of the data-set, of the data-item values, of the data-item values' correspondence with real-world phenomena, of the training-set, of the learning algorithm, and of the compatibility among them all, and to validate the inferences both logically and empirically. Unless and until those means are found, and are routinely applied, AI/ML and neural nets need to be regarded as unproven techniques that harbour considerable dangers to the interests of organisations, their stakeholders and individuals affected by inferences drawn from the data.

The author, in a series of jointly- and individually-authored works, has previously catalogued and assessed the issues (Wigan & Clarke 2013, Clarke 2016a, Clarke 2016b, Clarke 2018, Clarke & Taylor 2018 and Clarke 2019a). These analyses drew on, among other sources, Scherer (2016, esp. pp. 362-373), Yampolskiy & Spellchecker (2016) and Duursma (2018). This work culminated in a single sentence summary (Clarke 2019a, p.426):

> *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*

and in an examination of the following five underlying factors (pp.426-429):

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

Attention was drawn in Clarke (2019b) to the importance of undertaking multi-stakeholder risk assessment, with the following propositions put forward:

1 The responsible application of AI is only possible if stakeholder analysis is undertaken, firstly, in order to identify the categories of entities that are or may be affected by the particular project, and, secondly, to gain insight into those entities' needs and interests

2 The risk assessment processes commonly undertaken within an organisation reflect primarily the organisation's own interests, and hence are not adequate to reflect the interests of other stakeholders

3 The responsible application of AI depends on risk assessment processes being conducted from the perspective of each stakeholder group, to complement that undertaken from the organisation's perspective

Further, because risk assessment is merely the problem analysis phase, the requirements identified during that phase must be addressed in the risk management phase, culminating in an appropriate design. The design must then be carried through to implementation, and audit undertaken that the requirements have been satisfied.

---

4. Comprehensive Ethical Guidelines for AI/ML

The widespread concern about the impact of irresponsible activities in relation to big data, data analytics, AI generally and AI/ML in particular has manifested itself in a flurry of documents published by many different authors and organisations during the decade 2010-20.

In an earlier project undertaken by the author, a set of guidelines for responsible data analytics was proposed, reflecting the literature critical of various aspects of the big data movement, notably Bollier (2010), boyd & Crawford (2011), Lazer et al. (2014), Metcalf & Crawford (2016), King & Forder (2016) and Mittelstadt et al. (2016). These Guidelines are presented in Clarke (2018). In Clarke & Taylor (2018), the individual guidelines were mapped to a conventional business process for data analytics projects.

A more recent project drew on the wide array of documents that offered guidance for the responsible application of analytical techniques to large volumes of data. The research conducted in support of a series of articles on AI topics (Clarke 2019a, b, c) identified many previously-published sets of principles. A suite of 30 diverse sets of guidelines was selected. Exemplars were selected from those available in April 2019, and were favoured for inclusion primarily because of their apparent significance, based in particular on the sponsoring organisation, the substance or quantum of the contribution and/or the existence and coherence of rationale underlying their composition. Only documents in English were included – a pragmatic choice, but inevitably giving rise to geographic and cultural bias. Where multiple versions existed, generally only the most recent was included.

The set comprised 8 sets of general ethical principles applied to technology-rich contexts (excerpted in Clarke 2018b), and 22 documents whose focus was variously on AI generally, robotics, autonomous and intelligent systems, AI/ML specifically, and algorithmic transparency in particular (in Clarke 2018c). From those 30, a super-set of ethical guidelines was generated (Clarke 2019b). These comprised 50 principles that were evident in the sample, which were grouped under the 10 headings in Table 1:

Table 1: The 10 Themes Underlying the 50 Principles for Responsible AI

1	Assess Positive and Negative Impacts and Implications
2	Complement Humans
3	Ensure Human Control
4	Ensure Human Safety and Wellbeing
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 Obligations
10	Enforce, and Accept Enforcement of, Liabilities and Sanctions

Two other teams undertook similar studies in parallel with this author's project, with all three publications appearing in the space of nine months. A review of Zeng et al. (2019) and the source-documents that those authors analysed did not suggest the need for any re-framing or re-phrasing of the 10 Themes or the 50 Principles. A review of Jobin et al. (2019) found that, despite its formalised textual analysis of 84 documents, it omits more than a dozen of the 50 that appear in the set generated by the present study. The Jobin study did not detect any requirements relating to impact assessment, justification, stakeholder consultation and proportionality (9 principles), nor to complementariness to humans (2), and missed most aspects relating to human control of AI-based technology (5 principles).

The 30 documents drawn on in this author's study were published variously by governmental organisations (9), non-government organisations (6), academics (4), corporations and industry associations (7), joint associations (2) and professional associations (2). The scope was broad in geographical terms, with 11 from the USA, 9 from Europe, 6 from the Asia-Pacific, and 4 global in nature. In general, only the most recent version of documents was used, except in the case of the European Commission, whose late 2018 draft and early 2019 final versions evidence material differences and were both significant.

The process adopted in order to express the super-set of principles is described in Clarke (2019b). Decreasing returns to scale suggested that the flex-point had been passed and that little further benefit would be gained from extending the set of source-documents. A further analysis was prepared, in which the 50 principles are cross-referenced back from the super-set to the source-documents (Clarke 2019d, Clarke 2020a).

There are many ways in which the consolidated super-set of 50 Principles for Responsible AI can be applied. One way to use them, for example, is as a standard against which particular guidelines, codes and statutory requirements can be evaluated. The OECD published guidelines after the completion of the original study. When evaluated against the 50 Principles, the OECD document scored only 40%, despite the array of prior guidelines published long before the composition of that document (Clarke 2019e). A further document that can be analysed in this manner is the set of 72 requirements in Sherpa (2019b).

The question also arises as to how the super-set of 50 Principles might be themselves evaluated. One approach would be to consider each of the derived 50 Principles from various ethical viewpoints. Beyond the normative issues lie pragmatic questions, in particular: are the 50 Principles understandable and usable, and are they effective in achieving their purposes?

To achieve deep understanding in particular contexts, it is highly desirable to conduct *ex post facto* case study analysis and/or contemporaneous, embedded action research. This requires the acquisition and application of considerable financial resources and expertise; but see Sherpa (2019a). It is also challenging to find research partners. Many projects are more *ad hoc* and less well-developed than media reports suggest. Many fail, and their sponsors may wish to obscure the fact that they have failed. Many projects sail close to the wind in legal terms. Many are conducted surreptitiously because the sponsor risks drawing fire from stakeholders should their existence become known. Many projects are conducted behind closed doors because they have potential competitive value, and many more are obfuscated using potential competitive value as a pretext.

Given the resource-intensive nature of case study and action research, only a modest number of such projects can be performed. Deep empirical work accordingly needs to be complemented by approaches that are broader and shallower in nature. One approach is to analyse vignettes, by which is meant mini-case studies that are inspired by empirically-based examples but are filled out with speculative-but-tenable details. Where the analysis is not merely a point-in-time snapshot but proceeds along a timeline, the vignette is appropriately referred to as a scenario (Wack 1985, Schwartz 1991). Examples of the approach to big data analytics generally are provided in Clarke (2016a), and a discussion of the method is in Clarke (2015). See also Wigan & Clarke (2013).

In the final section of this article, a contribution is made to the application of the super-set of 50 Principles. The approach adopted is to compare the individual sources against the super-set, both individually and by grouping them on the basis of the category of entity that published them.

5. The Adequacy of Individual Sets of Ethical Guidelines

This section reports on the insights provided by the study into the potential effectiveness of each of the 30 source-documents used to create the super-set. Documents were scored liberally, in order to avoid unreasonably low scoring due to only partial or qualified coverage. This was particularly important in the case of those principles that have a degree of complexity or richness.

Some commonalities exist within the set of source documents. Overall, however, the main impression is of sparseness, with remarkably limited consensus. Despite the liberal scoring, each of the 30 documents reflected on average only about 10 of the 50 Principles. Moreover, only 3 source-documents achieved moderately high scores - 46% for Marx (1998), 58% for EC (2018) and 74% for EC (2019).

Apart from these three outliers, the mean was a very low 17%, range 8%-34%. When assessed from the other direction, each of the 50 Principles was reflected in, on average, just over 6 of the 30 documents (21%), range 1-15 of 30 but with one outlier, which was:

- 'Ensure people's physical health and safety ('nonmaleficence')' (4.1) in 24/30.

There was limited consensus even on quite fundamental requirements. For example:

- 'Ensure people's wellbeing ('beneficence')' (4.3) was stipulated in only 14/30;

- 'Ensure that effective remedies exist ...' (9.2) in 14/30;
- 'Conduct impact assessment ...' (1.4) in only 11/30 documents.

The purpose in this paper is to assess the comprehensiveness of individual instances of the 30 documents, and of sub-sets of the 30 based on the category of the entity that published them.

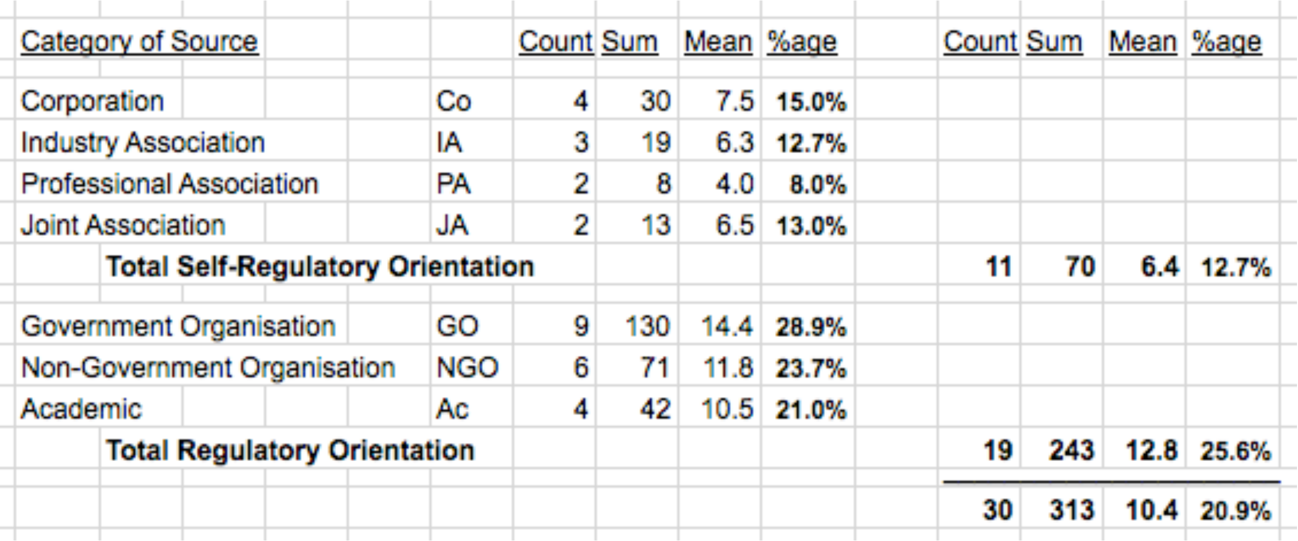
It was postulated that categories of entity were likely to be oriented towards different approaches to regulation, as follows:

- stronger orientation towards formal regulation would be shown by government organisations (9), NGOs (6) and academics (4), totalling 19 of the 30; whereas
- self-regulatory approaches would be favoured by corporations (4), industry associations (3), professional associations (2), and joint associations funded by industry (2), totalling 11.

Each set of guidelines was compared point-by-point against the super-set of 50 principles, and a score derived. Needless to say, there is ample scope for debate about the scoring method. However, because even partial and qualified coverage was scored, the method errs on the side of exaggerating conformance with the super-set rather than allocating unduly low scores. For the purposes being addressed in this paper, the scores for individual organisations were then combined into their categories, as outlined above.

In Table 2, the category summaries are shown. The full spreadsheet is available (Clarke 2020a). Across the sample of 30 sources, the average number of the 50 Principles they evidenced was 10.4 (21%). The categories of organisations postulated as having a stronger orientation towards formal regulation averaged 12.8 (26%), whereas those more likely to favour self-regulation averaged only 6.4 (13%).

Table 2: Number of the 50 Principles Evidenced in the Source-Documents



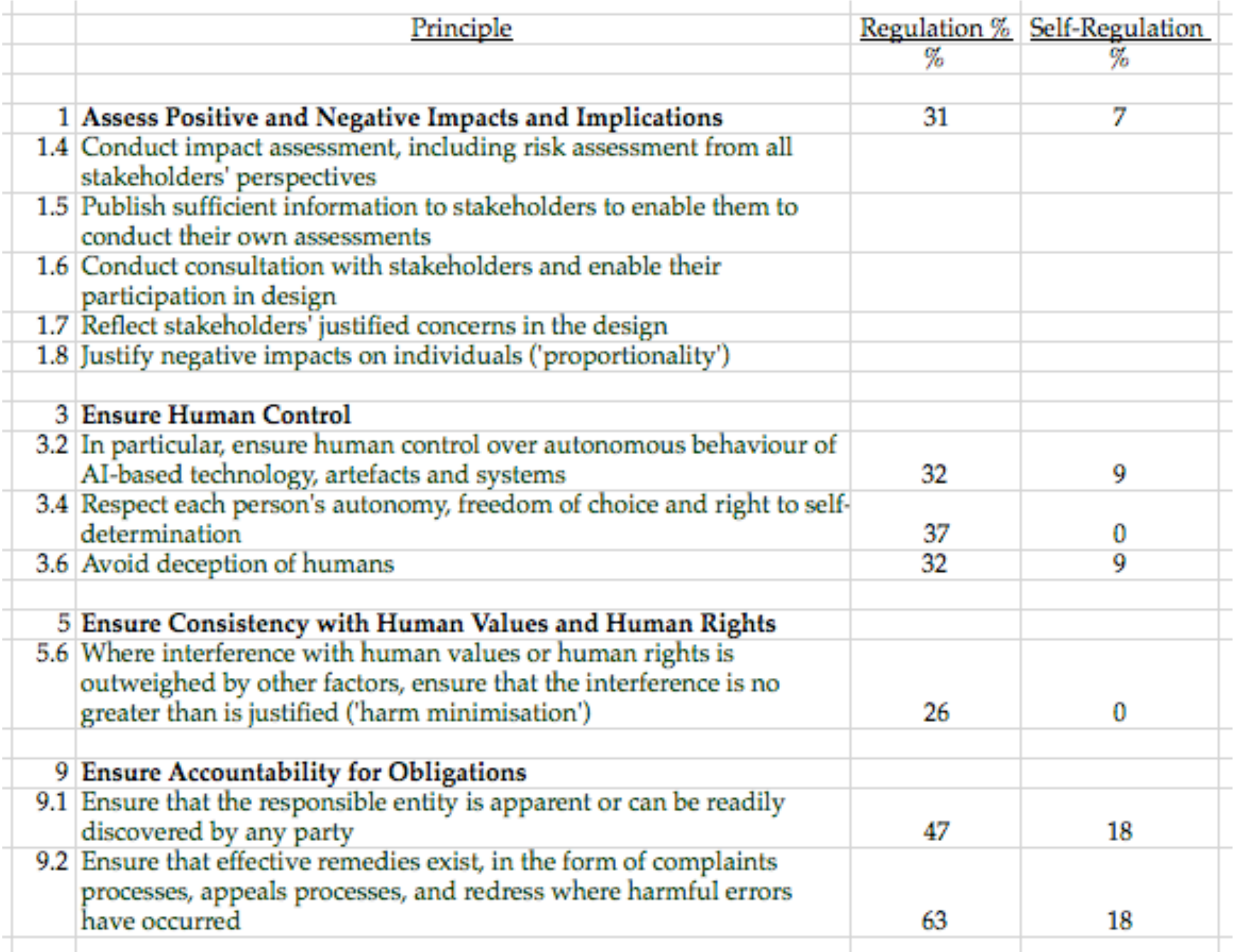
Category of Source		Count	Sum	Mean	%age	Count	Sum	Mean	%age
Corporation	Co	4	30	7.5	**15.0%**				
Industry Association	IA	3	19	6.3	**12.7%**				
Professional Association	PA	2	8	4.0	**8.0%**				
Joint Association	JA	2	13	6.5	**13.0%**				
Total Self-Regulatory Orientation						**11**	**70**	**6.4**	**12.7%**
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%**

Among the regulatory group, the lowest four scores were 4 and 5, but the highest were 17, 20, 28 and 37, respectively GEFA (2016), Marx (1998), EC (2018) and EC (2019). Of the 19, 6 more scored above the overall mean of 10.4 out of 50. On the other hand, only 2 of the 11 in the self-regulatory category exceeded the overall mean, and both of those only just, at 11.

The least-worst self-regulatory sub-category was Corporations, which scored 5, 7, 7 and 11 (range 10-22%, respectively Microsoft, IBM, Google and Sony). On the other hand, all four of them issued their documents in 2018-19. They therefore had plenty of opportunity to review other organisations' ethical guidelines, and hence it is reasonable to infer that they actively chose to fail.

A number of factors appeared to be reasonably mainstream among the 'regulation' group, yet were barely evident in the guidelines produced by organisations in the 'self-regulation' category. It would be useful to know which of those Principles, listed in Table 3, are actively opposed by proponents of the application of AI/ML techniques, as distinct from merely overlooked.

Table 3: Principles Subject to Major Divergence Between the Sub-Samples



	Principle	Regulation %	Self-Regulation
		%	%
1	**Assess Positive and Negative Impacts and Implications**	31	7
1.4	Conduct impact assessment, including risk assessment from all stakeholders' perspectives		
1.5	Publish sufficient information to stakeholders to enable them to conduct their own assessments		
1.6	Conduct consultation with stakeholders and enable their participation in design		
1.7	Reflect stakeholders' justified concerns in the design		
1.8	Justify negative impacts on individuals ('proportionality')		
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
9	**Ensure Accountability for Obligations**		
9.1	Ensure that the responsible entity is apparent or can be readily discovered by any party	47	18
9.2	Ensure that effective remedies exist, in the form of complaints processes, appeals processes, and redress where harmful errors have occurred	63	18

6. Conclusions

A super-set of Principles for Responsible AI was established through inspection of 30 significant exemplars. In comparison with that yardstick, one document is head and shoulders above the others (EC 2019), although with only a liberally-scored 74% of the derived, comprehensive super-set.

The other 18 organisations postulated as being oriented towards formal regulation are scattered across the score-range from 8% to 56%. Scores for organisations postulated as leaning towards self-regulation, on the other hand, fell in the range 4% to 22%.

For self-regulatory approaches based on ethical guidelines to succeed in protecting against the harms arising from AI/ML, a considerable number of conditions would need to be fulfilled. This study shows that self-regulation fails on the basic question of whether the guidelines that organisations set for themselves and their members are sufficiently comprehensive. This lends considerable weight to the expectation deriving from regulatory theory that dependence on self-regulation would be futile, assuming that the objective were to manage public risk.

It is challenging to draft and enact formal laws to regulate a new, poorly-understood and heavily-hyped technology. Considerable risk therefore exists that enthusiastic marketing may see a particular variant of technological determinism win through, at great risk to user-organisations and the public (Wyatt 2008). This is the converse of the 'precautionary principle', whose weak form asserts 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 (Wingspread 1998). The threats arising from irresponsible application of AI/ML appear to be sufficient to justify the strong form, to date seldom applied other than in environmental contexts (TvH 2006):

> *"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"*

Various approaches are possible to achieve protections against harm from AI/ML. In Clarke (2020b), it is argued that self-regulation is illusory, and statutory law a blunt weapon that lacks the flexibility needed to cope with the kind of new and changing but potentially impactful technology for which the precautionary principle was devised. An alternative approach holds promise, in the form of a co-regulatory framework. Key features are as follows. For a fuller presentation of the approach, see Clarke (2019c):

- a statutory framework to ensure that the regulatory objectives are achieved;
- a delegated authority such as a commission or independent regulatory agency, with supervisory powers;
- a code institution required to conduct consultative processes, including all stakeholders, to articulate and operationalise the abstractly-expressed objectives;
- code development resources;
- enforcement powers and resources; and
- enforcement obligations.

On the basis of the evidence arising from this study, the author contends that, in the absence of this level of regulatory action, ethical guidelines of any kind are primarily of value to corporations and government agencies in permitting them to proceed apace with dangerous applications of inherently dangerous technology.

Unsurprisingly, ethical guidelines issued by organisations motivated by the idea of self-regulation are of no value at all to the various publics that will suffer the harms arising from AI/ML. Disappointingly, however, almost all ethical guidelines issued by academics, non-government organisations and government organisations also fail badly against the derived, comprehensive super-set of guidelines. The sole instance that could be considered as being even moderately comprehensive is the EU's 'Ethics Guidelines for Trustworthy AI', of April 2019; and that will require a considerable body of law, resources and practices around it before it actually delivers protections.

Acknowledgements

This paper draws heavily on a series of previous publications and Working Papers, which are cited above and included in the Reference List. It benefited from useful suggestions by reviewers.

Reference List

Albus J. S. (1991) 'Outline for a theory of intelligence' IEEE Trans. Systems, Man and Cybernetics 21, 3 (1991) 473-509, at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.410.9719&rep=rep1&type=pdf

Anderson C. (2008) 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete' Wired Magazine 16:07, 23 June 2008, at http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory

Bollier D. (2010) 'The Promise and Peril of Big Data' The Aspen Institute, 2010, at http://www.ilmresource.com/collateral/analyst-reports/10334-ar-promise-peril-of-big-data.pdf

boyd D. & Crawford K. (2011) `Six Provocations for Big Data' Proc. Symposium on the Dynamics of the Internet and Society, September 2011, at http://ssrn.com/abstract=1926431

Chen H., Chiang R.H.L. & Storey V.C. (2012) 'Business Intelligence and Analytics: From Big Data to Big Impact' MIS Quarterly 36, 4 (December 2012) 1165-1188, at http://ai.arizona.edu/mis510/other/MISQ%2520BI%2520Special%2520Issue%2520Introduction%2520Chen-Chiang-Storey%2520December%25202012.pdf

Clarke R. (2015) 'Quasi-Empirical Scenario Analysis and Its Application to Big Data Quality' Proc. 28th Bled eConference, Slovenia, 7-10 June 2015, PrePrint at http://www.rogerclarke.com/EC/BDSA.html

Clarke R. (2016a) 'Big Data, Big Risks' Information Systems Journal 26, 1 (January 2016) 77-90, PrePrint at http://www.rogerclarke.com/EC/BDBR.html

Clarke R. (2016b) 'Quality Assurance for Security Applications of Big Data' Proc. EISIC'16, Uppsala, 17-19 August 2016, PrePrint at http://www.rogerclarke.com/EC/BDQAS.html

Clarke R. (2018a) 'Guidelines for the Responsible Application of Data Analytics' Computer Law & Security Review 34, 3 (May-Jun 2018) 467- 476, PrePrint at http://www.rogerclarke.com/EC/GDA.html

Clarke R. (2018b) 'Ethical Analysis and Information Technology' Xamax Consultancy Pty Ltd, July 2018, at http://www.rogerclarke.com/EC/GAIE.html

Clarke R. (2018c) 'Principles for AI: A SourceBook' Xamax Consultancy Pty Ltd, July 2018, at http://www.rogerclarke.com/EC/GAIP.html

Clarke R. (2019a) 'Why the World Wants Controls over Artificial Intelligence' Computer Law & Security Review 35, 4 (Jul-Aug 2019) 423-433, at https://doi.org/10.1016/j.clsr.2019.04.006, PrePrint at http://rogerclarke.com/EC/AII.html

Clarke R. (2019b) 'Principles and Business Processes for Responsible AI' Computer Law & Security Review 35, 4 (Jul-Aug 2019) 410-422, at https://doi.org/10.1016/j.clsr.2019.04.007, PrePrint at http://rogerclarke.com/EC/AIP.html

Clarke R. (2019c) 'Regulatory Alternatives for AI' Computer Law & Security Review 35, 4 (Jul-Aug 2019) 398-409, at https://doi.org/10.1016/j.clsr.2019.04.008, PrePrint at http://rogerclarke.com/EC/AIR.html

Clarke R. (2019d) 'Responsible AI Technologies, Artefacts, Systems and Applications: The 50 Principles Cross-Referenced to the Source DocumentsXamax Consultancy Pty Ltd, 2019, at http://www.rogerclarke.com/EC/GAIF-50Ps-XRef.html

Clarke R. (2019e) 'The OECD's AI Guidelines of 22 May 2019: Evaluation against a Consolidated Set of 50 Principles' Xamax Consultancy Pty Ltd, 2019, at http://www.rogerclarke.com/EC/AI-OECD-Eval.html

Clarke R. (2020a) 'The 50 Principles for Responsible AI Cross-Referenced to the Source Documents' Xamax Consultancy Pty Ltd, August 2020, at http://www.rogerclarke.com/EC/GAI-CrossTab.xls

Clarke R. (2020b) 'A Comprehensive Framework for Regulatory Regimes as a Basis for Effective Privacy Protection' In Review, Xamax Consultancy Pty Ltd, October 2020, at http://rogerclarke.com/DV/RMPP.html

Clarke R. & Taylor K. (2018) 'Towards Responsible Data Analytics: A Process Approach' Proc. Bled eConference, 17-20 June 2018, PrePrint at http://www.rogerclarke.com/EC/BDBP.html

Duursma (2018) 'The Risks of Artificial Intelligence' Studio OverMorgen, May 2018, at https://www.jarnoduursma.nl/the-risks-of-artificial-intelligence/

EC (2018) 'Draft Ethics Guidelines for Trustworthy AI' High-Level Expert Group on Artificial Intelligence, European Commission, 18 December 2018, at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57112

EC (2019) 'Ethics Guidelines for Trustworthy AI' High-Level Expert Group on Artificial Intelligence, European Commission, April 2019, athttps://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, at https://juliareda.eu/wp-content/uploads/2017/02/Green-Digital-Working-Group-Position-on-Robotics-and-Artificial-Intelligence-2016-11-22.pdf

Jobin A., Ienca M. & Vayena E. (2019) 'The global landscape of AI ethics guidelines' Nature Machine Intelligence 1 (September 2019) 389–399, at https://doi.org/10.1038/s42256-019-0088-2

King N.J. & Forder J. (2016) 'Data analytics and consumer profiling: Finding appropriate privacy principles for discovered data' Computer Law & Security Review 32 (2016) 696-714

Laney D. (2001) `3D Data Management: Controlling Data Volume, Velocity and Variety' Meta-Group, February 2001, at http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lazer D., Kennedy R., King G. & Vespignani A. (2014) 'The Parable of Google Flu: Traps in Big Data Analysis.Ó Science 343, 6176 (March 2014) 1203-1205, at https://dash.harvard.edu/bitstream/handle/1/12016836/The%20Parable%20of%20Google%20Flu%20%28WP-Final%29.pdf

McCarthy J. (2007) 'What is artificial intelligence?' Department of Computer Science, Stanford University, November 2007, at http://www-formal.stanford.edu/jmc/whatisai/node1.html

McCarthy J., Minsky M.L., Rochester N. & Shannon C.E. (1955) 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence' Reprinted in AI Magazine 27, 4 (2006), at https://www.aaai.org/ojs/index.php/aimagazine/article/viewFile/1904/1802

Marcus G. (2018) 'Deep Learning: A Critical Appraisal', arXiv, 2018, at https://arxiv.org/pdf/1801.00631.pdf

Markoff J. (2012) 'Scientists See Promise in Deep-Learning Programs' The New York Times, 23 November 2012, at https://www.nytimes.com/2012/11/24/science/scientists-see-advances-in-deep-learning-a-part-of-artificial-intelligence.html

Marx. G.R. (1998) 'An Ethics For The New Surveillance' The Information Society 14, 3 (August 1998) 171-185, at http://web.mit.edu/gtmarx/www/ncolin5.html

Metcalf J. & Crawford K. (2016) 'Where are human subjects in Big Data research? The emerging ethics divide' Big Data & Society 3, 1 (January-June 2016) 1-14

Mittelstadt B.D., Allo P., Taddeo M., Wachter S. & Floridi L. (2016) 'The ethics of algorithms: Mapping the debate' Big Data & Society 3, 2 (July-December 2016) 1-21

Russell S.J. & Norvig P. (2009) 'Artificial Intelligence: A Modern Approach' Prentice Hall, 3rd edition, 2009

Pathirage M. & Plale B. (2015) 'Fast Data Management with Distributed Streaming SQL' arXiv preprint arXiv:1511.03935, at http://arxiv.org/abs/1511.03935

Schroeck M., Shockley R., Smart J., Romero-Morales D. & Tufano P. (2012) `Analytics : The real world use of big data' IBM Institute for Business Value / Saïd Business School at the University of Oxford, October 2012, at http://www.ibm.com/smarterplanet/global/files/se__sv_se__intelligence__Analytics_-_The_real-world_use_of_big_data.pdf

Scherer M.U. (2016) 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' Harvard Journal of Law & Technology 29, 2 (Spring 2016) 354-400

Schwartz P. (1991) 'The Art of the Long View: Planning for the Future in an Uncertain World' Doubleday, 1991

Sherpa (2019a) 'Case Study Introduction and Overview' Sherpa Project, January 2019, at https://dmu.figshare.com/ndownloader/files/14277422

Sherpa (2019b) 'Guidelines for the Ethical Use of AI and Big Data Systems' Sherpa Project, undated but apparently of December 2019, at https://www.project-sherpa.eu/wp-content/uploads/2019/12/use-final.pdf

Simon H.A. (1960) 'The Shape of Automation' reprinted in various forms, 1960, 1965, quoted in Weizenbaum J. (1976), pp. 244-245

Simon H.A. (1996) 'The Sciences of the Artificial' 3rd ed. MIT Press. 1996

TvH (2006) 'Telstra Corporation Limited v Hornsby Shire Council' NSWLEC 133 (24 March 2006), esp. paras. 113-183, at http://www.austlii.edu.au/au/cases/nsw/NSWLEC/2006/133.htm

Wack P. (1985) 'Scenarios: Uncharted Waters Ahead' Harv. Bus. Rev. 63, 5 (September-October 1985) 73-89

Wigan M.R. & Clarke R. (2013) 'Big Data's Big Unintended Consequences' IEEE Computer 46, 6 (June 2013) 46-53, PrePrint at http://www.rogerclarke.com/DV/BigData-1303.html

Wingspread (1998) 'Wingspread Conference on the Precautionary Principle' Wingspread Statement on the Precautionary Principle, January 1998, at http://sehn.org/wingspread-conference-on-the-precautionary-principle/

Wyatt S. (2008) 'Technological Determinism Is Dead; Long Live Technological Determinism' in 'The Handbook of Science and Technology Studies' (eds. Hackett E.J., Amsterdamska O., Lynch M. & Wajcman J.), MIT Press, 2008 Chapter 7, pp.165-180

Yampolskiy R.V. & Spellchecker M.S. (2016) 'Artificial Intelligence Safety and Cybersecurity: a Timeline of AI Failures' arXiv, 2016, at https://arxiv.org/pdf/1610.07997

Zeng Y., Lu E. & Huangfu C. (2019) 'Linking Artificial Intelligence Principles' Proc. AAAI Workshop on Artificial Intelligence Safety (AAAI-Safe AI 2019), 27 January 2019, at https://arxiv.org/abs/1812.04814

Author Affiliations

Roger Clarke is Principal of Xamax Consultancy Pty Ltd, Canberra. He is also a Visiting Professor associated with the Allens Hub for Technology, Law and Innovation in UNSW Law, and a Visiting Professor in the Research School of Computer Science at the Australian National University.