The EC's Proposal for Regulation of AI:
Evaluation against a Consolidated Set of 50 Principles

Review Draft of 22 August 2021

Roger Clarke **

This document is at http://www.rogerclarke.com/EC/AIP-EC21.html

---

Abstract

In April 2019, the European Commission published a High-Level Expert Group's 'Ethics Guidelines for Trustworthy AI'. Two years later, in April 2021, the European Commission announced "new rules and actions for excellence and trust in Artificial Intelligence", with the intention to "make sure that Europeans can trust what AI has to offer". The announcement stimulated some enthusiasm among the public and public interest advocacy groups, which had been very concerned about the attempts by AI industry proponents to implement dangerous technologies.

This Working Paper presents the results of an assessment of the EC's 2021 Proposal. The primary benchmark used is a set of '50 Principles for Responsible AI'. This was consolidated from a large sample of 30 such documents in mid-2019. The '50 Principles' have previously been used to assess many 'ethical guidelines', including those published by the EC in 2019. A customised assessment method was necessary, both because the EC's Proposal is the first occasion that draft legislation has been published, and the central notion of 'AI system' is defined in a manner that is not consistent with conventional usage.

The results of the assessment do not support the proposition that the EC Proposal can deliver trustworthiness – although it could have public relations value to AI proponents by misleading the public into trusting AI. The effectiveness of the EC Proposal is characterised by a wide array of features that render it almost entirely valueless as a regulatory measure.

The fatal weaknesses include scope-limitations; unconventional uses of key terms; the absence of any protections whatsoever in respect of potentially very large numbers of 'AI systems'; a single, very weak requirement imposed on a small set of categories of system that are intended to interact directly with people; pitifully weak, vague and in many cases probably unenforceable protections in respect of a limited range of what the EC refers to as 'high-risk' systems; no requirements that inferences drawn by AI systems be capable of rational explanation; and no capacity for individuals or their advocates to seek corrections and redress for faulty inferencing.

The gulf that exists between the EC's 'Ethics Guidelines' of 2019 and the EC's Proposal of 2021 suggests that pleading by the AI industry, supported by government agency and corporate users, has resulted in the Expert Group being entirely ignored, and such a substantial emphasis being placed on industry stimulation that social considerations have been relegated to the role of constraints, resulting in a largely vacuous draft statute.

Contents

1. Introduction

During the period 2015-2020, Artificial Intelligence (AI) has been promoted with particular vigour. The nature of the technologies, combined with the sometimes quite wild enthusiasm of the technologies' proponents, have given rise to a considerable amount of public concern about AI's negative impacts. In an endeavour to calm those concerns, many organisations have published lists of 'principles for responsible AI'.

In Clarke (2019), a set of 30 such publications was analysed, resulting in a consolidated set of 10 Themes and 50 Principles. When compared against that consolidation, almost all documents to date are shown to have very limited scope. As at mid-2019, the sole contender that scored respectably was that by a High-Level Expert Group of the European Commission (EC) called 'Ethics Guidelines for Trustworthy AI' (EC 2019), and even that document only scored 37/50 (74%).

Two years later, on 21 April 2021, The European Commission (EC) published a 'Proposal for a Regulation on a European approach for Artificial Intelligence' (EC 2021). The EC's own term for the proposed statute is Artificial Intelligence Act (AIA). This Working Paper reports on an evaluation of the EC's Proposal against the consolidated 50 Principles.

2. Overview of the EC Proposal

The body of the EC's Proposal is long, and comprises multiple sections:

- an Explanatory Memorandum (EM), pp. 1-16
- a Preamble (Pre), in 89 numbered paragraphs, pp. 17-38
- the proposed Regulation (Reg), in 85 numbered Articles, pp. 38-88

The EC's Proposal is stated to be "a balanced and proportionate horizontal regulatory approach to AI that is limited to the minimum necessary requirements to address the risks and problems linked to AI, without unduly constraining or hindering technological development or otherwise disproportionately increasing the cost of placing AI solutions on the market" and "a proportionate regulatory system centred on a well-defined risk-based regulatory approach that does not create unnecessary restrictions to trade, whereby legal intervention is tailored to those concrete situations where there is a justified cause for concern or where such concern can reasonably be anticipated in the near future" (EM, p.3).

The motivation is quite deeply rooted in economics in general and innovation in particular. For example the purpose is expressed as being "to improve the functioning of the internal market by laying down a uniform legal framework" (Pre(1)). Compliance with human rights and data protection laws, and public trust, are treated as constraints to be overcome, e.g. the EC's Proposal is depicted as a "regulatory approach to AI that is limited to the minimum necessary requirements ..., without unduly constraining or hindering technological development or otherwise disproportionately increasing the cost of placing AI solutions on the market" (EM1.2, p.3).

The document is formidable. The style is variously eurocratic and legalistic. The 51 pp. of Articles require careful reading and cross-referencing, and need to be read within the context of the 22 pp. of Preamble, and taking into account the 16 pp. of (legally relevant) Explanatory Memorandum. Complicating the analysis, the underlying philosophy and the origins of the features are both somewhat eclectic. Serial readings and thematic searches were complemented by reference to publications that provide reviews and critiques of the document, in particular Veale & Borgesius (2021) and Greenleaf (2021).

The EC's Proposal is for a regulatory instrument, and hence its scope is considerably broader than Principles alone. In particular, it distinguishes four segments, establishes a prohibition on one of them, and applies differential principles to two of the other four. The distinctions are summarised in Table 1. It was therefore necessary to conduct four assessments, one in respect of each segment.

The term used in respect of Categories (2)-(4) is 'AI Systems', which is defined in Article 3(1). A different term used in relation to category **(1) Prohibited: 'AI Practices'**. The term 'AI practices' appears not to be defined. It is unclear what purpose the drafters had in mind by using a distinctly different term. It is even more unclear how the term will be interpreted by organisations, and, should it ever be relevant, by regulators or the courts. It can be reasonably anticipated that the use of the term results in loopholes, which may or may not have been intended.

In addition, the details of each category embody significant qualifiers, which have the effect of negating the provisions in respect of a considerable number of such 'AI practices'.

For example, the first is subject to a series of qualifying conditions: "subliminal techniques beyond [beneath?] a person's consciousness [i] in order to [ii] materially [iii] distort a person's behaviour [iv] in a manner that causes or is likely to cause that person or another person physical or psychological harm" (Art. 5.1(a)).

Table 1: The EU Proposal's Four 'Levels of Risk'

(1) Prohibited AI Practices (Art. 5, Pre 26-69)

(a) Subliminal techniques

(b) Exploitation of vulnerabilities of the disadvantaged

(c) Social scoring ("the evaluation or classification of the trustworthiness of natural persons ...")

(d) 'Real-Time' remote biometric identification in public places for law enforcement

(2) High-Risk AI Systems (Arts. 6-7, 8-51, Annexes II-VII, Pre 27-69, 84)

A 'high-risk AI system' is defined, in a complex manner, in Art. 6 and Annex III, as:

(a) Product or product safety components – but subject to the presumably very substantial exemptions declared in Art. 2.1. It appears that the drafters assume that the 6 nominated Regulations and 2 Directives are technologically-neutral and hence magically deliver equivalent protections to those in the 'AIA' provisions; and

(b) 21 very specific categories of AI systems within 8 "areas" (Annex III).

(3) 'Certain AI Systems' (Art. 52, Pre 70), also referred to in EM s.2.3 (p.7), ambiguously, as "non-high-risk AI systems"

The categories are declared and described as follows:

- "AI systems intended to interact with natural persons" (Art. 52.1), but with particular reference to categories of customer-facing automata currently referred to as "chatbots" (EM p.3)
- "an emotion recognition system" (Art. 52.2, but referred to in EM4 on p.14 as an AI system "used to detect emotions" – which is arguably a rather different description), except where "permitted by law to detect, prevent and investigate criminal offences, unless those systems are available for the public to report a criminal offence"
- "a biometric categorisation system" (Art. 52.2, but referred to in EM4 on p.14 as "an AI system to determine association with (social) categories based on biometric data" – which is distinctly different from the Art. 52 definition, because of 'determine' cf. 'categorise' and 'social categorisation' cf. categorisation in the abstract) , except where "permitted by law to detect, prevent and investigate criminal offences"
- "an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ('deep fake')" except where it is:
  - "authorised by law to detect, prevent, investigate and prosecute criminal offences" (Art. 52.3, but expressed slightly differently in Art. 1(c)); or
  - "is necessary for the exercise of the right to freedom of expression and the right of freedom of the arts and sciences ... and subject to appropriate safeguards for the rights and freedoms of third parties" (Art. 52.3)

It is alarming that the EM Proposal appears to embody legal authorisation of 'deep fake' techniques to "detect, prevent, investigate" (presumably as a means of misleading witnesses and suspects), and, even worse, to "prosecute" (which would represent the creation of false evidence).

(4) All Other AI Systems (Art. 69, EM 5.2.2, Pre 81-82)

These are referred to variously as "AI systems other than high-risk AI systems" (Art. 69), "low or minimal risk [uses of AI]" (EM 5.2.5), and "non-high-risk AI systems" (EM 5.2.7, Pre 81).

The second (exploitation of vulnerabilities) embodies an even more impressive suite of escape-clauses for corporate lawyers to take advantage of: "[i] exploits any of the vulnerabilities [ii] of a specific group of persons [iii] due to their age, physical or mental disability, [iv] in order to [v] materially [vi] distort the behaviour of a person [vii, and impenetrably] pertaining to that group [viii] in a manner that causes or is likely to cause that person or another person physical or psychological harm" (Art. 5.1(b)).

The third (social scoring) contains 121 words, and even more, and even more complex, qualifying conditions that the second (Art. 5.1(c)).

The fourth, 'biometric identification' excludes (i) retrospective identification, (ii) proximate rather than remote identification, (iii) authentication (1:1 matching), (iv) non-public places, and (v) all uses other than for law enforcement (Art. 5.1(d), 5.2-5.4).

Further, many more biometric identification applications are expressly excluded from the scope, if they are "strictly necessary" for:

- "targeted search for specific potential victims of crime"
- "the prevention of a specific, substantial and imminent threat to the life or safety of natural persons ..."
- "the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence [with a custodial sentence of at least three years]"

In all three cases, the EC Proposal authorises mass biometric surveillance, and does so in order to find needles in the haystack. Such applications are nominally subject to prior authorisation, but that can be retrospective if the use is claimed to be urgent. So mass biometric surveillance can be conducted at will, subject to the possibility of an *ex post facto* day of reckoning.

Some of the vast array of exceptions that escape categorisation as (1) 'Prohibited AI Practices' may fall within (2) 'High-Risk AI Systems', but many others likely will not, and hence escape the regulatory scheme entirely.

In relation to **(2) High-Risk AI Systems**, many of the 8 areas and 21 specific categories are subject to multiple and wide-ranging exclusion criteria, whose effect is to greatly reduce the number of AI systems that will actually be within-scope.

One example of an "area" is "Biometric identification and categorisation of natural persons" (III-1). The category is limited by the following criteria: "[i] intended to be used for [ii] the 'real-time' and [iii] 'post' [iv] remote biometric [v] identification [vi] of natural persons" (III-1(a)). This excludes systems used without intent, either 'real-time' or 'post' but not both, proximate rather than 'remote', and for authentication (1-to-1) rather than identification (1-among-many).

Another "area" is "Access to and enjoyment of essential private services and public services and benefits:" (III-5), but with the categories limited to "AI systems [i] intended to be used [ii] by public authorities or on behalf of public authorities [iii] to evaluate the eligibility of natural persons for public assistance benefits and services, [iv] as well as to grant, reduce, revoke, or reclaim such benefits and services (III-5(a)). This excludes systems used without intent, 'private services and benefits' in their entirety (e.g. privately-run school-bus services, even if 'essential') – despite the nominal inclusion of "private services" in the "area" – and use for evaluation but without at least one of grant, reduce, revoke, or reclaim (due to the use of the conjunction 'as well as'). Further, AI systems to evaluate creditworthiness or establish a credit score exclude "small scale providers for their own use" (Art. 5(b)).

For such AI systems as do not fit into the array of escape-clauses, the statutory obligations involve (Art. 8):

- A risk management system (Art. 9)
- Data governance (Art. 10)

- Technical documentation (Art. 11)
- Record-keeping (Art. 12)
- Transparency and provision of information to users (Art. 13)
- Human oversight (Art. 14)
- Accuracy, robustness and cybersecurity (Art. 15)
- Operational requirements (Arts. 16-29)

Although the EC is required to "assess the need for amendment of the list in Annex III once a year following the entry into force of this Regulation" (Art. 84.1), the wording of Art. 84.7 may or may not create an obligation to actually propose amendments to the list, and the EC may or may not comply, and the European Parliament may or may not enact such amendments as are proposed.

The question is complicated by Art. 7, which empowers the EC to add further high-risk AI systems into Annex III, but does not require it to do so. In any case, the categories of AI systems are limited to the eight areas already specified in Annex III, and the criteria to be applied are complex, long, and provide ample reasons to avoid adding AI systems to the list.

In the case of category **(3) 'Certain AI Systems'**, the descriptions of the four sub-categories are lengthy, and are readily interpreted as being definitional, and some specific exceptions are declared. Hence many such AI systems are likely to be excluded from scope.

This category is orthogonal to 'levels' (2) and (4). Hence any particular AI system in category (3) may fall into (3) only, or (3) and (2), or (3) and (4).

In relation to category **(4) All Other AI Systems**, the EC Proposal imposes no regulatory requirements whatsoever on categories of AI or its application that fall ouside 'levels' (1)-(3). It merely requires 'encouragement and facilitation' of voluntary industry and organisational codes: "Providers of non-high-risk AI systems should be encouraged to create codes of conduct intended to foster the voluntary application of the mandatory requirements applicable to high-risk AI systems. Providers should also be encouraged to apply on a voluntary basis additional requirements" (Pre para. 81).

This is non-regulation. Further, the reference-point against which the appropriateness of such codes might be measured is indicated as "the mandatory requirements applicable to high-risk AI systems" . As the following analysis shows, this is a very limited set of principles. It is unclear why the reference-point is not instead the EC's own 'Ethics Guidelines for Trustworthy AI' (EC 2019). The cumulative effect of these factors is further discussed in s.4.1.

---

3. The Evaluation Process

This document reports on an assessment of the EC's Proposal against the consolidated set of 50 Principles, which were published in Clarke (2019) and are available in HTML, and in PDF.

The text of the EC Proposal was examined, with primary focus on the proposed legislation (pp.38-88), but with reference also to the Explanatory Memorandum (pp.1-16) and the Preamble (pp.17-38). Elements were noted whose coverage relates to those in the consolidated set of 50 Principles.

The information arising from the examination is as follows:

- The EC Proposal's contributions to the 50 Principles are extracted, and annotated. The details are provided in:
  - Appendix 1A – High-Risk AI Systems
  - Appendix 1B – 'Certain AI Systems'
  - Appendix 1C – All Other AI Systems
- In respect of High-Risk AI Systems, the consolidated set of 50 Principles is reproduced, with annotations inserted indicating text within the EC Proposal that corresponds with them. See:
  - Appendix 2 – High-Risk AI Systems

As a complement to those analyses, it is necessary to take into account the multi-dimensional scope of applicability of the provisions. The applicability is to some extent common across the four segments, and to some extent specific to each segment. In addition, a preliminary assessment was undertaken of the extent to which the EU's proposal, if enacted in its present form, would represent effective regulation. This varies considerably across the four segments.

Previous assessments of 32 different sets of 'principles for responsible AI' considered each set against each of the 50 consolidated Principles, and scored either 1 point or 0 point for each. During those assessments, the analysis "scored documents liberally, recognising them as delivering on a Principle if the idea was in some way evident, even if only some of the Principle was addressed, and irrespective of the strength of the prescription" (Clarke 2019, p.416).

All 32 of those sets were mere 'ethical guidelines', and were not intended to be enacted into law. The EC Proposal under consideration here, on the other hand, is expressly a draft statute, to be considered for enactment by the European Parliament. Whereas sets of unenforceable 'ethical' guidelines can be reasonably evaluated in isolation, statutory requirements are situated within a legal context. The EU's relatively strong human rights and data protection laws need to be factored in – although it is far beyond the scope of the study reported here to delve deeply into questions of their scope and effectiveness in the context of 'AI' 'practices' and 'systems'.

The 'liberal' scoring is less appropriate to the formal regulatory context. This assessment accordingly varied the scoring approach. Rather than a binary value (0 or 1), it assigns a score on an 11-point range {0.0, 0.1, 0.2, ..., 0.9, 1.0}. In order to maintain some degree of comparability with the previous 32 assessments, a second, binary score was assigned by rounding up, i.e. 0.0 rounds to 0, but all scores in the range 0.1 to 1.0 round to 1.

No genuinely 'objective' procedure is practicable. The score is assigned subjectively, based on a combination of the extent of coverage, the scope of exemptions, exceptions and qualifications, and the apparent practicability of enforcement. The scores were assigned by the author, and reviewed by the author again some weeks afterwards in order to detect instances of inappropriate scores, and adapted accordingly. Needless to say, each person who applies this approach to the data will come up with variations, which may result in both systematic and stochastic differences in assessments.

In order to provide a quantitative summation, the 32 previous assessments added the individual scores (with a maximum of 50), and expressed the result as a percentage. This of course implicitly weights each of the 50 Principles identically, and implicitly weights each of the 10 Themes according to the number of Principles within that Theme. The same approach is adopted here. The rounded / binary scores (0 or 1) result in a percentage that enables a reasonable degree of comparability of the EC Proposal's coverage against the 32 'ethical guidelines'.

Addition of the scores on the more granular, 11-point scale provides an estimation of the EC's Proposal's efficaciousness. As no other Proposal has as yet been assessed (or, indeed, seen), there is not yet any other instance that this measure can be compared against.

4. The Results

The main body of the assessment is in s.4.2, supported by Appendices 1A-1C and 2. This reports on the coverage of The 50 Principles that the EC Proposal achieves. Because the Proposal is for law to be enacted, it is necessary to precede that with an examination of the law's scope of applicability, and to follow that with some observations on the proposed law's effectiveness in relation to the primary objective of stimulating economic activity, and the secondary or qualifying purpose of establishing a regulatory instrument.

4.1 Scope of Applicability

The definition of what it is that the law is to apply to is the subject of several somewhat tortuous segments of the EC's Proposal. They are extracted in Appendix 3. For the proposed law to apply:

- four scope-conditions must all be satisfied, relating to the artefact, the entity, the geographical location, and the timeframe; AND
- one scope-condition must NOT be satisified, relating to the purpose.

The relevant category of artefact is defined as 'AI systems'.

The sub-term **'AI' is used in a pragmatic manner**, but aspects of it may be foreign to some people. It is defined to include several sub-sets of the field of AI as that somewhat vague term is conventionally understood. One of those is "logic-based approaches", and another is "knowledge-based approaches" including rule-based expert systems. The other sub-set has been in recent years referred to as AI/ML (machine learning). No other aspects of AI (such as pattern recognition, natural language understanding, robotics or cyborgisation) are included in the definition.

However, the term AI is defined to also include further categories of 'approaches and techniques' that are not conventionally referred to as AI: "Statistical approaches, Bayesian estimation, search and optimization methods". These pre-date the coinage of the term 'AI' in 1955, and have been more commonly associated with operations research and data mining / 'data analytics'.

The term **'AI system' is defined as "software** that is developed [using AI approaches and techniques, in the sense just discussed] ..." (Art. 3(1)). The term 'system' conventionally refers to a collection of interacting elements that are usefully treated together, as a whole. Moreover, those elements, while they include software, may include other categories as well, particularly hardware, and, at least in the socio-technical context, people and perhaps organisations as well. It is feasible to conceive the term, as used in the EC Proposal, as being intended to encompass all forms in which software presents, whether or not integrated into a system, and whether or not the system comprises elements other than software. On the other hand, there is ample scope for legalistic debate in and out of the court-room as to what is and is not an 'AI system', and hence for loopholes to be exploited.

A further limitation in Art. 3(1) is that the artefact must be able to "generate outputs", and it appears that these must reflect or fulfil "human-defined objectives". It may also prove to be somewhat problematic that the examples of "outputs" extend to "decisions influencing the environments they interact with", but no mention is made of actuators, nor to any other term that relates to autonomous action in the real world. It would seem both incongruous and counterproductive for **AI-driven robotics** to be **excluded** from the regulatory regime; yet, on the surface of it, that appears to be the case.

The field of view adopted in the EC proposal has a considerable degree of coherence to it, in that it declares its applicability to a number of categories of data analytics techniques and approaches, whose purpose is to draw inferences from data. Unfortunately, it uses the term 'AI systems' to refer to that set. This is misleading, and potentially seriously so. Misunderstandings are bound to arise among casual readers, the popular media and the

general public, but quite possibly also among many people in relevant professional and regulatory fields.

Moreover, the term is arguably so misleading that the use of the term **'AI Act'** could be seen as a materially excessive claim in relation to the scope of the Proposal, because a great deal of AI is out-of-scope. Added to that (and beneficially), a considerable amount of non-AI is within-scope. A more descriptive term would be **'Data Analytics Act'**.

Further scope-exemptions arise in the case of **'High-Risk AI Systems'**. See s.2 and Table 1. This is by virtue of the complex expressions in Arts. 2.1(a), 2.1(b), 2.2 – intersecting with 6 - 2.3 and 2.4. This possibly releases from any responsibility a great many instances of AI systems that are recognised as embodying high risk.

Two relevant categories of **entity** are defined. The notion of **'provider'** is defined in a sufficiently convoluted manner that scope may exist for some entities that provide relevant software or services into the EU to arrange their affairs so as to be out-of-scope.

The term **'user of an AI system'** is less problematic. A 'natural person' is included, but the exclusion clause for "used in the course of a personal non-professional activity" would appear to have the effect that a considerable amount of abusive behaviour by individuals is out-of-scope.

The **geographical location** scope-definition is convoluted, and hence the law might apply in circumstances in which **any** of the provider, the provider's action **or** the system is within the EU, and in which **any** of the user, the system **or** the use of the output is within the EU. However, that may remain unclear unless and until the courts have ruled on the meanings of the provisions.

The **timeframe** scope-definition is unclear. It presumably applies from some date a defined period after enactment, and is otherwise without constraints.

An express exclusion exists where the **purposes** are exclusively military.

The primary source used in the present study as a basis for evaluating the EC Proposal is The 50 Principles (Clarke 2019). These relate to AI as a whole, and do not go into detail in relation to Data Analytics. Because the EC Proposal defines 'AI Systems' not in a conventional manner but as something similar to '{advanced} data analytics', it is necessary to supplement The 50 Principles with an additional reference-point in the form of 'Guidelines for the Responsible Application of Data Analytics'. This was developed from the literature and appeared in the same journal as The 50 Principles (Clarke 2017).

4.2 Coverage of the Principles

This section summarises the results of the assessment of the extent to which the EC Proposal satisfies the reference-point, the consolidated 50 Principles. Because the EC Proposal divides AI Systems into 4 'levels', this analysis is of necessity divided into 4 sub-sections.

(1) Prohibited AI Practices

In the case of Prohibited AI practices, the Principles are by definition not relevant. However, a great many AI systems that appear to fall into a Prohibited category are not prohibited, because of the many exclusions and exemptions whose presence is drawn to attention in section 2 above.

All of those excluded and exempted AI systems fall into either category (2) or (4), and may fall into category (3) as well. Hence the extent to which the EC Proposal satisfies the 50 Principles in respect of those AI systems depends on which category/ies they fall into.

(2) High-Risk AI Systems

The majority of the relevant passages are in Chapter 2 (Requirements) Arts. 8-15 and Chapter 3 (Obligations) Arts. 16-29, but a number of other, specific Articles are also relevant.

The following is a summary of the lengthy assessment in Appendices Appendix 1A and Appendix 2, and the statistical analysis in Annex 1:

- as described in section 3, the first assessment adopts a liberal approach, scoring 1 point for each of the 50 Principles if the idea is in some way evident, even if only partially or weakly:
    - **the EU Proposal's coverage of the 50 Principles is remarkably sparse**, making a contribution of some kind to only 50% of them. Although this is better than all but 2 of the 32 'ethical guidelines' evaluated to date, it is still very low;
    - **the EU Proposal's coverage is markedly lower than that of the EC's own Ethical Guidelines (EC 2019)**, which score 74%. Its coverage is sufficiently different from that of the Guidelines to suggest that the Guidelines were not a material influence on the composition of the 2021 Proposal;
    - **among the 50 Principles' 10 Themes, the worst failings were foundational issues**, with 1 (Assess Positive and Negative Impacts and Implications) at 22%, 2 (Complement Humans) at 0%, 3 (Ensure Human Control) at 57%, and 4 (Ensure Human Safety and Wellbeing) at 33%. Also scoring badly are 7 (Embed Quality Assurance) at 33%, and 10. (Enforce, and Accept Enforcement of, Liabilities and Sanctions) at 50%;
- the second assessment assigns a subjective score for the extent of the coverage of each of the 50 Principles, on an 11-point scale: {0.0, 0.1, 0.2, ... 1.0}:
    - **even in respect of the 25 Principles to which it makes a contribution, the EU Proposal is highly inadequate**, totalling only 14.7 /50 (29%), with only 3/50 scoring 1.0, a further 14/50 scoring between 0.5 and 0.9, and the remaining 8/50 scoring below 0.5. This constitutes a Serious Fail;
    - **the foundational Themes 1-4 score an even more Serious Fail**, with 4%, 0%, 33% and 28%, for a total of 20%. The only 3 of the 10 Themes with a Pass-level score were 8 (Exhibit Robustness and Resilience – 68%), 5 (Ensure Consistency with Human Values and Human Rights – 60%), and 9 (Ensure Accountability for Obligations – 50%).

The disjunction between the EC Proposal (EC 2021) and the earlier 'Ethics Guidelines for Trustworthy AI' (EC 2019) is striking. Key expressions in the earlier document, such as 'Fairness', 'Prevention of Harm', 'Human Autonomy', 'Human agency', 'Explicability', 'Explanation', 'Well-Being' and 'Auditability', are nowhere to be seen in the body of the Proposal. The term 'stakeholder participation' occurs a single time (as a merely optional feature in organisations' creation process for voluntary codes). The term 'auditability' occurs, but not in a manner relating to providers or users, i.e. not as a regulatory concept.

The magnitude of the shortfall may be gauged by considering key Principles that are excluded from the protections, even for the (possibly quite low) proportion of high-risk AI systems that the EC Proposal has defined to be within-scope. Table 2 lists key instances.

Despite the acknowledgement by the EC Proposal that substantial harm may arise from these AI systems, those responsible for the development and deployment of such systems do not need to consult with affected parties, nor enable them to participate in the design, nor take any notice even where concerns are expressed, nor even ensure that negative impacts are proportionate to the need.

AI systems can be implemented with decision-making and action-taking capabilities, without consideration as whether they should instead be conceived as complementary technology, and as decision aids. Unless they fall into the narrow terms of 'certain

systems', as discussed in section 4.2(3) immediately below, high-risk AI systems are permitted to deceive humans. Organisations can impose AI systems on individuals as a condition of contract, and even as a condition of receiving a public benefit, irrespective of the risks the person thereby faces.

There is no obligation to contribute to human wellbeing. There is no obligation to avoid the manipulation of the vulnerable. There is not only no need to explain decision rationale to people, but also no need to even be capable of doing so. The quality assurance requirements are limited, and lack even a need to ensure that the techniques used are legitimate to apply to the particular kind of data that is available, or that the data is of sufficient quality to support decision-making. No safeguards are needed, no controls to ensure the safeguards are working, and no audits to ensure the controls are being applied. Yet this purports to be effective regulation of high-risk applications of powerful technologies, many of which are new, experimental and poorly-understood.

Table 2: Missing Principles for High-Risk AI Systems

Excerpts from The 50 Principles (Clarke 2019)

1. Assess Positive and Negative Impacts and Implications
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')
2. Complement Humans
2.1 Design as an aid, for augmentation, collaboration and inter-operability
2.2 Avoid design for replacement of people by independent artefacts or systems, except where those artefacts or systems are demonstrably more capable than people, and even then ensuring that the result is complementary to human capabilities
3. Ensure Human Control
3.6 Avoid deception of humans
3.7 Avoid services being conditional on acceptance of AI-based artefacts and systems
4. Ensure Human Safety and Wellbeing
4.3 Contribute to people's wellbeing ('beneficence')
4.6 Avoid the manipulation of vulnerable people , e.g. by taking advantage of individuals' tendencies to addictions such as gambling, and to letting pleasure overrule rationality
6. Deliver Transparency and Auditability
6.3 Ensure that people are aware of inferences, decisions and actions that affect them, and have access to humanly-understandable explanations of how they came about
7. Embed Quality Assurance
7.1 Ensure effective, efficient and adaptive performance of intended functions
7.3 Justify the use of data, commensurate with each data-item's sensitivity
7.6 Ensure inferences are not drawn from data using invalid or unvalidated techniques
7.8 Impose controls in order to ensure that the safeguards are in place and effective
7.9 Conduct audits of safeguards and controls
8. Exhibit Robustness and Resilience
8.3 Conduct audits of the justification, the proportionality, the transparency, and the harm avoidance, prevention and mitigation measures and controls

AI systems of course do not exist in a legal vacuum, and organisational behaviour that exploits the enormous laxness of the EC's Proposal will in due course be found to be in breach of existing laws. That, however, will require litigation, considerable resources, and considerable time, and, during that time, potentially very substantial harm can be inflicted on people, without any ability for them to understand, avoid or obtain redress for, unjustified decisions, actions and harm.

(3) 'Certain AI Systems'

A single Article, of about 250 words, imposes a very limited transparency requirement on four highly-specific categories of AI systems (Art. 52). It applies only to AI systems that interact with humans, 'emotion recognition systems', ' biometric categorisation systems' and 'deep fake systems', but in each case with substantial qualifications and exceptions outlined in Table 1.

The analysis in Appendix 1B shows that the Article merely requires that the people subjected to such AI systems be informed that it is an AI system of that kind. This is a very limited contribution to Principle 6.1, and adds an infinitesmal score to the totals summarised in the previous section. For a great many AI systems, on the other hand, this would be the sole requirement of the regulatory regime.

(4) All Other AI Systems

The very last Article in the EC Proposal, and hence probably an afterthought, is a requirement of the EC and EU member states to "encourage and facilitate the drawing up of codes of conduct intended to foster the voluntary application to AI systems other than high-risk AI systems of the requirements set out in Title III, Chapter 2" (Art. 69.1). The analysis in Appendix 1C suggests this is close to valueless.

In section 4.2(2) above, it was noted that even within-scope high-risk AI systems are free from any requirements in respect of a wide range of what various sets of 'Ethical Guidelines for Responsible AI' declare as being Principles. The myriad AI systems that fall outside the scope of the EC Proposal suffer not only from all of those inadequacies, but also from the absence of the quite basic protections that are applicable to high-risk systems, as indicated in Table 3.

In respect of most AI systems, user organisations are not required to conduct an assessment of the impacts on and implications for affected parties. They are subject to no requirement to ensure human control, even over autonomous behaviour. They have no obligations to avoid, prevent or mitigate harm. AI systems are permitted to be used in secret, yet there are no requirements even that the decision rationale made can be reconstructed, let alone explained to those who suffer harm as a result. Quality assurance measures are optional, and no testing in real-world contexts is needed. There is no requirement for robustness even against passive threats let alone active attacks. No accountability mechanisms are imposed.

If AI were to deliver even a small proportion of the magic that is promised, existing laws will be grossly inadequate to cope with its ravages. The law is slow, expensive, unpredictable and unadaptive. In any case, many of these Principles are not currently established in law, because previous technologies were less powerful, and their use was intermediated by human beings rather than the functions being performed autonomously by artefacts. It will require long debate in abstruse terms before series of courts, contested expert evidence, and many false starts and appeals before some semblance of order, and protections against unreasonable decisions, can be restored.

Table 3: Additional Missing Principles for All Other AI Systems

Excerpts from The 50 Principles (Clarke 2019)

1. Assess Positive and Negative Impacts and Implications
1.4 Conduct impact assessment, including risk assessment from all stakeholders' perspectives
3. Ensure Human Control
3.1 Ensure human control over AI-based technology, artefacts and systems
3.2 In particular, ensure human control over autonomous behaviour of AI-based technology, artefacts and systems
3.5 Ensure human review of inferences and decisions prior to action being taken
4. Ensure Human Safety and Wellbeing
4.1 Ensure people's physical health and safety ('nonmaleficence')
4.4 Implement safeguards to avoid, prevent and mitigate negative impacts and implications
6. Deliver Transparency and Auditability
6.1 Ensure that the fact that a process is AI-based is transparent to all stakeholders
6.2 Ensure that data provenance, and the means whereby inferences are drawn from it, decisions are made, and actions are taken, are logged and can be reconstructed
7. Embed Quality Assurance
7.2 Ensure data quality and data relevance
7.7 Test result validity, and address the problems that are detected
8. Exhibit Robustness and Resilience
8.1 Deliver and sustain appropriate security safeguards against the risk of compromise of intended functions arising from both passive threats and active attacks, commensurate with the significance of the benefits and the potential to cause harm
8.2 Deliver and sustain appropriate security safeguards against the risk of inappropriate data access, modification and deletion, arising from both passive threats and active attacks, commensurate with the data's sensitivity
8.4 Ensure resilience, in the sense of prompt and effective recovery from incidents
9. Ensure Accountability for Obligations
9.1 Ensure that the responsible entity is apparent or can be readily discovered by any party
9.2 Ensure that effective remedies exist, in the form of complaints processes, appeals processes, and redress where harmful errors have occurred

4.3 Effectiveness as a Regulatory Instrument

The analysis immediately above suggests that the EC's Proposal:

- creates no protections whatsoever in respect of potentially very large numbers of 'AI systems' that fall outside the document's narrow scope-definitions of 'prohibited', 'high-risk', and 'certain' AI systems, or that fit within express and implied exemptions, despite the considerable impact the unregulated systems are likely to have

- prohibits very few categories of AI system
- creates highly inadequate, vague and in many cases probably unenforceable protections in respect of a limited range of what it calls 'high-risk' systems, and provides no means whatsoever for people negatively affected by AI systems to pursue corrections and redress, limiting action to EU member-state's regulatory agency
- imposes a very weak transparency requirement on a small set of categories of system that are intended to interact directly with people

The EC's Proposal is not a serious attempt to protect the public. It is very strongly driven by economic considerations and administrative convenience for business and government, with the primary purposes being the stimulation of the use of AI systems. The public is to be lulled into accepting AI systems under the pretext that protections exist. The social needs of the affected individuals have been regarded as a constraint not an objective. The many particularities in the wording attest to close attention being paid to the pleadings of advocates for the interests of government agencies, corporations and providers of AI-based services. The draft statute seeks the public's trust, but fails to deliver trustworthiness.

4.4 Effectiveness as a Stimulatory Instrument

The analysis has been conducted with a primary focus on the EC Proposal's regulatory impact, but mindful of its "twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of such technology" (EM 1.1, p.1). The very substantial failure of the EC Proposal from a regulatory perspective makes clear that its role as a stimulant for technology and economic activity has dominated the design. The strong desire to achieve the public's trust is apparent throughout the EM and Preamble; but the shallowness of the offering is apparent throughout the analysis presented here.

There are many indicators of the importance accorded to innovation, in the form of the entire omission of even quite basic regulatory measures, in the vast array of exemptions from scope even in the case of so-called 'high-risk AI systems', in the completeness of the exemption of AI systems that are defined to be outside scope, in the manifold exceptions that apply even to those systems that are within-scope, and in the large numbers of loopholes provided, whether by error or intent.

The Proposal even contains an explicit authorisation in support of providers and user organisations, which compromises existing legal protections: "Personal data lawfully collected for other purposes shall be processed for the purposes of developing and testing ... innovative AI systems ... for safeguarding substantial public interest [in relation to crime, public safety and public health and environmental quality] ... where ... requirements cannot be effectively fulfilled by processing anonymised, synthetic or other non-personal data" (Art. 54.1).

This appears to authorise what would otherwise be expropriation of personal data in breach of data protection laws. It is in a draft law whose purpose is ostensibly to "facilitate the development of a single market for lawful, safe and trustworthy AI applications" (EM1.1, p.3), based on "Article 114 of the Treaty on the Functioning of the European Union (TFEU)" (Pre(2), p.17), which allows the EU to regulate those elements of private law which create obstacles to trade in the internal market".

5. Interpretation

The EC's announcement in April 2021 (EC 2021) proclaimed that "The new AI regulation will make sure that Europeans can trust what AI has to offer". The analysis reported here shows that to have been a wild exaggeration. Other parts of the Press Release were much closer to the truth, however, such as the declared intention of "strengthening AI uptake, investment and innovation across the EU".

Public policy will be best served by communicating the complete inadequacy of the EC's Proposal, to the public, the media, and importantly the European Parliament. In the event that 'AI systems' begin to deliver on their promises, the existence of an enactment such as this would do irreparable harm to the people who are subjected to the enthusiasm and carelessness of largely uncontrolled providers and user organisations.

Appendix 1A: High-Risk AI Systems, and The 50 Principles

This section contains extracts from (EC 2021), interpretive comment, and [cross-references to the corresponding elements of The 50 Principles enclosed within square-brackets]

These provisions only apply to those High-Risk AI Systems that do not enjoy an exemption, including by virtue of any of Arts. 2.1(a), 2.1(b), 2.2, 2.3 and 2.4, but with the meaning of 2.2 intersecting with provisions in Art. 6 in ways that appear to be difficult to determine

Note also that the, possible many, exemptions of AI systems are exemptions-of-the-whole, that is to say that none of the provisions apply, and hence exempted high-risk AI systems are only subject to the non-regulation of level (4) All Other, plus the possibility of the very limited transparency requirement for 'level' (3) 'Certain AI Systems'

Risk Management System (Art. 9)

"A risk management system shall be established, implemented, documented and maintained ... " (Art. 9.1). A lengthy recipe is expressed, broadly reflecting contemporary risk assessment and risk management practices. The responsibility is implicitly addressed to providers and appears not to be addressed to users. It does not expressly require any testing in real-world contexts.

In addition, the Art. 9 provisions are subject to Art. 8.2, which heavily qualifies the provisions by making compliance with them relative to "the intended purpose". Not only is this sufficiently vague as to represent a very substantial loophole, it entirely ignores any other possible uses, including already-known actual uses, and their impacts and implications.

Fundamentally, however, the provision is misdirected. The notions of risk assessment and risk management are understood by organisations to adopt the perspective of the organisation itself, i.e. this is about 'risk-to-the-organisation assessment and management'. The notions of Multi-Stakeholder Risk Assessment and Management are not unknown. See, for example, s.3.3 of (Clarke 2019); but they remain at this stage a long way from mainstream understanding and use.

The term 'impact assessment' is commonly used when organisations are required to consider the dis/benefits and risks faced by users and usees of IT systems generally. The term 'impact assessment' is used 13 times in the EC Proposal, but it does not relate not to responsibilities of providers and users. All but one relates to the process conducted by the EC in developing the proposal, and the remaining one relates to Data Protection Impact Assessments under the GDPR. Moreover, the mentions of the term 'impact' in relation to affected individuals and fundamental rights are all in discursive prose or obligations on the EC itself. The term does not occur in passages that relate to risk assessment and management by providers or user organisations.

A few passages could be read as requiring examination of broader interests. The statement is made that "Risks ... should be calculated taking into account the impact on rights and safety" (EM3.1, p.8); but this is only in a report on the EC's own stakeholder consultations. A single relevant instance of the expression "risk to" appears in the EC's Proposal: "Title III contains specific rules for AI systems that create a high risk to the health and safety or fundamental rights of natural persons" (EM 5.2.3, p.13). An exception relates to impacts on children: "When implementing the risk management system ..., specific consideration shall be given to whether the high-risk AI system is likely to be accessed by or have an impact on children" (Art. 9.8).

Nothing else in Art. 9 suggests to the reader that the work is to reflect the interests of multiple stakeholders, and in particular the interests of the affected parties. It is very

difficult to believe that one sentence in the Explanatory Memorandum and a single children-specific clause would result in providers dramatically shifting their understanding of the scope of risk assessment and management. It would require a very broad reading of the EC Proposals by the courts for 'risk assessment and risk management' to be interpreted as requiring providers and user organisations to conduct 'impact assessments' on individuals affected by use of AI systems.

In short, if the objective was to protect the public, **this is a highly ineffectual provision**.

[Art. 9 generally makes some contribution to P1.4 ("Conduct impact assessment, including risk assessment from all stakeholders' perspectives"), but only a small contribution because risk assessment is conducted from the perspective of the provider, perhaps with the interests of user organisations also in mind. An exception exists in relation to the reflecting the interests of children, by virtue of Art. 9.8.]

[Art. 9.4 corresponds to P4.4 ("Implement safeguards to avoid, prevent and mitigate negative impacts and implications").]

Data Governance and Management Practices (Art. 10)

"Appropriate data governance and management practices shall apply for the development of high-risk AI systems [generally]...[concerning in particular] (Art. 10.6, 10.2):

(a) "the relevant design choices;

(b) data collection;

(c) relevant data preparation processing operations, such as annotation, labelling, cleaning, enrichment and aggregation;

(d) the formulation of relevant assumptions, notably with respect to the information that the data are supposed to measure and represent;

(e) a prior assessment of the availability, quantity and suitability of the data sets that are needed;

(f) examination in view of possible biases;

(g) the identification of any possible data gaps or shortcomings, and how those gaps and shortcomings can be addressed".

[Art. 10.2 articulates the first part of P7.2 ("Ensure data quality ..."), but does not address the second part ("Ensure .. data relevance"). It makes no contributions to P7.4 (re data security safeguards), nor to P7.3 ('justification for the use of sensitive data').]

[Art. 10.2(f) fails to fulfil even the data-related aspects of P5.1 ("Be ... impartial ... avoid unfair discrimination and bias ...", because it requires only examination, and fails to actually require it be avoided.]

[When compared with the Guidelines for the Responsible Application of Data Analytics (Clarke 2017), the articulation in Art. 10.2 sub-sections is vague, and provides coverage of few of the 9 data-related Guidelines 2.2-2.8, 3.3, 3.5:

- Art. 10.2(c) makes some contribution to G2.3 (re data merger) and G2.4 (re data scrubbing)
- Art. 10.2(d) has close affinity to one of three threshold tests of data in G3.5, relating to "reliable correspondence with the real-world systems about which inferences are to be drawn"
- Art. 10.2(e), specifically re "suitability of the data sets", makes some contribution to G2.2 (stipulations about what needs to be understood about each source of data)]

Some additional requirements are imposed on a sub-set of "high-risk AI systems", those that "make use of techniques involving the training of models with data" (Art. 10.1).

"Training, validation and testing data sets" are required to "be relevant, representative, free of errors and complete" and to "have the appropriate statistical properties" (Art. 10.3).

[This further articulates the first part of P7.2 ("Ensure data quality ..."), and an element of G3.5 ("satisfy threshold tests ... in relation to data quality").]

"Training, validation and testing data sets shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, behavioural or functional setting within which the high-risk AI system is intended to be used" (Art. 10.4).

[This addresses G2.1 ("Understand the real-world systems about which inferences are to be drawn and to which data analytics are to be applied").]

Art 10.5 contains a very complex 99-word sentence whose core expression appears to be "may process special categories of personal data ... subject to appropriate safeguards ... where anonymisation may significantly affect the purpose pursued". The effect appears to be to authorise, in some contexts, for some purposes, highly sensitive personal data to be used without anonymisation, and merely subject to some other, less effective protections ("security and privacy-preserving measures, such as pseudonymisation, or encryption"). This appears to be a designed-in loophole to override protections in other laws – and potentially a quite substantial loophole.

[This appears to authorise what would otherwise be breaches of existing safeguards implementing P7.3 ('justification for the use of sensitive data') and G2.5 ("ensure ... de-identification (if the purpose is other than to draw inferences about individual entities)"). Because it provides an exemption, it also appears to be a negation of P1.8 ("Justify negative impacts on individuals ('proportionality')").]

Technical Documentation (Art. 11)

"[T]echnical documentation ... shall be drawn up before that system is placed on the market or put into service and shall be kept up-to date [and] drawn up in such a way [as] to demonstrate that the high-risk AI system complies with the requirements set out in [Arts. 8-15]" (Art. 11.1)

[This is merely an enabler of parts of P7 ("Embed quality assurance"). It addresses a small part of P10.2 ("Comply with [regulatory processes] ...").]

Record-Keeping (Art. 12)

"High-risk AI systems shall be designed and developed with capabilities enabling the automatic recording of events ('logs') ... [conformant with] recognised standards or common specifications ... [ensuring an appropriate] level of traceability ... " (Arts. 12.1-3). Cross-references to Arts. 61 and 65 appear to add nothing to the scope or effectiveness of the provision.

For remote biometric identification AI systems, some specific capabilities are listed, generally consistent with logging conventions (Art. 12.4).

[This implements only a small element within P6.2, which states " Ensure that data provenance, and the means whereby inferences are drawn, decisions are made, and actions are taken, are logged and can be reconstructed". Instead of "means", the EC Proposal requires only "events", and it omits the vital criterion: that the inferencing and decision-making processes can be reconstructed.

[The provision appears not to make any contribution to G4.6, which requires that "mechanisms exist whereby stakeholders can access information about, and if appropriate complain about and dispute interpretations, inferences, decisions and actions", because it is expressed in the passive voice, and the only apparent rights of access are by the provider themselves and, under limited circumstances explained in Art. 23 and lengthily in Art. 65, the relevant national competent authority and/or national market surveillance authority.]

Transparency to Users (Art. 13)

"High-risk AI systems shall be designed and developed in such a way [as] to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately" (Art. 13.1).

"High-risk AI systems shall be accompanied by instructions for use ...", including "its intended purpose" and its risk profile (Arts. 13.2-3). Nothing in the provision appears to require the user to apply the instructions for use, or constrain the user to only use it for the system's "intended purpose".

[Art. 13.1 makes a limited contribution to P6.2 ("Ensure that ... the means whereby inferences are drawn ... can be reconstructed "). The contribution is very limited in that transparency is limited to the user organisation and no other stakeholders, and "sufficiently transparent" is far less than a requirement that "the means whereby inferences are drawn [and] decisions made" be communicated even to the user organisation.

[Similarly, and very importantly, it fails to require the provider to enable the user organisation to comply with G4.9: "the rationale for each inference is [to be] readily available to [those affected] in humanly-understandable terms".]

[Arts. 13.2-3 merely facilitate communication of information along the supply-chain. They enable the possibility of the user organisation applying the tool in a manner that manages the risks to the people affected by its use, but in themselves the provisions do nothing to ensure responsible use of AI systems in the terms of The 50 Principles.]

[The provision of information makes a contribution to the capability of the user organisation to fulfil G3.2: "Understand the ... nature and limitations of data analytic tools that are considered for use".]

Human Oversight (Art. 14)

"High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons" (Art. 14.1). The term "effectively overseen" may (but may not) encompass detection of "anomalies, dysfunctions and unexpected performance" (14.4(a)), awareness of the possibility of "automation bias" (14.4(b)), an unclear expression 'correct interpretation of the output' (14.4(c)), and an ability "to decide ... not to use the high-risk AI system or [to] otherwise disregard, override or reverse the output" (14.4(d)) – but without any obligation to ever take advantage of that ability.

[These provisions are enablers for P3.1 ("Ensure human control over AI-based artefacts and systems"), but whereas the inclusion of such features is a (qualified) obligation of providers, this provision alone does nothing to ensure the features are effectively designed, effectively communicated to user organisations and thence to users, and applied, appropriately or even at all.]

The possible need for an "ability to intervene ... or interrupt through a 'stop' button or a similar procedure" (Art. 14.4) – again without any obligation to ever apply it – appears to contemplate automated decision-making and action. This is in contrast to the apparent exclusion of robotic action from scope, noted in section 4.1 as arising from the failure of the Art. 3(1) definition of AI system to go beyond "[data] outputs" to encompass actions in the real world.

Art. 14.5 applies solely to "AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons". It appears to state that measures required by Art. 14.3 in those cases are to "ensure that ... no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons". It is unclear what the expression "the identification resulting from the system" is intended to mean, let alone what it will be taken to mean by the many people who are expected to read and apply the provisions and/or other people's interpretations of the provisions.

[These provisions may make small contributions to P3.1 ("Ensure human control over AI-based artefacts and systems") and P3.2 ("ensure human control over autonomous behaviour"). One reason that the contributions are so limited is that they only relate to the existence of means of human oversight, and do not actually require user organisations to apply those means. Another is that the poor standard of the drafting results in an absence of clarity about what, if anything, this Article requires be done in particular circumstances.

[Art. 14.5, which is applicable to only 1 of the 20 forms of 'high-risk AI system', fails to impose any responsibility on user organisations relating to human review. So these provisions do not satisfy P3.5 ("Ensure human review of inferences and decisions prior to acting on them"). (Art. 29 relating to user responsibilities is addressed below).]

Accuracy, Robustness and Cybersecurity (Art. 15)

"High-risk AI systems shall be designed and developed in such a way that they achieve, in the light of their intended purpose, an appropriate level of accuracy, robustness and cybersecurity ..." (Art. 15.1) and are to be "resilient as regards errors, faults or inconsistencies" (Art. 15.3).

[These provisions appear to be based on inadequate understanding of the terms used. Accuracy is an attribute of data rather than of inferences or decisions, for which the criteria are validity and reasonableness. The term 'robustness' refers to the ability to function effectively despite environmental perturbance, and 'resilience' refers to the capacity to recover, and/or the rapid achievement of recovery, from loss of service as a result of environmental perturbance. The vague term 'cybersecurity' may refer to assurance of service, or to any, some or all of assurance of sustained quality of service or of data, or to assurance of access to data only by authorised entities for authorised purposes.

[These Articles could be interpreted to mean a very wide range of things, but in practice can be largely ignored by providers because they are not sufficiently coherent to contribute to any of P6 requirements ("Embed quality assurance"), nor even the P8 requirements ("Exhibit robustness and resilience").]

"High-risk AI systems that continue to learn ... shall be developed in such a way [as] to ensure that possibly biased outputs due to outputs used as an input for future operations ('feedback loops') are duly addressed with appropriate mitigation measures" (Art. 15.3).

[The sentiment expressed is positive, but the expression is so clumsy and unclear, and the scope so limited, that ample excuse exists for ignoring the requirement on the basis that it is unclear how it could be operationalised. It accordingly contributes only a little to P4.4 ("Implement safeguards to avoid, prevent and mitigate negative impacts and implications"), P7.1 ("Ensure effective, efficient and adaptive performance of intended functions"), and P7.6 ("Ensure that inferences are not drawn from data using invalid or unvalidated techniques"). It similarly offers little in relation to G3.2 and G3.4.]

Procedural Provisions Imposed on Providers (Arts. 16-28)

A range of procedural obligations are imposed on providers, including a small number that appear to be additional to the limited requirements noted above.

Providers are required to have a quality management system in place, and it is to include "examination, test and validation procedures to be carried out before, during and after the development of the high-risk AI system" (Art. 17.1(d)).

However, the quality of the quality management system is qualified, in that it "shall be proportionate to the size of the provider's organisation" (Art. 17.2). This creates a substantial loophole whereby arrangements can be made for risk-prone AI systems to be provided by small organisations with less substantial obligations than larger providers.

[This contributes to P7 ("Embed quality assurance"), but generally without articulation. It does address P7.7 ("Test result validity"), but without expressly requiring that problems

that are detected are addressed, and it is subject to a vague qualification in relation to the operator's size.]

"Providers of high-risk AI systems which consider or have reason to consider that a high-risk AI system which they have placed on the market or put into service is not in conformity with this Regulation shall immediately take the necessary corrective actions to bring that system into conformity, to withdraw it or to recall it, as appropriate" (Art. 21).

However, the provision does not make clear who determines whether an AI system is non-compliant, nor how and to what extent the requirement is enforceable.

[The provision makes no material contribution to P9.1 ("Ensure that the responsible entity is apparent or can be readily discovered by any party"), because none of the parties affected by the AI system, and no regulatory agencies, appear to be involved. Nor does it contribute in any way to P9.2 ("Ensure that effective remedies exist, in the form of complaints processes, appeals processes, and redress where harmful errors have occurred").]

Moreover, the provider is absolved of all responsibilities, despite the fact that the system may continue to be used for the original "intended Purpose" as well as the 'modified intended purpose' (Art. 28.2).

[This appears to seriously compromise such protections as were afforded by the provisions noted above.]

A Provider's Responsibilities May Shift to Users (Art. 28)

"[A] user ... shall be subject to the obligations of the provider under Article 16 [if] they modify the intended purpose of a high-risk AI system ... [or] make a substantial modification to the high-risk AI system" (Art. 28.1).

The effect of Art. 16 appears to be to apply many of the above obligations, arising from Arts. 8-26. However, the drafting quality is deplorable, and years of litigation would be necessary to achieve reasonable clarity about what Art. 16 does and does not require of user organisations.

[To the extent that the user organisation has a less distant relationship with the individuals affected by the use of the AI system (although not necessarily a close one), this may improve the contribution of some of the provisions to some of The 50 Principles and some of the Guidelines. However, the scope for user organisations to dodge responsibilities is enormous, even greater than that afforded to providers.]

User Responsibilities (Art. 29)

"Users of high-risk AI systems shall use such systems in accordance with the instructions of [sic: should be 'for'?] use accompanying the systems, pursuant to paragraphs 2 and 5" (Art. 29.1).

[Multiple weaknesses in the Art. 13 provisions in relation to 'instructions for use' were noted earlier. These weaknesses are compounded by the failure to impose any obligations on the user organisation in relation to transparency to those affected by use of the AI system. This provision accordingly appears to add little to the paltry protections contained in the earlier Articles.]

"[T]o the extent the user exercises control over the input data, that user shall ensure that input data is relevant in view of the intended purpose of the high-risk AI system" (Art. 29.3).

[Superficially, this seems to make a contribution to the second part of P7.2 ("Ensure ... data relevance"). However, it creates a massive loophole by permitting a user organisation to avoid responsibility for only using data that is relevant, on the illogical basis that they don't 'exercise control' over that data. The text is vague, and hence fulfilment of the condition is likely to be very easy to contrive. So the EC Proposal permits irrelevant data to cause harm to affected individuals, without any entity being liable for that harm.]

User organisations are to monitor and report on risks as defined in Art. 65(1) (Art. 29.4); but the scope of the designated EU Directive, and hence of Art. 65(1), appears to be limited to "harm" to "health and safety", neither of which appears to be defined in that Directive.

User organisations are also to report on "any serious incident or any malfunctioning ... which constitutes a breach of obligations ... ", as per Art. 62 (Art. 29.4). The scope of this reporting obligation is anything but clear, as is its enforceability.

[It is not clear that these provisions make a material contribution to P9.1 ("Ensure that the responsible entity is apparent or can be readily discovered by any party"), because none of the parties affected by the AI system are involved. Nor does it contribute in any way to P9.2 ("Ensure that effective remedies exist, in the form of complaints processes, appeals processes, and redress where harmful errors have occurred").]

"Users of high-risk AI systems shall keep the logs automatically generated by that high-risk AI system, to the extent such logs are under their control" (Art. 29.5).

[The corresponding obligation imposed on providers was noted as being inadequate. The imposition on user organisations makes even less contribution to P6.2 ("Ensure that the means whereby inferences are drawn, decisions made and actions are taken are logged and can be reconstructed"). That is because it enables the user organisation to absolve themselves of any responsibility to have control over logs.]

Machinery Provisions (Arts. 30-51)

Arts 30-51 specify eurocratic processes for the processing of notifications, standards, conformity assessment, certificates and registration. They are machinery provisions, and do not appear to not contain any substantive contributions to protections.

Governance and Enforcement (Arts. 56-68)

Arts 56-68 and 71-72 specify governance arrangements.

"Providers shall establish and document a post-market monitoring system in a manner that is proportionate to the nature of the artificial intelligence technologies and the risks of the high-risk AI system" (Art. 61). Providers are to provide access to market surveillance authorities under specified conditions (Art. 64). The market surveillance authority can require a provider to "withdraw the AI system from the market or to recall it ..." (Arts. 67-68). Obligations under Arts. 62 and 65.1 were referred to above (because they were referenced by Art. 29.4).

Art. 71.1 requires that there be "[effective, proportionate, and dissuasive] penalties ... [and] all measures necessary to ensure that they are properly and effectively implemented".

[The creation of the possibility that some relevant parties may become aware of who the responsible entity is represents only a small contribution to P9.1 ("Ensure that the responsible entity is apparent or can be readily discovered by any party").]

[Arts. 63--68 and 71-72 make a more substantial contribution to P9.2 ("Ensure that effective remedies exist ..."), but affected individuals and their advocates are excluded from the scheme.]

[The provisions within the EC Proposal do not appear to satisfy P10.1 ("Ensure that complaints, appeals and redress processes operate effectively").]

[The contribution to P10.2 ("Comply with processes") is modest. The market surveillance authority has powers, but the regime appears to create no scope for enforcement of any aspects of the EC Proposals by individuals or by public interest advocacy organisations.

A generic issue that arises is that the provisions are generally procedural in nature, not outcome-based. They read more like a generalised International Standards Organisation document written by industry, for industry, than a regulatory instrument.

Appendix 1B: 'Certain AI Systems', and The 50 Principles

This section contains extracts from (EC 2021), interpretive comment, and [cross-references to the corresponding elements of The 50 Principles inside square-brackets]

Note that this applies only to those 'Certain AI Systems' that do not enjoy an exemption, including by virtue of any of Arts. 2.1(a), 2.1(b), 2.3 and 2.4

A very limited transparency requirement applies to 'certain systems'. The requirement applies irrespective of whether the AI system in question is or is not also a 'high-risk AI system' as defined in Art. 6 and Annex III (Art. 52.4. See Table1).

Category 1, 'interaction systems', are to be "designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use" (Art. 52.1).

However, "This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence" (Art. 52.1).

[This makes a small contribution to P6.1 ("Ensure that the fact that a process is AI-based is transparent to all stakeholders"). The implementation is very limited because it only applies to the relatively small proportion of AI systems that interact directly with humans, and does not apply to those that impact humans without any direct interaction.

[It fails to make any contribution to P6.3 ("Ensure that people are aware of inferences, decisions and actions that affect them, and have access to humanly-understandable explanations of how they came about"), because the rights accruing to affected people are no more than knowing that is the case, i.e. **the EC Proposal authorises the suppression from those affected by it of what the AIS system does, how it does it, and what its impacts are**.]

[In addition, the absence of any other provisions has the effect of authorising AI systems that breach all of the many other 48 Principles that are not adequately protected under existing heads of law.]

Categories 2, 'emotion recognition systems' and 3, ' biometric categorisation systems', are subject to the requirement "shall inform of the operation of the system the natural persons exposed thereto" (Art. 52.2).

Category 4, 'deep fake systems', are subject to the requirement "shall disclose that the content has been artificially generated or manipulated" (Art. 52.3).

[These implement P6.1 ("Ensure that the fact that a process is AI-based is transparent to all stakeholders"), and make contributions to P6.3 whose effectiveness in unclear because of the vague expression "inform of the operation of the system" and the limitation to the fact of artificiality, not the nature or extent of artificiality, nor the means whereby the artificiality was achieved.]

[It is remarkable that, in all four categories of AI system, only a single 1 of 50 Principles is addressed, and the many other Principles affected by them have been ignored, including:

- P3.3 ("Respect each person's autonomy, freedom of choice and self-determination")
- P3.6 ("Avoid deception of humans")
- P4.2 ("Ensure people's psychological safety, by avoiding negative effects on any individual's mental health, inclusion in society, worth, standing in comparison with other people, or emotional state")
- P4.5 ("Avoid violation of trust")
- P4.6 ("Avoid the manipulation of vulnerable people ...")
- P6.5 ("Ensure data quality and data relevance") and
- P6.6 ("Deal fairly with people (faithfulness, fidelity)".]

Appendix 1C: All Other AI Systems, and The 50 Principles

This section provides extracts from (EC 2021), interpretive comment, and [cross-references to the corresponding elements of The 50 Principles inside square-brackets]

Note that this even these empty provisions may not apply to those Other AI Systems that enjoy an exemption, including by virtue of any of Arts. 2.1(a), 2.1(b), 2.3 and 2.4

The very last Article in the EC Proposal, and hence probably an afterthought, is a requirement of the EC and EU member states to "encourage and facilitate the drawing up of codes of conduct intended to foster the voluntary application to AI systems other than high-risk AI systems of the requirements set out in Title III, Chapter 2" (Art. 69.1).

Very limited public benefit is likely to accrue from the anaemic set of requirements even in respect of that sub-set of 'high-risk AI systems' as are subject to the provisions, or to some of them. It appears unlikely that any value at all would arise from the publication of 'codes' by corporations and industry associations that are voluntary and unenforceable.

It is particularly noteworthy that the EC did not even see fit to draw attention to its own 'Ethics Guidelines for Trustworthy AI' (EC 2019). These were ostensibly developed and published with the declared purpose of being "to promote Trustworthy AI" (p.2) and "articulate a framework for achieving Trustworthy AI based on fundamental rights" (p.6); yet in a mere 2 years they appear to have been relegated to history. The reasonable supposition is that corporate lobbying successfully achieved the massive dilution apparent in the EC Proposal, and this was carried over even as a basis for voluntary, unenforceable and inherently ineffective self-regulation.

This Article is very difficult to interpret as anything other than the cheapest possible form of window-dressing.

Appendix 2: The 50 Principles and High-Risk AI Systems

Annotated extract of The 50 Principles (Clarke 2019)
See here for a PDF version of the extract

Principles that are evident in the EC Proposal (even if only partly covered, or weakly expressed) are **in bold-face type**. The relevant segments of text from the EC Proposal are shown *in italics*, followed by cross-references to the locations in which that text occurs.

The majority of the relevant passages are in Chapter 2 (Requirements) Arts. 8-15 and Chapter 3 (Obligations) Arts. 16-29, but a number of other, specific Articles are also relevant.

The following Principles apply to each entity responsible for each phase of AI research, invention, innovation, dissemination and application.

1. Assess Positive and Negative Impacts and Implications

1.1 Conceive and design only after ensuring adequate understanding of purposes and contexts

"Training, validation and testing data sets shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, behavioural or functional setting within which the high-risk AI system is intended to be used" (Art. 10.4).

Art. 10.4 contains reference to context, but it only applies to "training, validation and testing data sets", and not to broader aspects of the design, and not to the operation nor to the use of AI systems. A score of 0.1 is assigned.

1.2 Justify objectives

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

1.3 Demonstrate the achievability of postulated benefits

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

1.4 Conduct impact assessment, including risk assessment from all stakeholders' perspectives

"A risk management system shall be established, implemented, documented and maintained ... " (Art. 9.1), and "When implementing the risk management system ..., specific consideration shall be given to whether the high-risk AI system is likely to be accessed by or have an impact on children" (Art. 9.8)

Due to the provisions' vagueness, mis-direction to organisational risk assessment rather than impact assessment from the perspective of the people affected, and limitation of consideration of impact to that on children, a score of 0.3 is assigned.

1.5 Publish sufficient information to stakeholders to enable them to conduct their own assessments

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

1.6 Conduct consultation with stakeholders and enable their participation in design

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

1.7 Reflect stakeholders' justified concerns in the design

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

1.8 Justify negative impacts on individuals ('proportionality')

Mentions of proportionality abound in relation to the potential impacts of regulation on organisations, but nothing was found that imposes any such requirement in relation to potential impacts on people affected by High Risk AI Systems, including in Art. 9 (Risk Management).

1.9 Consider alternative, less harmful ways of achieving the same objectives

Nothing was found that imposes any such requirement, including in Art. 9.4 (Risk Management).

2. Complement Humans

2.1 Design as an aid, for augmentation, collaboration and inter-operability

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

2.2 Avoid design for replacement of people by independent artefacts or systems, except in circumstances in which those artefacts or systems are demonstrably more capable than people, and even then ensuring that the result is complementary to human capabilities

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

3. Ensure Human Control

3.1 Ensure human control over AI-based technology, artefacts and systems

"High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons"(Art. 14.1), possibly encompassing detection of "anomalies, dysfunctions and unexpected performance" (14.4(a)), awareness of the possibility of "automation bias" (14.4(b)), 'correct interpretation of the output' (14.4(c)), and an ability "to decide ... not to use the high-risk AI system or [to] otherwise disregard, override or reverse the output" (14.4(d)).

Because this is a (qualified) obligation only of providers, this provision alone does nothing to ensure the features are effectively designed, effectively communicated to user organisations and thence to users, and applied, appropriately or even at all. A score of 0.5 is assigned.

3.2 In particular, ensure human control over autonomous behaviour of AI-based technology, artefacts and systems

"High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons"(Art. 14.1).

"The measures ... shall enable the individuals to whom human oversight is assigned ... as appropriate to the circumstances ... to ... (e) be able to intervene on the operation of the high-risk AI system or interrupt the system through a 'stop' button or a similar procedure" (Art. 14.4).

The first provision could be interpreted as observation without control. The second adds control to the requirement, but is subject to the unclear and potentially substantial qualification "as appropriate to the circumstances". A score of 0.7 is assigned.

3.3 Respect people's expectations in relation to personal data protections, including:

- their awareness of data-usage
- their consent
- data minimisation
- public visibility and design consultation and participation
- the relationship between data-usage and the data's original purpose

The EU Proposal asserts that "The proposal is without prejudice and complements the General Data Protection Regulation (Regulation (EU) 2016/679)" (EM s.1.2).

However, "To the extent that it is strictly necessary for the purposes of ensuring bias monitoring, detection and correction in relation to the high-risk AI systems, the providers of such systems may process special categories of personal data referred to in Article 9(1) of Regulation (EU) 2016/679, Article 10 of Directive (EU) 2016/680 and Article 10(1) of Regulation (EU) 2018/1725, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons, including technical limitations on the re-use and use of state-of-the-art security and privacy-preserving

measures, such as pseudonymisation, or encryption where anonymisation may significantly affect the purpose pursued" (Art. 10.5).

Nothing in the Requirements and Obligations draws the attention of providers and users to the GDPR. Nothing in the GDPR requires public visibility, public design consultation or public design participation. A score of 0.9 is assigned.

3.4 Respect each person's autonomy, freedom of choice and right to self-determination

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

3.5 Ensure human review of inferences and decisions prior to action being taken

"For [AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons] ... ensure that ... no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons" (Art. 14.5).

The clause applies only to 1 of 20 categories of High-Risk AI systems, and not at all to any of the great many other-than-high-risk AI systems, and it is in qualified by the unclear expression "[action or decision taken] on the basis of the identification resulting from the system". A score of 0.2 has been assigned.

3.6 Avoid deception of humans

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

3.7 Avoid services being conditional on the acceptance of AI-based artefacts and systems

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

4. Ensure Human Safety and Wellbeing

4.1 Ensure people's physical health and safety ('nonmaleficence')

"[I]nstructions for use ... shall specify ... any known or foreseeable circumstance, related to the use of the high- risk AI system in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, which may lead to risks to the health and safety ..." (Art. 13.2-3).

"Human oversight shall aim at preventing or minimising the risks to health, safety ... that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter" (Art. 14.2).

"Any ... user or other third-party shall be considered a provider for the purposes of this Regulation and shall be subject to the obligations of the provider under Article 16, in any of the following circumstances: ...
(b) they modify the intended purpose of a high-risk AI system already placed on the market or put into service;
(c) they make a substantial modification to the high-risk AI system" (Art. 28.1).

However, the complexity of wording ensures the existence of considerable uncertainty, a great deal of scope for regulatory prevarication, and large numbers of loopholes. A score of 0.7 is assigned.

4.2 Ensure people's psychological safety, by avoiding negative effects on their mental health, emotional state, inclusion in society, worth, and standing in comparison with other people

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

4.3 Contribute to people's wellbeing ('beneficence')

Nothing was found that imposes any such requirement.

4.4 Implement safeguards to avoid, prevent and mitigate negative impacts and implications

"In identifying the most appropriate risk management measures, the following shall be ensured ... elimination or reduction of risks ... adequate mitigation ..." (Art. 9.4).

"High-risk AI systems that continue to learn ... shall be developed in such a way to ensure that possibly biased outputs due to outputs used as an input for future operations ('feedback loops') are duly addressed with appropriate mitigation measures" (Art. 15.3).

The second is clumsy, unclear and may have little impact. However, the first of the two provides coverage of the Principle, so a score of 1.0 is assigned.

4.5 Avoid violation of trust

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

4.6 Avoid the manipulation of vulnerable people , e.g. by taking advantage of individuals' tendencies to addictions such as gambling, and to letting pleasure overrule rationality

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management).

5. Ensure Consistency with Human Values and Human Rights

5.1 Be just / fair / impartial, treat individuals equally, and avoid unfair discrimination and bias, not only where they are illegal, but also where they are materially inconsistent with public expectations

The term 'discrimination' occurs 26 times in the Explanatory Memorandum and Preamble, but not at all in the Principles. Similarly, the terms 'just' and 'justice'are used in preliminary text but not in the Principles.

"Appropriate data governance and management practices shall apply ...[concerning in particular] ... (f) examination in view of possible biases" (Art. 10.2).

" ... enable the individuals to whom human oversight is assigned to do ... as appropriate to the circumstances (b) remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias')" (Art. 14.4).

"High-risk AI systems that continue to learn ... shall be developed in such a way to ensure that possibly biased outputs due to outputs used as an input for future operations ('feedback loops') are duly addressed with appropriate mitigation measures" (Art. 15.3).

The first two passages only require vigilance, in the second case qualified by "as appropriate to the circumstances", not avoidance of bias. The third applies only to machine learning (ML) applications, and appears to authorise bias in that it requires only mitigation not prevention.

A score of 0.3 is assigned.

5.2 Ensure compliance with human rights laws

"[I]nstructions for use ... shall specify ... any known or foreseeable circumstance, related to the use of the high-risk AI system in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, which may lead to risks to ... fundamental rights" (Art. 13.2-3).

"Human oversight shall aim at preventing or minimising the risks to ... fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter" (Art. 14.2).

"Any ... user or other third-party shall be considered a provider for the purposes of this Regulation and shall be subject to the obligations of the provider under Article 16, in any of the following circumstances: ...
(b) they modify the intended purpose of a high-risk AI system already placed on the market or put into service;
(c) they make a substantial modification to the high-risk AI system" (Art. 28.1).

However, the complexity of wording ensures the existence of considerable uncertainty, a great deal of scope for regulatory prevarication, and large numbers of loopholes. Further, nothing in the Requirements and Obligations reminds providers and users of their obligations in relation to human rights. A score of 0.7 is assigned.

5.3 Avoid restrictions on, and promote, people's freedom of movement

As per P5.2, a score of 0.7 is assigned.

5.4 Avoid interference with, and promote privacy, family, home or reputation

As per P5.2, a score of 0.7 is assigned.

5.5 Avoid interference with, and promote, the rights of freedom of information, opinion and expression, of freedom of assembly, of freedom of association, of freedom to participate in public affairs, and of freedom to access public services

As per P5.2, a score of 0.7 is assigned.

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')

"Human oversight shall aim at preventing or minimising the ... that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter" (Art. 14.2).

"Any ... user or other third-party shall be considered a provider for the purposes of this Regulation and shall be subject to the obligations of the provider under Article 16, in any of the following circumstances: ...
(b) they modify the intended purpose of a high-risk AI system already placed on the market or put into service;
(c) they make a substantial modification to the high-risk AI system" (Art. 28.1).

Nothing requires any evaluation to be undertaken as to whether interference with human values or human rights is outweighed by other factors. Moreover, the complexity of wording ensures the existence of considerable uncertainty, a great deal of scope for regulatory prevarication, and large numbers of loopholes. A score of 0.5 is assigned.

6. Deliver Transparency and Auditability

6.1 Ensure that the fact that a process is AI-based is transparent to all stakeholders

Art. 13 requires transparency of the operation of AI systems to users, but not to those affected by the system. Users have no obligations relating to transparency. A small proportion of High Risk AI systems are also subject to the Art. 52 transparency provisions. A score of 0.3 is assigned.

6.2 Ensure that data provenance, and the means whereby inferences are drawn from it, decisions are made, and actions are taken, are logged and can be reconstructed

"High-risk AI systems shall be designed and developed with capabilities enabling the automatic recording of events ('logs') ... [conformant with] recognised standards or common specifications ... [ensuring an appropriate] level of traceability ... " (Arts. 12.1-3).

"Users of high-risk AI systems shall keep the logs automatically generated by that high-risk AI system, to the extent such logs are under their control" (Art. 29.5).

These provisions address only one small part of the Principle, logging, and even then only in respect of "events", not of "means whereby inferences are drawn from it, decisions are made, and actions are taken". Further, the criterion applied is only "traceability", which is far less stringent than "means ... can be reconstructed", and user organisations are invited to make arrangements such that the logs are not under their control.

"High-risk AI systems shall be designed and developed in such a way [as] to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately" (Art. 13.1).

This provision requires sufficient transparency to enable interpretation and appropriate use, but this is far less stringent than "ensure ... means ... can be reconstructed".

A score of 0.3 is assigned.

6.3 Ensure that people are aware of inferences, decisions and actions that affect them, and have access to humanly-understandable explanations of how they came about

Nothing was found that imposes any such requirement, including in Art. 9 (Risk Management) and Art. 12 (Record-Keeping).

GDPR Art. 13.2(f), Art. 14.2(g) and Art.15.1(h) require personal data controllers to "provide the data subject with ... information ... [including, in the case of] the existence of automated decision-making ... meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject". Further, Art. 22 creates a qualified "right not to be subject to a decision based solely on automated processing". Some commentators read the combination of these provisions as somehow implying a right to an explanation for decisions whether made soon after or long after data collection, despite the absence of any such expression in the Article.

The drafting complexity is such that a wide array of qualifications and escape clauses exist, and analysis suggests that such an optimistic interpretation is unjustified (Wachter et al. 2017). In addition, these GDPR articles apply only to "a decision based solely on automated processing", and not at all where the decision is reviewed (however cursorily) by a human being. It also needs to be appreciated that, in the case of decisions based on opaque and a-rational AI/ML inferencing techniques such as neural networks, it is not possible to undertake a meaningful review.

In this assessment, the position is adopted that existing EU law does not provide any right to a humanly-understandable explanation for decisions arising from AI systems, and hence P6.3 is not satisfied.

7. Embed Quality Assurance

7.1 Ensure effective, efficient and adaptive performance of intended functions

Nothing was found that imposes any such requirement, because Art. 17 (Quality Management System) does not relate quality to the AI system's intended functions, and hence an AI system can perform very poorly but not be in breach of Art. 17.

7.2 Ensure data quality and data relevance

"Appropriate data governance and management practices shall apply [whether or not the AI system makes use of] techniques involving the training of models with data ..." (Art. 10.2, 10.6) .

Where the system makes use of "techniques involving the training of models with data", "training, validation and testing data sets" are required to "be relevant, representative, free of errors and complete" and to "have the appropriate statistical properties" (Art. 10.3).

Art. 10.2 articulates the first part of P7.2 ("Ensure data quality ..."), but, when compared with Guidelines for the Responsible Application of Data Analytics, the articulation is not comprehensive. Art. 10.3 provides a further, very specific extension to that articulation.

A score of 0.4 is assigned (60% of 0.7 of the Principle, relating to data quality).

"[T]o the extent the user exercises control over the input data, that user shall ensure that input data is relevant in view of the intended purpose of the high-risk AI system" (Art. 29.3).

Because a user organisation can avoid responsibility for only using data that is relevant, irrelevant data is permitted to cause harm to affected individuals, without any entity being liable for that harm.

A score of 0.1 is assigned (1/6th of 0.3 of the Principle, relating to relevance).

Overall, a score of 0.5 is assigned.

7.3 Justify the use of data, commensurate with each data-item's sensitivity

Nothing was found that imposes any such requirement. "Training, validation and testing data sets shall be subject to appropriate data governance and management practices [including] (a) the relevant design choices" (Art. 10.2) is too limited and too vague to contribute to the Principle.

7.4 Ensure security safeguards against inappropriate data access, modification and deletion, commensurate with its sensitivity

"High-risk AI systems shall be designed and developed in such a way that they achieve, in the light of their intended purpose, an appropriate level of ... cybersecurity ... appropriate to the relevant circumstances and the risks" (Art. 15-1,4).

The vague term 'cybersecurity' may refer to assurance of service, or toany, some or all of assurance of sustained quality of service or of data, or to assurance of access to data only by authorised entities for authorised purposes. A score of 0.7 is assigned.

7.5 Deal fairly with people ('faithfulness', 'fidelity')

Nothing was found that imposes any such requirement.

7.6 Ensure that inferences are not drawn from data using invalid or unvalidated techniques

Nothing was found that imposes any such requirement, including in Art. 17 (Quality Management System), which requires only "written policies, procedures and instructions ... [regarding] (b) techniques, procedures and systematic actions ..." (Art. 17.1), and imposes no actual quality requirements in relation to the techniques used to draw inferences.

7.7 Test result validity, and address the problems that are detected

"Training, validation and testing data sets shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, behavioural or functional setting within which the high-risk AI system is intended to be used" (Art. 10.4).

Art. 10.4 contains an oblique reference to validity in a context, but it only applies to "training, validation and testing data sets", and not to the overall design, or the operation nor to the use of AI systems, let alone to the validity of specific inferences.

"Providers ... shall put a quality management system in place [including] examination, test and validation procedures to be carried out before, during and after the development of the high-risk AI system ..." (Art. 17.1(d)). However, "the implementation ... shall be proportionate to the size of the provider's organisation" (Art. 17.2).

Because the first provision is weakened by the second, and there is no express requirement that problems that are detected are addressed, the net effect is less than full correspondence with the Principle. An overall score of 0.8 is assigned.

7.8 Impose controls in order to ensure that the safeguards are in place and effective

Nothing was found that imposes any such requirement. The EC Proposal uses the term 'controls' to refer to 'safeguards', and contains nothing about control arrangements to ensure that the intended safeguards are in place, operational and effective.

7.9 Conduct audits of safeguards and controls

Nothing was found that imposes any such requirement.

8. Exhibit Robustness and Resilience

8.1 Deliver and sustain appropriate security safeguards against the risk of compromise of intended functions arising from both passive threats and active attacks, commensurate with the significance of the benefits and the potential to cause harm

"High-risk AI systems shall be designed and developed in such a way that they achieve, in the light of their intended purpose, an appropriate level of ... robustness ... and perform consistently in those respects throughout their lifecycle" (Art. 15.1). A score of 1.0 is assigned.

8.2 Deliver and sustain appropriate security safeguards against the risk of inappropriate data access, modification and deletion, arising from both passive threats and active attacks, commensurate with the data's sensitivity

"High-risk AI systems shall be designed and developed in such a way that they achieve, in the light of their intended purpose, an appropriate level of ... cybsersecurity ... and perform consistently in those respects throughout their lifecycle" (Art. 15.1).

The vague term 'cybersecurity' may refer to assurance of service, or to assurance of any, some or all of sustained quality of service or of data, or to assurance of access to data only by authorised entities for authorised purposes. A score of 0.7 is assigned.

8.3 Conduct audits of the justification, the proportionality, the transparency, and the harm avoidance, prevention and mitigation measures and controls

Nothing was found that imposes any such requirement.

8.4 Ensure resilience, in the sense of prompt and effective recovery from incidents

"High-risk AI systems shall be resilient as regards errors, faults or inconsistencies that may occur within the system or the environment in which the system operates, in particular due to their interaction with natural persons or other systems ... [and against] attempts by unauthorised third parties to alter their use or performance by exploiting the system vulnerabilities ... appropriate to the relevant circumstances and the risks" (Art. 8.4-3-4). A score of 1.0 is assigned.

9. Ensure Accountability for Obligations

9.1 Ensure that the responsible entity is apparent or can be readily discovered by any party

"Providers shall establish and document a post-market monitoring system in a manner that is proportionate to the nature of the artificial intelligence technologies and the risks of the high-risk AI system" (Art. 61).

Because the system is for the provider and the market surveillance authority alone, and no provision is made for accessibility even by user organisations, let alone people affected by the system and their advocates, this makes only a small contribution to the Principle. A score of 0.3 is assigned.

9.2 Ensure that effective remedies exist, in the form of complaints processes, appeals processes, and redress where harmful errors have occurred

Although Ch.3, Arts. 63-68 is headed "Enforcement", Arts. 71-72 are also relevant. "Where, in the course of ... evaluation, the market surveillance authority finds that the AI system does not comply with the requirements and obligations laid down in this Regulation, it shall without delay require the relevant operator to take all appropriate corrective actions to bring the AI system into compliance, to withdraw the AI system from the market, or to recall it within a reasonable period, commensurate with the nature of the risk, as it may prescribe" (Art. 65.2 et seq. See also Art. 67).

"Member States shall lay down the rules on penalties, including administrative fines, applicable to infringements of this Regulation and shall take all measures necessary to ensure that they are properly and effectively implemented. The penalties provided for shall be effective, proportionate, and dissuasive" (Art. 71.1). (Maximum) penalties are prescribed (Arts. 71.3-5). See also Art. 72.

However, nothing in the EC Proposal requires providers or users to even receive and process complaints, let alone deal with problems that the public notifies to them. Moreover, no scope exists for affected individuals to initiate any kind of action, and hence the public appears to be entirely dependent on action being taken by each national 'market surveillance authority'.

And overall score of 0.7 is assigned.

10. Enforce, and Accept Enforcement of, Liabilities and Sanctions

10.1 Ensure that complaints, appeals and redress processes operate effectively

Nothing was found that imposes any such requirement.

10.2 Comply with external complaints, appeals and redress processes and outcomes, including, in particular, provision of timely, accurate and complete information relevant to cases

"[T]echnical documentation ... shall be drawn up before that system is placed on the market or put into service and shall be kept up-to date [and] drawn up in such a way [as] to demonstrate that the high-risk AI system complies with the requirements set out in [Arts. 8-15]" (Art. 11.1)

This makes a contribution to the provider's and user organisation's capability to participate in enforcement processes. However in itself it falls far short of an enforcement regime.

Under specified circumstances and conditions, "market surveillance authorities shall be granted access" by providers (Art. 64.1-.2). The market surveillance authority can require a provider to "withdraw the AI system from the market or to recall it ..." (Arts. 67-68).

Although the market surveillance authority has powers, the provisions create no scope for enforcement of any aspects of the EC Proposals by individuals, or by public interest advocacy organisations. A score of 0.4 is assigned.

Appendix 3: The Scope of Applicability of the EC's Proposal

This Appendix extracts key passages in the EC's Proposal that determine the scope of the regulatory regime.

For the proposed law to apply:

- four scope-conditions must all be satisfied, relating to the artefact, the entity, the geographical location, and the timeframe; AND
- one scope-condition must NOT be satisified, relating to the purpose

INCLUSIONS

1. Artefact - What things are within-scope?

"AI Systems", viz. "software that is developed with one or more [specified] techniques and approaches ... AND can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with" - Art. 3(1)

The specified "techniques and approaches" are Machine learning approaches ..., Logic- and knowledge-based approaches ... and Statistical approaches, Bayesian estimation, search and optimization methods - Annex I

EXCEPT "high-risk AI systems that are safety components of products or systems, or which are themselves products or systems, falling within the scope of [eight] acts" - Art. 2.2

A feasible interpretation is that this encompasses all forms in which 'software' presents, including:

- components such as elements within code-libraries
- independent programs
- suites of programs
- programs embedded in hardware, and
- systems comprising hardware, software and people that contain any element developed with a specific technique or approach (at least to the extent that such an element is involved in the determination or performance of a relevant action?)

However, the courts may decide otherwise.

2. Entity - Who is within-scope?

2.1 Providers

"providers placing on the market or putting into service AI systems ..." – Art. 2.1(a)

Provider "means a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge" - Art. 3(1), second occurrence - [sic: should be 3(2)].

Combining the definition with the scope declaration, a provider is an entity that satisifies **all** of these conditions:

- develops an AI system or has one developed for it;
- intends placing it on the market or putting it into service ...;
- actually places it on the market or puts it into service.

EXCEPT "public authorities in a third country" and "international organisations" that have relevant agreements with the EU or part thereof - Art. 2.4

EXCEPT 'placing on the market' means "the first making available of an AI system on the Union market" - Art. 3(9)

cf. 'making available on the market' means "any supply of an AI system for distribution or use on the Union market in the course of a commercial activity, whether in return for payment or free of charge" - Art. 3(10)

The use in the definition in 2.1(a) and 3(2) of 'placing' rather than 'making available' creates the possibility of subsequent offerors of any particular AI system, e.g. purchasers of licences and agencies, not being within-scope of the Regulation.

EXCEPT 'putting into service' means "the supply of an AI system for first use directly to the user or for own use on the Union market for its intended purpose" - Art. 3(11)

The use in the definition in 2.1(a) and 3(2) of 'putting into service', and in 3(11) of 'first' together appear to exclude from the scope the subsequent supply of previously-supplied AI systems.

2.2 Users

"users of AI systems" – Art. 2.1(b)

User means "any natural or legal person, public authority, agency or other body using an AI system under its authority ..." - Art. 3(4)

EXCEPT "where the AI system is used in the course of a personal non-professional activity" - Art. 3(4)

3. Geographical Location - Where is within-scope?

3.1 Providers

"providers [offering or deploying] AI systems irrespective of whether those providers are established within the Union or in a third country" – Art. 2.1(a)

"providers ... of AI systems that are located in a third country, where the output produced by the system is used in the Union" – Art. 2.1(c)

Clarification is needed as to whether that which has to be located within the EU to be within-scope is the provider, the system, or both. The 2.1(c) formulation might apply only to the system.

The 2.1(a) formulation might apply only to the provider's action.

A feasible interpretation is that the geographical scope-condition is satisfied if **any** of the provider, the provider's action **or** the system is within the EU; but the courts may decide otherwise.

3.2 Users

"users of AI systems located within the Union" – Art. 2.1(b)

"users ... in a third country, where the output produced by the system is used in the Union" – Art. 2.1(c)

Clarification is needed as to whether that which has to be located within the EU to be within-scope is the user, the system, the use, or two or more of them.

A feasible interpretation is that the geographical scope-condition is satisfied if **any** of the user, the system **or** the use of the output is within the EU; but the courts may decide otherwise.

4. Timeframe – When is within-scope?

No provision is apparent in the EC Proposal relating to the date when the law would come into force.

Nor is there any apparent provision for a period of applicability or a sunset clause.

EXCLUSION

5. Purpose - What uses are out-of-scope?

"... developed or used exclusively for military purposes" - Art. 2.3

'intended purpose' means "the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation" - Art. 3(12)

References

Clarke R. (2017) '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. (2019) 'Principles and Business Processes for Responsible AI' Computer Law & Security Review 35, 4 (2019) 410-422, PrePrint at http://www.rogerclarke.com/EC/AIP.html.
See also the current versions of The 50 Principles in HTML, and The 50 Principles in PDF

EC (2019) 'Ethics Guidelines for Trustworthy AI' High-Level Expert Group on Artificial Intelligence, European Commission, April 2019, at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

EC (2021) 'Proposal for a Regulation on a European approach for Artificial Intelligence' European Commission, 21 April 2021, at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=75788

Primary Document (107 pp.) at https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF

Annexes (16 pp.) at https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_2&format=PDF

Press Release at https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682

Greenleaf G.W. (2021) 'The 'Brussels effect' of the EU's 'AI Act' on data privacy outside Europe' Privacy Laws & Business 171 (June 2021) 1-7

Veale M M. & Borgesius F.Z. (2021) 'Demystifying the Draft EU Artificial Intelligence Act' SocArXiv, 6 July 2021, at https://osf.io/preprints/socarxiv/38p5f/

Wachter S., Mittelstadt B. & Floridi L. (2017) 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' International Data Privacy Law 7, 2 (May 2017) 76-99, at https://academic.oup.com/idpl/article/7/2/76/3860948

Author Affiliations

Roger Clarke is Principal of Xamax Consultancy Pty Ltd, Canberra. He is also a Visiting Professor in Cyberspace Law & Policy at the University of N.S.W., and a Visiting Professor in the Research School of Computer Science at the Australian National University.
