SHERPA

Regulatory options for AI and big data

R Rodrigues, A Panagiotopoulos, B Lundgren, S Laulhé Shaelou, A Grant

D3.3: 20 December 2019

This project has received funding from the
European Union's Horizon 2020 Research and Innovation Programme
Under Grant Agreement no. 786641

Document Control

Deliverable	D3.3 Report on regulatory options
WP/Task Related	WP3
Delivery Date	20 December 2019
Dissemination Level	PU
Lead Partner	Trilateral Research: Rowena Rodrigues, Adam Panagiotopoulos, David Wright, Tally Hatzakis
Contributors	University of Central Lancashire Cyprus (Stéphanie Laulhé Shaelou, Amy Grant); University of Twente (Björn Lundgren, Kevin Macnish, Mark Ryan); Andreas Andreou (AHR), Thamar Zijlstra, Marlou Bijlsma (NEN) (inputs, comments and feedback)
Reviewers	Bernd Stahl (De Montfort University) Alexey Kirichenko (F-Secure) University of Central Lancashire Cyprus SHERPA Stakeholder Advisory Board (feedback provided as listed in Annex 1)
Abstract	This report reviews various regulatory options that support the ethical and/or responsible development of smart information systems (AI and big data). Its insights will be useful to policymakers as a guide to making policy and regulatory decisions. It presents snapshots of policy and other stakeholder perspectives on the regulation of AI and big data and discusses how AI challenges regulation. It looks at EU aspirations for better law-making and presents key considerations for regulating AI and big data.
Key Words	regulation, regulatory options, regulatory gaps, AI, smart information systems, analysis

Revision History

Version	Date	Author(s)	Reviewer(s)	Notes
0.1	May 2019	Rowena Rodrigues	Task 3.3 team	Outline
0.2	July 2019	Rowena Rodrigues	Task 3.3 team	Develop outline; start research
0.3	6 December 2019	Rowena Rodrigues, Adam Panagiotopoulos, Stéphanie Laulhé Shaelou, Amy Grant, Björn Lundgren	Bernd Stahl Alexey Kirichenko Stakeholder Board	Second draft
0.4	20 December 2019	Rowena Rodrigues	DMU	Final version

Table of Contents

Executive Summary 5

List of figures 8

List of tables 8

List of acronyms/abbreviations 9

Glossary of terms 11

1. Introduction 13

2. Methodology and scope 15

3. Policy positions and perspectives on regulation of AI and big data 18

3.1 International policy level 18

3.2 EU policy level 20

3.3 National policy level 21

3.4 General analysis and conclusion 25

4. Regulatory options 26

4.1 Identification of options 26

4.2 Assessment of options: analysis of findings and results 29

4.3 How AI challenges regulation and EU aspirations for better law-making 62

5. Conclusion: key considerations for regulating AI and big data 69

References 72

Annexes 78

1. List of stakeholders consulted 78

2. Scoping paper 79

3. Final list of options studied 79

4. Individual assessments of proposed options 83

4.1. Moratorium on the development of 'lethal autonomous robotics'/offensive LAWS 83

4.2. Binding Framework Convention for AI 88

4.3. Legislative framework for oversight over the human rights compliance of AI systems 94

4.4. Legal framework for public authorities to carry out HRIAs 100

4.5. Convention on human rights in the robot age 106

4.6. CEPEJ European Ethical Charter on the use of AI in judicial systems 111

4.7. International Artificial Intelligence Organization 117

4.8. Global legal AI and/or robotics observatory 124

4.9. EU-level special list of robot rights 129

4.10. Adoption of common Union definitions: cyber physical systems, autonomous systems, smart autonomous robots 134
4.11. Creating electronic personhood status for autonomous systems 138
4.12. Establishment of a comprehensive Union system of registration of advanced robots 143
4.13. General fund for all smart autonomous robots/individual fund 147
4.14. Mandatory consumer protection impact assessment 152
4.15. EU Taskforce of field specific regulators for AI/big data 157
4.16. Algorithmic Impact Assessments under the GDPR 161
4.17. Voluntary/mandatory certification of algorithmic decision systems (ADS) 170
4.18. DEEP FAKES Accountability Act (H.R. 3230) 176
4.19. Algorithmic Accountability Act of 2019 (HR 2231) 185
4.20. Directive on Automated Decision-Making (Canada) 190
4.21. US Food and Drug Administration regulation of adaptive AI/ML technology 195
4.22. New statutory duty of care for online harms (UK Government) 199
4.23. Redress by design mechanisms for AI 205
4.24. Register of algorithms used in government 210
4.25. Digital Authority 218
4.26. Independent cross-sector advisory body (Centre for Data Ethics and Innovation) 222
4.27. FDA for algorithms 230
4.28. US Federal Trade Commission to regulate robotics 235
4.29. Using anti-trust regulations to break up big tech and appoint regulators 238
4.30. Three-level obligatory impact assessments for new technologies 244
4.31. Regulatory sandboxes 248
5. Views of other stakeholders on AI and regulation 253
5.1 Academia 253
5.2 Media 255
5.3 Industry and professional associations 256
5.4 Civil society 258
5.5 The public 260

Executive Summary

Purpose and scope of the report

This report reviews various regulatory options that support the ethical and/or responsible development of smart information systems (AI and big data). Its insights will be useful to policymakers as a guide to making policy and regulatory decisions. It presents snapshots of policy and other stakeholder perspectives on the regulation of AI and big data and discusses how AI challenges regulation. It also looks at EU aspirations for better law-making and presents key considerations for regulating AI and big data. The study adopted a wide, inclusive understanding of 'regulatory options' to cover **proposals for laws, bodies** and other **regulatory tools and mechanisms**. Well-established legislation has been excluded.

31 options were reviewed: **8 at the international level; 9 at the EU-level; 11 national,** and **3 cross-overs**.

International	EU-level	National	Cross-over
1.Moratorium on LARs/LAWS 2.Binding Framework Convention for AI 3.Legislative framework for independent and effective oversight 4.Legal for human rights impact assessments (HRIAs) on AI systems 5.Convention on human rights in the robot age 6.CEPEJ European Ethical Charter 7.International Artificial Intelligence Organization 8.Global legal AI and/or robotics observatory	1.EU-level special list of robot rights 2.Adoption of common Union definitions 3.Creating electronic personhood status for autonomous systems 4.Establishment of a comprehensive Union system of registration of advanced robots 5.General fund for all smart autonomous robots 6.Mandatory consumer protection impact assessment 7.EU Taskforce of field specific regulators for AI/big data 8.Algorithmic Impact Assessments under the GDPR 9.Voluntary/mandatory certification of algorithmic decision systems	1.DEEP FAKES Accountability Act 2.Algorithmic Accountability Act 3.Directive on Automated Decision-Making 4.US Food and Drug Administration regulation of adaptive AI/ML technology 5.New statutory duty of care for online harms 6.Redress by design mechanisms for AI 7.Register of algorithms used in government 8.Digital Authority 9.Independent cross-sector advisory body (CDEI) 10.FDA for algorithms 11.US Federal Trade Commission to regulate robotics	1.Using anti-trust regulations to break up big tech and appoint regulators 2.Three-level obligatory impact assessments for new technologies 3.Regulatory sandboxes

Stakeholder positions and perspectives on regulation of AI and big data

Although there are disagreements, important actors seem to **aim for harmonized rules**. However, there is great **variation as to the specificity** of regulatory proposals. Most **push for a heavier rather than a lighter touch**, but there are clear disagreements. Proposals often **combine risk-based**

approaches with principle-based regulation. There is an understanding among industrial proponents that **regulations are needed, but there is disagreement and ambiguity** about self-regulation, co-regulation, or full regulation. The most common worries are that **a heavy-touch will restrict innovation**, while **a light-touch will leave individuals and society exposed to risks to fundamental values or human rights**. The challenge for any regulations is how to promote good AI development and use, how to minimize the creation of bad AI or misuse of AI-technology, and how to increase its security (reliability and resilience). Proposals for regulations almost always address ethical concerns and human rights.

Regulatory options study results

The reviewed proposals aim to directly regulate AI and suggest governance mechanisms for it. Solutions to the risks of AI are suggested either as autonomous and independent, or as supporting measures to the existing legal and technological *status quo*.

Regulatory trends

There are three main 'regulatory trends':

- a **commonly recognised need for AI regulation, soft or hard,** and, ideally at a supra-national level;
- proposals for the **creation of a regulatory agency/body**;
- calls to review **the existing legal framework and either revise it to address the challenges and risks of AI or provide for specific legal acts or other instruments** (such as frameworks and codes of conduct and tools) to specifically govern AI.

Limitations, risks and challenges for the adoption and implementation of the reviewed options

- **Limitations** include, broad scope, lack of specific features such as transparency, over-focus on specific criteria such as high focus on bias and discrimination, and neglect of other fundamental rights and freedoms, resource constraints;
- **Risks** include, considering options as panacea or replacement of existing frameworks, privatisation of regulation and scrutiny, conflicts with intellectual property rights, negative impact on human rights, mission creep;
- **Challenges** include, confusion and ill-applied measures, resistance from stakeholders, operational burdens, sustainability, political will.

Human rights and ethics

Nearly half of the reviewed options explicitly support human rights. Others might have this effect more indirectly, and some do not address this (and/or human rights falls outside their scope or has not been sufficiently defined). The most discussed human rights and freedoms include: **data protection, dignity, equality, freedom from discrimination, privacy, and the right to life.**

Key **ethical principles** that featured repeatedly in many of the reviewed proposals include: **fairness, transparency, accountability, prohibition or minimisation of bias, privacy, prevention or reduction of harm, respect for human rights, democracy or democratic governance, human autonomy, rule of law and human safety.**

Non-feasible options

Options identified as non-feasible, or which have drawn criticism and are potentially most likely to be affected by future developments, include the proposal for anti-trust regulations, and the proposal for the US DEEP FAKES Accountability Act.

Most promising options

- The three international-level options that look most promising are: the *Binding Framework Convention*, the CEPEJ *European Ethical Charter*, and the Legislative Framework for independent and effective oversight.
- At the EU-level, the general fund for smart robots and the Common Union registration of robots fared extremely well; with algorithmic impact assessments under the General Data Protection Regulation (GDPR), and voluntary/mandatory certification of ADS not far behind.
- At the national level, the most promising are redress by design, followed by proposed 'specific' legislation. Bodies such as the CDEI and Digital Authority also look promising, as does the proposal for a register of algorithms used in government.

How AI challenges regulation and EU aspirations for better law-making

AI challenges regulation in various ways and this should be recognised and addressed when regulating AI. Timing **AI and big data** regulation well (though this is very challenging) will contribute to its effectiveness and meaningfulness. The report illustrates how this might work at different stages of the AI application/system lifecycle. Based on the EU aspirations for better law-making, we recommend **greater wisdom/prudence** to be applied with regard to regulatory decision-making in the AI context, and **regulatory serenity**.

Key considerations for regulating AI and big data

- *Striking a balance between enabling beneficial AI and risk mitigation*

Striking a balance between enabling beneficial technologies and risk mitigation is complex, not always possible, and requires policymakers and legislators to understand the differential nature of AI and big data risks. It also requires an understanding of how AI actors will respond to the regulatory actions and incentives. The **possibility of regulatory failure** should also be considered – the amplification of risks due to reckless or casual and unconsidered adoption of laws to regulate AI, or even the adoption of bad AI laws.

- *Smart mixing for good results*

The challenge is to find a smart mix of instruments (i.e., technical, standards, law and ethical) in consultation with stakeholders to facilitate responsible innovation. The protection of ethical principles and human rights calls for a mix of voluntary, interventionist, and facilitative regulatory measures. Regulation also needs to be agile.

- *Super-security for high-risk/high-impact AI*

Given the high risks of non-obvious/hidden security vulnerabilities or malicious manipulation of AI to cause serious harm and threats to life and society, there is a need to actively discuss and work on regulatory options that support **super-secure AI where it has high likelihood and high severity of risk/impact on rights and freedoms of individuals and especially the vulnerable.** Security could be seen as a standalone requirement and regulatory focus (a regulatory aim in itself), and could be considered as the means to achieve the safeguarding of ethics and human rights.

What next?

A policy brief taking into account the results of this Study is also planned. SHERPA will use the results of the report to support its Delphi study on ethics and human rights. The report will also feed into planned focus groups in 2020 where stakeholders will discuss regulatory options further. After this, SHERPA recommendations for action by various stakeholders will be finalised.

List of figures

Figure 1: Identified regulatory options (proposals for laws, bodies and other mechanisms)
Figure 2: Stakeholders benefitted
Figure 3: Reviewed options and human rights
Figure 4: Featured ethical principles – top scorers
Figure 5: Other ethical principles featuring in the studied proposals
Figure 6: Impact on innovation
Figure 7: International options: scores
Figure 8: EU-level options: scores
Figure 9: National options: scores
Figure 10: Cross-over options: scores

List of tables

Table 1: List of acronyms/abbreviations
Table 2: Glossary of terms
Table 3: Finalised list of criteria used in analysis of regulatory options
Table 4: How AI challenges different forms of regulation
Table 5: When to regulate

List of acronyms/abbreviations

Abbreviation	Explanation
ACM	Association for Computing Machinery
ADM	Automated Decision-Making
ADS	Algorithmic decision systems
AI	Artificial intelligence
AIA	Algorithmic Impact Assessment
AI HLEG	High-Level Expert Group on Artificial Intelligence
BDVA	Big Data Value Association
CAHAI	Ad-hoc Committee on Artificial Intelligence (Council of Europe)
CDEI	Centre for Data Ethics and Innovation
CEPEJ	Council of Europe European Commission for the efficiency of justice
CoE	Council of Europe
COFECE	Comisión Federal de Competencia Económica
DPA	Data protection authority
EC	European Commission
ECHR	European Convention on Human Rights
EDPB	European Data Protection Board
ENISa	European Union Agency for Cybersecurity
EU	European Union
EurAI	European Association of AI
FDA	Food and Drug Administration
GDPR	General Data Protection Regulation
HRIA	Human rights impact assessments
IAIO	International Artificial Intelligence Organisation
MSMEs	Ministry of Micro, Small and Medium Enterprises
ICDPPC	International Conference of Data Protection and Privacy Commissioners
IEEE	Institute of Electrical and Electronics Engineers

Abbreviation	Explanation
INAI	Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales
ISO	International Organization for Standardization
ITU	International Telecommunication Union
LARs	Lethal autonomous robotics
LAWS	Lethal Autonomous Weapons Systems
ML	Machine Learning
OECD	Organisation for Economic Co-operation and Development
R&D	Research and development
SDG	Sustainable Development Goals
SIENNA	Stakeholder-Informed Ethics for New technologies with high socio-ecoNomic and human rights impAct project
SIS	Smart information systems
SME	Small and Medium Enterprise
SSRN	Social Science Research Network
UAI	Unboxing Artificial Intelligence
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
US	United States

Table 1: List of acronyms/abbreviations

Glossary of terms

Term	Explanation
Co-regulation	Establishment of a legislated framework that declares requirements, enforcement processes and sanctions, and allocates powers and responsibilities to appropriate regulatory agencies, but delegates development and maintenance of the detailed obligations to an independent body, comprising representatives of all stakeholder groups, including the various categories of the affected public.[1] According to the *Interinstitutional Agreement on Better Law-making* (2003)[2]: Co-regulation "means the mechanism whereby a Community legislative act entrusts the attainment of the objectives defined by the legislative authority to parties which are recognised in the field (such as economic operators, the social partners, non-governmental organisations, or associations)."
Full regulation	An approach where binding legal rules are used to specify the behaviour required of organisations or individuals. It is appropriate to address activities with potentially serious risks of impacts for the economy, the environment or individuals, and where legal certainty and enforcement backed by legal sanctions are necessary. It may also be the only available option if there is no scope for "softer" self-regulatory actions by business organisations or when such approaches have failed.[3] Examples: EU legislation backed by enforcement by an independent regulator; national legislation backed by enforcement by a regulatory body, e.g., national legislation on autonomous vehicles.
General regulation	Something which regulates something more generally and widely, e.g., the *General Data Protection Regulation*, versus something sector/domain/object-specific.
Harmonisation	Aims to create consistency of laws, regulations, standards and practices.
Heavy touch/heavy-handed approach	Directly opposed to light-touch (see below), includes use of greater regulation, enforcement and control. This approach is more prescriptive.
Light-touch approach	Refers to policy approaches that rely on private markets more than regulation or measures, to create a "minimal regulatory environment". Largely a hands-off approach. This approach is less prescriptive.
Principles-based regulation:	An approach that moves away from reliance on detailed, prescriptive rules and relies more on high-level, broadly-stated rules or principles to set the standards by which regulated firms must conduct business.[4]

[1] Clarke, Roger, "Regulatory Alternatives for AI", 9 Feb 2019. http://www.rogerclarke.com/EC/RAI.html
[2] No longer in force.
[3] https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-18_en_0.pdf
[4] See http://eprints.lse.ac.uk/62814/1/__lse.ac.uk_storage_LIBRARY_Secondary_libfile_shared_repository_Content_Black,%20J_Principles%20based%20regulation_Black_Principles%20based%20regulation_2015.pdf

Term	Explanation
Regulation	The act or process of controlling by rule of restriction (Black's Law Dictionary).
Regulatory option	In this context, widely and broadly scoped as including both governance and/or legislative or legal regulation proposals, but excluding technical proposals, codes of conduct, ethical codes and standardisation options/proposals. These are covered elsewhere in SHERPA.
Risk-based approach	Involves targeting enforcement resources on the basis of assessments of the risks posed by a regulated person or firm to the regulator's objectives; risk-based regulation offers an evidence-based means of targeting the use of resources and of prioritising attention to the highest risks, in accordance with a transparent, systematic and defensible framework.[5]
Stakeholder	A relevant actor (persons, groups or organisations) who: (1) might be affected by the project; (2) have the potential to implement the project's results and findings; (3) have a stated interest in the project fields; and, (4) have the knowledge and expertise to propose strategies and solutions in the fields of SIS and artificial intelligence (AI)
Self-regulation	The possibility for economic operators, and the social partners, non-governmental organisations or associations to adopt amongst themselves and for themselves common guidelines at European level (particularly codes of practice or sectoral agreements (as defined in the 2003 *EU Interinstitutional Agreement on Better Law-making*)
Standard	Standards are approved by a recognized body which is responsible for establishing rules, guidelines or characteristics for products or related processes and production methods. Compliance is not mandatory. They may also deal with terminology, symbols, packaging, marking and labelling requirements.[6]
SIS	The combination of Artificial Intelligence and big data analytics.

Table 2: Glossary of terms

[5] See Black, Julia, and Robert Baldwin, "Really responsive risk-based regulation," *Law & Policy,* 32.2, 2010, pp. 181-213

[6] Source: WTO TBT booklet, https://www.wto.org/english/res_e/publications_e/tbttotrade_e.pdf

1. Introduction

This report explores and provides insights on various regulatory options that support the ethical and responsible development of smart information systems. Its insights will be useful to policymakers as a guide to making policy and regulatory decisions. SIS (the combination of Artificial Intelligence (AI) and big data analytics) have the potential to bring great benefits to society, and at the same time to cause great disruption, especially adversely affecting human rights. Regulation might help alleviate some of these concerns, but, is not without its challenges.[7] As Etzioni says, "The difficulty of regulating AI does not absolve us from our responsibility to control AI applications. Not to do so would be, well, unintelligent".[8]

There is substantial existing academic and policy literature on the regulation of AI and big data.[9] This focuses broadly on the regulation of AI, and big data[10], as well as on more specific aspects, e.g., safeguards for automated decision-making, algorithmic accountability[11], cyber skirmishes[12]).

This report takes this into account, and builds upon previous SHERPA findings, particularly the case studies, scenarios, and *Task 1.5: Findings and the identification and analysis of the challenges and regulatory gaps* (law and compliance); it also takes into account the international and legal analysis of AI and robotics in SIENNA (completed in March 2019).[13]

The report:

- Examines some regulatory options, specifically new proposals relevant to AI and big data and highlights the advantages, risks and challenges, and obstacles to the implementation of such options, their chances of success, the roles of relevant actors, and impacts on AI and big data stakeholders.
- Explores whether and how present and proposed regulatory interventions relating to AI and big data adhere to European aspirations for better regulation.
- Analyses whether embedding ethics and human rights in AI and big data is best served by an interventionist, or a flexible and facilitative approach.

---

[7] For instance, legislation can be slow, or sometimes 'knee-jerk' and hastily implemented. What aspects need further regulation (algorithmic design/coding/outcomes or impacts), and how should this be achieved?

[8] Etzioni, Oren, "Point: Should AI Technology Be Regulated?: Yes, and Here's How", *Communications of the ACM,* December 2018, Vol. 61 No. 12, pp. 30-32.

[9] A search on SSRN using key words "regulation AI big data" threw up 36 results; on LawArxiv it threw up 1,173 results.

[10]https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/DTM_Big%20Data%20v1_0.pdf; Cumbley, Richard, and Peter Church, "Is "big data" creepy?" *Computer Law & Security Review* 29.5 (2013): 601-609; Mayer-Schonberger, Viktor, and Yann Padova, "Regime change: enabling big data through Europe's new data protection regulation", *Colum. Sci. & Tech. L. Rev.* 17, 2015: 315. Sokol, D. Daniel, and Roisin Comerford,"Antitrust and Regulating Big Data," *Geo. Mason L. Rev.* 23, 2015, p. 1129.

[11] Kaminski, Margot E., "Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability", *Southern California Law Review*, Vol. 92, No. 6, 2019; U of Colorado Law Legal Studies Research Paper No. 19-9.

[12] Taddeo, Mariarosaria and Floridi, Luciano, "Regulate Artificial Intelligence to Avert Cyber Arms Race", *Nature* 556, 296-298 (2018); doi: 10.1038/d41586-018-04602-6.

[13] SIENNA, *D4.2: Analysis of the legal and human rights requirements for AI and robotics in and outside the EU*, 2019.

The report provides useful insights for policy-makers and advances the discussion on the regulatory proposals. It will also feed into the SHERPA Delphi study and subsequent activities of the project (e.g., the focus groups where stakeholders will discuss regulatory options further).

Structure of the report

Section 2 discusses the report's methodology and scope. Section 3 presents snapshots of policy positions and perspectives on the regulation of AI and big data, with views, gaps identified and their regulatory recommendations (supported by snapshot views of other stakeholders in Annex 5). Section 4 identifies and examines various regulatory proposals for new laws, regulatory bodies and other tools and mechanisms, made in relation to AI and big data, and examines these based on set criteria. It also discusses how AI challenges regulation and EU aspirations for better-law making. Section 5 presents key considerations for regulating AI and big data.

Please note, this report focuses on AI *and* big data (the scope of SHERPA); where AI is used exclusively this is generally taken into account. Further, we recognise that the functions of law and ethics are different, but can play out complementarily (informing, influencing, facilitating governance) to support the achievement of societal goals especially in the context of AI and big data.

Relation and substantive connection with ongoing SHERPA activities

SHERPA Task 3.2 focussed on developing guidelines for research and innovation in and with SIS, to provide ethical guidelines, along the lines of a code of conduct. This report thus excludes from its scope proposals for ethical guidelines and codes of conduct. However, ethical guidelines in themselves may not be sufficient to address and/or alleviate ethical risks from AI and big data. There might be legal issues that are difficult to deal with, or in practice are acknowledged but not implemented, and require regulation.

The synergy between SHERPA Task 3.3 (whose results this report documents) and Task 3.4 is important. SHERPA Task 3.4 explores the need for and feasibility of standardization for AI and big data. Building on existing good practices, research results from other SHERPA work packages, and other input from stakeholders, Task 3.4 will draft a report on the feasibility of standardization that will complement this work. As much standardization is already taking place, it will be important to link ongoing efforts. This will also provide greater assurance that the results will be sustainably embedded in formal standardization. Standards are a tool for self-regulation but can also be used by the regulator. Standards are developed to guide and/or facilitate the implementation of legal principles (e.g., ISO 13485:2016 Medical devices — Quality management systems — Requirements for regulatory purposes; ISO 27000 series that helps with GDPR compliance; ISO 26000 for compliance with social responsibility). Legislation can also mandate the use of standards and/or certification mechanisms.

SHERPA Task 3.5 looks at technical options and interventions, which again have a deep connection with the law. Task 3.6 (terms of reference for a new regulator) will also consider and draw from the work done this report.

2. Methodology and scope

Our research focussed on general regulatory options for AI and/or big data, especially those that include ethics, security and human rights (in line with the focus of the SHERPA project). We have noted more specific issue-based literature (e.g., covering a single topic in detail). We have focussed to a lesser extent on technical regulation (e.g., regtech), ethical guidelines or standards (only including these if highly relevant), as this is within the scope of other SHERPA tasks, i.e., Task 3.2, Task 3.4., Task 3.5. The search terms used in the research underlying this report included: ***regulation/law/human rights + AI/big data analytics/smart information systems*** (period covered: unless otherwise specified, 2009-2019). We primarily focussed on developments at the European Union level and in the Member States; but as AI and big data analytics have global dimensions we have also taken a wider approach (especially in the options study) to make this analysis more internationally relevant.

The methodology and scope of Section 3 (*Policy positions and perspectives on regulation of AI and big data)* is outlined in that section.

Identification and analysis of regulatory options for AI and big data analytics

Based on SHERPA prior work, the SIENNA legal analysis[14], and our research in this task (e.g., via requests for information, personal interviews), we first identified some regulatory options (proposals for law, bodies and other mechanisms) for regulating AI and big data. The team prepared a list of criteria for assessment of the regulatory options, which were internally discussed and revised. The criteria were drawn in part from the European Union Better Regulation objectives and Toolbox[15] questions, regulatory impact assessment templates, and other research analysing regulatory options (e.g., Clarke's "Regulatory Alternatives for AI" 2019)[16]. The criteria were tested in July 2019 against two options, the proposal for the Digital Authority, and the Council of Europe proposal for human rights impact assessment law, and refined via internal discussions. The preliminary options and criteria for examination of individual options were shared with the SHERPA Advisory Board and the project policy officer in the form of a scoping paper (see Annex 2) in September 2019 for feedback. It was also shared via the SHERPA website, social media and newsletter, and was finalised in September 2019. We received 12 responses in total, from the SHERPA Advisory Board and others and the criteria to be used in assessment were refined further; we also added in new options for study. We then analysed each option using desktop research, supplemented by requests for information from stakeholders connected with the options, legal academics, policymakers and industry experts where feasible, using the following criteria.

Criteria/touch points
1. Outline its relevance/connection to AI and big data analytics: What does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures? Give an application example.
2. What is its basis (on which the regulatory option is created - law? if yes, which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?
3. Purpose/objective/what need does the option fulfil?
4. What gap does it address?
5. What added value does it have?

[14] SIENNA, *D4.2: Analysis of the legal and human rights requirements for AI and robotics in and outside the EU*, 2019.
[15] https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-1_en_0.pdf
[16] http://www.rogerclarke.com/EC/RAI.html

Criteria/touch points
6. What are the limitations, risks and challenges? *(Internal note: look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why?
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. citizens, b. public administrations, c. Businesses, and particularly SMEs.
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]
11. Whose rights and/or interests does this option neglect?
12. Does it explicitly support or adversely affect human rights? If yes, which ones? If not, how might it boost human rights?
13. How does it address ethics and ethical principles? Which ones?
14. Does it explicitly consider gender dimensions? How (e.g., in the composition of the agency/body, consideration of gender equality, gender neutrality)?
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.
16. What provisions are there for regular review and update?
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments (e.g., technological, policy changes, social demands)?
18. Will it adversely impact the ability for businesses and others to innovate? *[If yes elaborate]*
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis).
20. Any other implementation challenges (especially those not covered above e.g., complexities)?
21. Based on this study, how likely is this option to succeed? (1 – Extremely unlikely, 2 – Unlikely 3 – Neutral, 4 – Likely 5 – Extremely likely)?
22. Overall conclusion (what are the factors critical to its adoption and/or success?).

Table 3: Finalised list of criteria used in analysis of regulatory options

Some of the key challenges in preparing the report included:

(1) *Identifying all the relevant options* - it was not possible to be comprehensive given the limited remit of the task and the breadth of the topic, but we addressed this challenge by issuing the scoping paper after the preliminary research on identifying options and getting feedback to ensure we had covered the significant ones.

(2) *The breadth of the topic and parallel work on the subject* - this report was not intended to be a treatise on the legal regulation of AI and big data. Instead, it complements the policy and legal discussions on the subject with snapshots of stakeholder positions, and detailed options analysis, along with critical insights into the key considerations.

(3) *How we understand 'regulation' and what this means for AI and big data*. There are different conceptualisations of 'regulation'[17]. Given other SHERPA work on technical, standardisation and

[17] See e.g., Black J., "Critical Reflections on Regulation", 27 *Australian Journal of Legal Philosophy* (2002) 1; Brownsword R. & M. Goodwin, *Law in Context: Law and the Technologies of the Twenty-First Century: Text and Materials,* Cambridge University Press, 2012.

ethical codes/guidelines, this report focussed widely on proposals for legislation, regulatory bodies and other regulation-supporting mechanisms.

3. Policy positions and perspectives on regulation of AI and big data

This section presents an overview of current positions and perspectives on the regulation of AI and big data. The overview identifies key policy bodies/organisations active in determining, facilitating or promoting AI/big data regulation; the types of legislation, regulations, and regulatory options they are advocating for, in what sector, and why. The overview focuses on ethical considerations, especially human rights, and aims to describe whether there is a particular type of regulatory approach/position being advocated/favoured/strongly pushed (e.g., risk-based, principles-based, self-regulation, co-regulation, full regulation, harmonised, light-touch, heavy-touch, general, sector specific, combinations).

3.1 International policy level

At the international level, the key organisations active in determining, facilitating or promoting AI/big data regulation include: the United Nations (UN), Council of Europe (CoE), and the Organisation for Economic Co-operation and Development (OECD).[18] The output from these organisations illustrates a joint understanding that regulations on AI technology must be harmonised,[19] and there is a shared ideal within the available proposals concerning the fundamental principles, including, for example, human rights.

Within the UN, the International Telecommunication Union (ITU) aims to provide "a neutral platform for government, industry and academia to build a common understanding of the capabilities of emerging AI technologies and consequent needs for technical standardization and policy guidance".[20] The UN has also created the Centre for Artificial Intelligence and Robotics.[21] The UN strives to influence AI regulation through international cooperation and dialogue.[22] The UN also related AI to its *Sustainable Development Goals* (SDGs), by identifying AI as a tool for social good and its central role in achieving the SDGs.[23] However, there are also discussions on regulations for harmful AI technology (e.g., killer robots).[24]

---

[18] There are various other organizations that affect AI regulation, but in this overview we have focused on work that is AI specific, setting aside the fact that, for example, human rights agreements set limits for AI-technology. See, e.g., SIENNA D4.2, for an overview of human rights declarations and agreements, and how they map on to AI-relevant principles.

[19] As pointed out by one of our stakeholder board members, in general for internationally operating companies, regulating at the highest, international level might be preferred over more local regulation. Abiding by a single internationally accepted set of rules would have a lower regulatory burden than different sets of rules for different geographies or countries or provinces.

[20] ITU, "Artificial Intelligence". https://www.itu.int/en/ITU-T/AI/Pages/default.aspx

[21] UNICRI, "UNICRI Centre for Artificial Intelligence and Robotics". http://www.unicri.it/in_focus/on/UNICRI_Centre_Artificial_Robotics

[22] E.g., ITU, "AI for Good Global Summit". https://aiforgood.itu.int/

[23] ITU, "Artificial Intelligence", op. cit.; and UN, *Sustainable Development Goals*. https://sustainabledevelopment.un.org/?menu=1300

[24] E.g., Motoyama, S., "Inside the United Nations' Effort To Regulate Autonomous Killer Robots: Meet the UN diplomat heading up the coming 'killer robot' conference", *The Verge*, 27 August 2018. https://www.theverge.com/2018/8/27/17786080/united-nations-un-autonomous-killer-robots-regulation-conference; and Gayle, D. "UK, US and Russia among those opposing killer robot ban" *The Guardian*, 29 March 2019. https://www.theguardian.com/science/2019/mar/29/uk-us-russia-opposing-killer-robot-ban-un-ai

In August 2018, the *Report of the Special Rapporteur to the General Assembly on AI and its impact on freedom of opinion and expression* recommended sector-based regulations because comprehensive legalization of AI may lead to "lack of detail with overly restrictive or overly permissive provisions". The report focuses on the priority of states (and companies) to "ensure that AI is developed in keeping with human rights standards", which implies a principle-based approach.[25]

In May 2019, the OECD adopted the *Principles on Artificial Intelligence*[26], with two sections of recommendations. The first is a flexible set of "Principles for responsible stewardship of trustworthy AI", specifically: "*i)* inclusive growth, sustainable development and well-being; *ii)* human-centred values and fairness; *iii)* transparency and explainability; *iv)* robustness, security and safety; and *v)* accountability". The second section, "National policies and international co-operation for trustworthy AI", recommends adherence to the aforementioned principles alongside investments in a "digital ecosystem", "policy environment", preparation for "labour market transformation", and "international co-operation for trustworthy AI".[27] The latter part, as well as the collaboration in itself, indicates a desire, as well as an international appreciation of the need, for harmonised regulations. The guidelines promote a combination of a risk-based (with a focus on safety and risk management)[28] and a principle-based (with a focus on human rights) approach.[29] Guided by the OECD recommendations, the G20 adopted principles for human-centered AI in June 2019.[30] The *G20 Ministerial Statement on Trade and Digital Economy*[31] outlines that "Policies, regulations, or the removal of regulatory barriers can contribute to and accelerate economic growth, and inclusive development by developing countries as well as MSMEs". [32] They state that, "governments should promote a policy environment that supports an agile transition from the research and development stage to the deployment and operation stage for trustworthy AI systems. To this effect, they should consider using experimentation to provide a controlled environment in which AI systems can be tested, and scaled-up, as appropriate."[33] They also advise that, "Governments should review and adapt, as appropriate, their policy and regulatory frameworks and assessment mechanisms as they apply to AI systems to encourage innovation and competition for trustworthy AI." [34]

Simultaneously, the Council of Europe released *Unboxing Artificial Intelligence: 10 steps to protect Human Rights* (UAI) based on a large set of previous reports by the CoE.[35] UAI includes both recommendations and obligations to:

---

[25] UN, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. https://undocs.org/A/73/348; brief summary available at https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx

[26] OECD, *Recommendation of the Council on Artificial Intelligence*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[27] OECD, *Recommendation of the Council on Artificial Intelligence*, op. cit.

[28] OECD, *Recommendation of the Council on Artificial Intelligence*, e.g., 1.4 Robustness, security and safety.

[29] OECD, *Recommendation of the Council on Artificial Intelligence*, e.g., 1.2 Human-centred values and fairness.

[30] OECD, *What are the OECD Principles on AI?*. https://www.oecd.org/going-digital/ai/principles/; and G20, *G20 Ministerial Statement on Trade and Digital Economy*, 8-9 June 2019. https://www.mofa.go.jp/files/000486596.pdf

[31] G20, op. cit. 2019

[32] G20, op. cit. 2019

[33] G20, op. cit. 2019

[34] G20, op. cit. 2019

[35] "[t]he European Ethical Charter on the use of artificial intelligence in judicial systems, the Guidelines on Artificial Intelligence and Data Protection, the Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes and the Study on the human rights dimensions of automated data processing techniques and possible regulatory implications" (The Council of Europe Commissioner for Human Rights *Unboxing Artificial Intelligence: 10 steps to protect Human Rights*, p. 6. https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64)

- regularly "carry out human rights impact assessments (HRIAs)" (i.e., before and after procurement), on AI systems used by public authorities, including a requirement on self-assessment and external reviews of both the system and how it is used[36];
- open procurement processes[37];
- implementation of the *UN Guiding Principles on Business and Human Rights and the Recommendation* (CM/Rec(2016)3), and other measures necessary to protect human rights (ECHR) relative to actions by all AI actors in the private sector[38];
- requirement on transparency and information of AI usage in public service[39];
- independent oversight, with "the power to intervene in circumstances where they identify (a risk of) human rights violations occurring", and the requirement for both the public and private sector to provide all information necessary for oversight[40];
- absolute prevention and mitigation of discrimination risks[41];
- implementation of "the modernised *Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data* ('Convention 108+')", and that all usage (including development) of AI must "fully secure a person's right to privacy"[42];
- protection of freedom of expression, assembly and association, and a right to work (implying the need for diversity of ideas, sector-based requirements, and plans for, e.g., reschooling of workers)[43];
- human control of the system, responsibility and accountability assigned to a natural or legal person, and remedies for human rights violations[44]; and the promotion of AI literacy[45].

The guidelines end with a checklist of "Do's" and "Don'ts", illustrating the strict heavy-touch nature of the requirements.[46]

3.2 EU policy level

In 2017, the EU Parliament called on the European Commission to, amongst other things, create an EU Agency for Robotics and Artificial Intelligence.[47] The EU Commission did not consider the creation of a new agency necessary, but instead proposed the creation of "a high-level advisory body on robotics and artificial intelligence which could provide knowledge and expertise to the Commission". This became the High-Level Expert Group on Artificial Intelligence ("AI HLEG").[48]

In 2019, the AI HLEG released *Ethics Guidelines for Trustworthy AI*, which calls for an appropriate governance and regulatory framework. By "appropriate", the AI HLEG means a framework that

[36] The Council of Europe Commissioner for Human Rights, op. cit., pp. 7-8.
[37] The Council of Europe Commissioner for Human Rights, op. cit., pp. 8-9.
[38] The Council of Europe Commissioner for Human Rights, op. cit., p. 9.
[39] The Council of Europe Commissioner for Human Rights, op. cit., pp. 9-10.
[40] The Council of Europe Commissioner for Human Rights, op. cit., pp. 10-11.
[41] The Council of Europe Commissioner for Human Rights, op. cit., p. 11.
[42] The Council of Europe Commissioner for Human Rights, op. cit., pp. 11-12.
[43] The Council of Europe Commissioner for Human Rights, op. cit., pp. 12-13.
[44] The Council of Europe Commissioner for Human Rights, op. cit., pp. 13-14.
[45] The Council of Europe Commissioner for Human Rights, op. cit., pp. 14-15.
[46] The Council of Europe Commissioner for Human Rights, op. cit., pp. 17-23.
[47] European Parliament, *Civil Law Rules on Robotics*, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.pdf
[48] European Parliament, Follow up to the European Parliament resolution of 16 February 2017 on civil law rules on robotics. http://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2017/11-20/A8-0005-2017_EN.pdf

promotes socially valuable AI development and deployment, ensures and respects fundamental rights, the rule of law and democracy, while safeguarding individuals and society from unacceptable harm.[49]

The AI HLEG guidelines are built on three components which all require that a system should be: lawful ("complying with all applicable laws and regulations"), ethical ("ensuring adherence to ethical principles and values"), and robust (from both a technical and social perspective).[50] However, these principles can be in conflict (e.g., the law can sometimes require unethical actions, and vice versa).

The AI HLEG guidelines propose a risk-based approach to regulation. For AI applications that generate "unacceptable" risks or pose threats of harm that are substantial, a precautionary principle-based approach should be adopted instead.[51]

Beyond an ethical and legal framework, the European Commission expects to increase investments (private and public) and make preparations for AI's socio-economic effects. Furthermore, the European Parliament has called for the EC to make an assessment of AI's impact. In February 2019, Parliament also adopted its own report, *A Comprehensive European industrial policy on artificial intelligence and robotics*.[52]

3.3 National policy level

Below, we summarise some of the most recent trends (2017-2019) in national regulations for AI in 5 selected countries: Australia, China, Germany, Mexico, and the USA. While we wanted to select a major economy in each inhabited continent, at the time of writing (August 2019), no African country has yet finalized a national AI policy, nor has any South American country (hence the inclusion of Mexico and the USA).[53] In summary, while there are a lot of overlaps between the selected countries, there is a large variation in how detailed the specific policy and regulatory proposals are, varying from an overall need for modifications to very specific proposals. All surveyed nations aim to adhere to ethical principles, but only some give specific principles for ethical AI.

Australia

In April 2019, The Department of Industry, Innovation and Science released a report prepared by the Commonwealth Scientific and Industrial Research Organisation.[54] The report sets out eight core principles for AI: 1) generate net-benefits; 2) "Do no harm", do not deceive, and minimise negative outcomes; 3) "Regulatory and legal compliance"; 4) ensure privacy protections; 5) fairness (i.e., non-discrimination), including non-bias in training data; 6) "Transparency & Explainability" (individuals

[49] High-Level Expert Group on AI, *Ethics Guidelines for Trustworthy AI*, European Commission, 2018. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

[50] High-Level Expert Group on AI, op. cit., p. 2.

[51] High-Level Expert Group on AI, op. cit., p. 37.

[52] European Parliament, *Legislative Train Schedule: Connected Digital Single Market*, 2018, "https://www.europarl.europa.eu/legislative-train/theme-connected-digital-single-market/file-artificial-intelligence-for-europe

[53] See, e.g., The Law Library of Congress, Global Legal Research Directorate, "Regulation of Artificial Intelligence in Selected Jurisdictions", January 2019. https://www.loc.gov/law/help/artificial-intelligence/regulation-artificial-intelligence.pdf, pp. 119-132; "Kenya Govt unveils 11 Member Blockchain & AI Taskforce headed by Bitange Ndemo", *The Kenyan Wall Street*, February 28, 2018. https://kenyanwallstreet.com/kenya-govt-unveils-11-member-blockchain-ai-taskforce-headed-by-bitange-ndemo/; Besaw, C. & J. Filitz, "AI & Global Governance: AI in Africa is a Double-Edged Sword", United Nations University Centre for Policy Research. https://cpr.unu.edu/ai-in-africa-is-a-double-edged-sword.html; and, Artificial Intelligence for Development, "A roadmap for artificial intelligence for development in Africa", 8 May 2019. https://ai4d.ai/blog-africa-roadmap/.

[54] Dawson, D, E Schleiger, J Horton, J McLaughlin, C Robinson, G Quezada, J Scowcroft and S Hajkowicz, *Artificial Intelligence: Australia's Ethics Framework*. Data61 CSIRO, Australia, 2019.

must be informed when an algorithm impacts them and about what information is used in decision-making); 7) "Contestability" (i.e., a process that allows individuals "to challenge the use or output of the algorithm"; and 8) "Accountability" ("People and organisations responsible for the creation and implementation of AI algorithms should be identifiable and accountable for the impacts of that algorithm, even if the impacts are unintended").[55]

China

China set its overall AI plan in 2017.[56] It has been complemented by a three-year plan for 2018-2020.[57] The 2017 plan starts with a description of the strategic situations in which, for example, AI is seen as "a new engine of economic development"[58] and a technology that "is indispensable for the effective maintenance of social stability."[59] The plan lists four basic principles: technological leadership; systematic layouts (which "give full play to the advantages of the socialist system"); market-orientated (including market and ethical regulations); and open source.[60] The plan then lists strategic goals at three five-year marks (2020, 2025, and 2030). This includes (as a sub-goal at the first mark) the focus to: "establish initially artificial intelligence ethics norms, policies and regulations of some areas."[61] This is followed (at the next mark) by: "We shall make initial establishment of artificial intelligence laws and regulations, ethical norms and policy systems, and form artificial intelligence safety assessment and control capabilities."[62] And at the final mark: "We shall form a number of the world's leading artificial intelligence technology innovation and personnel training bases, and create comprehensive laws and regulations, ethics and policy system of artificial intelligence."[63]

The rest of the plan includes a combination of sector-specific concerns and, for example, assurance measures including: "Develop laws and regulations and ethical norms that promote the development of AI"[64]; "Improve the key policies that support AI development"[65]; "Establish standards and the intellectual property system for AI technology"[66]; "Establish safety supervision and evaluation systems for AI"[67]; "Vigorously strengthen training for the labour force working in AI"[68]; and "Carry out a wide range of AI science activities"[69]. Finally, the implementation section aims to be guided by public opinion and to inform the public.[70] The three-year plan mentions regulatory aims once, under the assurance measure of development environment optimization: "Carry out research on relevant

[55] Dawson et al, op. cit., p. 6.
[56] *New Generation of Artificial Intelligence Development Plan*, State Council Document No. 35, translation available at: https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf
[57] *Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry (2018–2020)*, translation available at: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-government-outlines-ai-ambitions-through-2020/
[58] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 2.
[59] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 2.
[60] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 4.
[61] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 5.
[62] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 6.
[63] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 7.
[64] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 25.
[65] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 25.
[66] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 26.
[67] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 26.
[68] *New Generation of Artificial Intelligence Development Plan*, op. cit., pp. 26-27.
[69] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 27.
[70] *New Generation of Artificial Intelligence Development Plan*, op. cit., p. 28.

policies, laws and regulations of AI and create a good environment for the healthy development of the industry."[71]

Germany

Germany released its *Artificial Intelligence Strategy* in November 2018.[72] The strategy consists of a large set of proposals ranging from different forms of investments to ethical and regulatory suggestions. The strategy starts with three goals and fourteen sub-goals:

1) "make Germany and Europe a leading centre for AI"[73];
2) "a responsible development and use of AI which serves the good of society[74]; and
3) "integrate AI in society in ethical, legal, cultural and institutional terms in the context of a broad societal dialogue and active political measures".[75]

The strategy also includes a set of fields of actions, which reveals the desire for harmonised regulations, for example by establishing a "German observatory for artificial intelligence and […] support the establishment of similar observatories at European and international level"[76], through "initiating a European and transatlantic dialogue on the human-centric use of AI in the world of work"[77], and through engaging in dialogue to "if possible reach agreement on joint guidelines with other leading regions"[78].

Germany will also, for example, "examine ways of auditing AI for use in companies"; "promote research regarding explainability and accountability"[79]; "promote research and development" to protect privacy[80]; harness opportunities in biotech, food production, and develop AI for benefits in

[71] *Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry (2018–2020)*, op. cit.
[72] The Federal Government. *Artificial Intelligence Strategy*, November 2018. https://www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie_engl.pdf
[73] The Federal Government, op. cit., p. 8. This includes investments in research, national collaborations, and to "establish the right framework conditions to create value from applications of AI in Germany, and to focus our efforts on developing the benefits of AI for our citizens – both at an individual and at societal level", and goals to benefit society, "to become new top exports", "strictly observing data security and people's right to control their personal data", and protection "from manipulation and misuse and to prevent risks to public security in the best way possible".
[74] The Federal Government, op. cit., p. 9. This includes "responsible use of AI", "adhering to ethical and legal principles consistent with our liberal democratic constitutional system throughout the process of developing and using AI", taking into account "recommendations of the Data Ethics Commission", "a European solution for data-based business models", and "to raise awareness on the part of the relevant stakeholders" "regarding ethical and legal limits of the use of artificial intelligence and to examine whether the regulatory framework needs to be further developed in order for it to guarantee a high level of legal certainty", and "to demand and foster compliance with ethical and legal principles throughout the process of developing and using AI".
[75] The Federal Government, op. cit., p. 9. This includes ensuring that AI is "people-focused, especially with regard to the use of AI in the world of work", "enabling self-determination, providing security and protecting health", "representing diversity", improving the ability to participate in working life for people with disabilities, and "to improve security, efficiency and sustainability", "whilst also promoting social and cultural participation, freedom of action and self-determination", and "utilise the potential of AI for sustainable development" to achieve Agenda 2030, and "to create a policy environment for AI applications that creates and maintains diversity and guarantees the necessary scope for the development of cultural and media freedoms."
[76] The Federal Government, op. cit., p. 26
[77] The Federal Government, op. cit., p. 26.
[78] The Federal Government, op. cit., p. 41.
[79] The Federal Government, op. cit., p. 16.
[80] The Federal Government, op. cit., p. 16.

other areas (such as healthcare, environment, and climate)[81]; develop technologies for civil security; "improve resilience […] against attacks"[82]; improve people's AI skills[83]; "review the legal framework" and "assess how AI systems can be made transparent, predictable, and verifiable"[84]; set standards[85]; "set up a communication strategy for AI", and "support dialogue between social partners on sustainable integration of AI into the world of work."[86]

Mexico

In June 2018, The British Embassy in Mexico, Oxford Insights, and C Minds, released the White Paper, *Towards An AI Strategy in Mexico: Harnessing the AI Revolution*[87], in collaboration with the Mexican Government, which has since adopted it as an official strategy.[88] The strategy ranges over various topics, and the section on ethics and regulations includes two main goals, 1) "Bring data assets inside the scope of competition law (COFECE)", and 2) "Create a Mexican AI Ethics Council (current and next administration)".[89] The latter also includes two sub-goals: "Set guidelines and limits which reflect Mexican values", and "Award a quality mark for AI companies who abide by the standards."[90] The strategy also includes ethics considerations in other sections, such as "Data infrastructure ("Maintain a resilient open data infrastructure"; "Create Mexican training data to inform Applications (next administration)"; and "Protect personal privacy (next administration, INAI)").[91] The strategy also sets up goals (or recommendations) for example related to investments, collaborations, and education.[92]

USA

In February 2019, an executive order set out goals for AI in the USA.[93] This has since resulted in *The national artificial intelligence research and development strategic plan: 2019 update* in June 2019. The report sets out the USA's *National AI R&D Strategic Plan*, including eight priorities/strategies, setting out goals for AI research investment; human-AI collaboration; understanding and addressing "the ethical, legal, and societal implications of AI"; ensure safety and security; "shared public datasets and environments for AI training and testing"; measurements and evaluation of AI through standards and benchmarks; "strategically foster an AI-ready workforce"; and "Expand public-private partnerships".[94] The executive order has also resulted in *A Plan for Federal Engagement in Developing Technical Standards and Related Tools*.[95]

---

[81] The Federal Government, op. cit., pp. 16-20.
[82] The Federal Government, op. cit., pp. 17-18.
[83] The Federal Government, op. cit., pp. 29-30.
[84] The Federal Government, op. cit., pp. 37-38.
[85] The Federal Government, op. cit., pp. 39-40.
[86] The Federal Government, op. cit., p. 45.
[87] Martinho-Truswell, Emma, Hannah Miller, Isak Nti Asare, André Petheram, Richard Stirling, Constanza Gómez Mont, and Cristina Martinez, " Towards an AI strategy in Mexico: Harnessing the AI Revolution", 2018. https://docs.wixstatic.com/ugd/7be025_e726c582191c49d2b8b6517a590151f6.pdf
[88] Gobierno de México, *Estrategia de Inteligencia Artificial MX 2018*. https://www.gob.mx/mexicodigital/articulos/estrategia-de-inteligencia-artificial-mx-2018
[89] Martinho-Truswell et al, 2018. op. cit., p. 37.
[90] Martinho-Truswell et al, 2018. op. cit., p. 37.
[91] Martinho-Truswell et al, 2018. op. cit., p. 36.
[92] Martinho-Truswell et al, 2018. op. cit., pp. 32-35.
[93] The President, *Maintaining American Leadership in Artificial Intelligence*. Executive Order 13859 of 11 February 2019. Federal Register 84(31), pp. 3967-3972. https://www.govinfo.gov/content/pkg/FR-2019-02-14/pdf/2019-02544.pdf
[94] The Select Committee On Artificial Intelligence Of The National Science & Technology Council, *The National Artificial Intelligence Research And Development Strategic Plan: 2019 Update*. A Report, 2019. https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf, p. iii
[95] NIST, "U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools, 2019." https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf

3.4 General analysis and conclusion

In summary, although there are disagreements, important actors seem to **aim for harmonized rules** (as indicated, e.g., by the OECD guidelines and the major companies behind Partnership on AI). Most proposals **advocate a heavier rather than lighter touch**, but there are clear disagreements. Proposals often **combine risk-based approaches with principle-based regulation** (perhaps because, e.g., safety concerns are often risk-related, while some human rights require a principle-based approach). There is an understanding among industrial proponents that **regulations are needed, but there is disagreement and ambiguity** as to whether this should be self-regulation, co-regulation, or full regulation. Unsurprisingly, the most common worry is that **a heavy-touch will restrict innovation**, while a **light-touch will leave individuals and society exposed to risks to fundamental values or human rights.** The challenge for any regulation is how to promote good AI development and use, and how to minimize the creation of bad AI or misuse of AI-technology and increase its security (reliability and resilience).

Proposals for regulations almost always address ethical concerns and human rights. However, there is great **variation as to the specificity** of such proposals. Some broadly note that guidelines should be ethically appropriate or aligned with human rights. While this might imply that technically there are few regulatory gaps, it is problematic to promote regulations without clearly identifying how potential ethical and human rights problems can be dealt with. Such broad formulations are simply too unspecified to provide guidance as to how ethical issues, trade-offs, or human rights should be handled or balanced in practice. This is either an indication that the process is not fully developed to address ethical concerns in regulations, or that ethical considerations are side-lined for other priorities. Still, some proposals address ethical concerns with a fairly detailed recognition of some of the challenges we are facing. Of course, it is also important to recognize that some challenges require prioritization, in the sense that different values or goals sometimes directly or indirectly contradict each other[96]. Indeed, whether the more substantive proposals are fully coherent or imply regulatory conflicts requires a more detailed analysis.

[96] For example, how to achieve transparency and protect privacy.

4. Regulatory options

There are various approaches to AI and big data regulation. Self-regulatory approaches might include, for example, the use of industry standards[97] and certifications (e.g., relating to safety/security of AI systems), the use of codes of ethical practice or guidelines, or host take-down of illegal or harmful content. Co-regulatory approaches include, for example, the application of a government-approved industry code monitored by an independent body with the threat of sanctions, or a formal self-regulator recognised by regional/national institutions. Full regulatory approaches include legislation backed by enforcement by an independent regulator, or national legislation backed by enforcement by a regulatory body, e.g., the *General Data Protection Regulation* (GDPR), or Italian *Smart Road Decree no. 90*. A hybrid example that might include a mix of approaches or not cleanly fit into a single approach might be an independent European body, which contributes to the consistent application of rules throughout a region and promotes cooperation between the region's authorities, e.g., the European Data Protection Board (EDPB).

This section looks at various regulatory options - limited to proposals for laws, regulatory bodies and other regulatory tools for AI and big data. This analysis takes place within the framework of national, European and/or international law. For instance, the fact that in the EU many laws are harmonised and/or approximated in the name of the principle of free movement, gives rise to potentially unique regulatory options, including in the field of AI and data.

Section 4.1 further outlines the scope and methodology used and presents the list of regulatory options that were studied. Section 4.2 presents the analysis of the options studied. Section 4.3 discusses how AI challenges regulation and EU aspirations for better law-making.

4.1 Identification of options

Scope and methodology

The study identified a number of regulatory options (proposals for laws, bodies and other regulatory tools and mechanisms*)* for analysis. This was based on a desktop review, which included a scan of international and European policy documents and academic literature (during July-August 2019), and the scoping paper feedback from the SHERPA Stakeholder Advisory Board (in September 2019). The options were identified based on, (a) their regulatory nature, (b) their connection to SIS (AI and/or big data), (c) their potential connection to ethics and human rights, (d) their active discussion at different levels. We looked at the work of international, EU and national bodies active in the AI/big data policy and regulatory space and supplemented this with research from SIENNA[98] and SHERPA (work on case studies, scenarios and human rights challenges, and section 4 of this report). These options have been proposed by a variety of stakeholders: elected officials, policymakers, regulators, the research community, civil society, projects active in the area (e.g., SIENNA and SHERPA) based on reviews and analysis of legal issues and/or human rights challenges of AI and big data. The initial list of options

[97] E.g., IEEE P2802 - *Standard for the Performance and Safety Evaluation of Artificial Intelligence Based Medical Device: Terminology*. https://standards.ieee.org/project/2802.html; IEEE, P7006 - *Standard for Personal Data Artificial Intelligence (AI) Agent* (2018). https://standards.ieee.org/project/7006.html

[98] SIENNA, *D4.2: Analysis of the legal and human rights requirements for AI and robotics in and outside the EU*, 2019.

identified changed based on the scoping paper feedback, where some were suggested to be no-go options, and in some cases there was no data found to analyse.

As the area of AI and big data forms a very volatile cross-disciplinary combination of innovation, policies and technologies, multiple initiatives surface worldwide, not always accompanied by sufficient substance and processes. Thus, it was deemed prudent to adopt a wide understanding of inclusive 'regulatory options' to cover **proposals for laws, bodies** and other **regulatory tools and mechanisms**. We considered it necessary to study a spectrum of proposals and not adopt a too restrictive approach for two reasons:

1. There is **no silver bullet approach** to regulating AI and big data.
2. As pointed out by one of the SHERPA Advisory Board Members in their feedback to our scoping paper, given the specific effects of AI-based algorithms, it is possible that classical legal systems are unable to address their effects (citing the example of liability law).

Well-established legislation has been excluded from the scope of this study, as there is a wealth of serious research and analysis which this study does not seek to duplicate. What is now important given the pushes and pulls towards regulating AI and big data is to consider how new (or relatively new and under-examined) proposals would positively or adversely affect the regulatory future of AI and big data, and impact on stakeholders. We recognise the regulatory options for AI and big data as a constant 'moving target' and do not claim to be comprehensive in this report, also given the study had a limited duration.

The Figure below depicts the options studied. **31 options** were individually analysed, **8 at the international level, 9 at the EU-level, 11 national and 3 cross-overs**. The full list is presented in Annex 3 with individual assessments presented in Annex 4 of this report.

International	EU-level	National	Cross-over
1.Moratorium on LARs/LAWS 2.Binding Framework Convention for AI 3.Legislative framework for independent and effective oversight 4.Legal for human rights impact assessments (HRIAs) on AI systems 5.Convention on human rights in the robot age 6.CEPEJ European Ethical Charter 7.International Artificial Intelligence Organization 8.Global legal AI and/or robotics observatory (SIENNA)	1.EU-level special list of robot rights 2.Adoption of common Union definitions 3.Creating electronic personhood status for autonomous systems 4.Establishment of a comprehensive Union system of registration of advanced robots 5.General fund for all smart autonomous robots 6.Mandatory consumer protection impact assessment 7.EU Taskforce of field specific regulators for AI/big data 8.Algorithmic Impact Assessments under the GDPR 9.Voluntary/mandatory certification of algorithmic decision systems (ADS)	1.DEEP FAKES Accountability Act 2.Algorithmic Accountability Act 3.Directive on Automated Decision-Making 4.US Food and Drug Administration regulation of adaptive AI/ML technology 5.New statutory duty of care for online harms 6.Redress by design mechanisms for AI 7.Register of algorithms used in government 8.Digital Authority 9.Independent cross-sector advisory body (CDEI) 10.FDA for algorithms 11.US Federal Trade Commission to regulate robotics	1.Using anti-trust regulations to break up big tech and appoint regulators 2.Three-level obligatory impact assessments for new technologies 3.Regulatory sandboxes

Fig 1: Identified regulatory options (proposals for laws, bodies and other mechanisms)

The above list does not include the following options from our preliminary search that were not included for further analysis:

- **Canada-France: Global Partnership on Artificial Intelligence (GPAI)** proposed by the governments of Canada and France: no proposal has been developed on this to date. It was replaced by a study of the *Directive on Automated Decision-Making* (Canada), as this might be of value to both EU and national regulators.
- **Super-regulator:** feedback from the Advisory Board suggested that this might manifest in the forms of an EU or national Task Force; we examined the proposal for a task force of field-specific regulators at the EU-level.
- **Estonian AI liability law:** the proposal did not go forward as Estonia decided it would like to build on the EU framework.[99]
- **New rules governing the free flow of non-personal data in the Union** (European Parliament) - Regulation (EU) 2018/1807) applicable as of 28 May 2019, is intended

[99] https://digi.geenius.ee/rubriik/uudis/eesti-riigi-it-juht-siim-sikkut-kui-riigieelarves-oleks-teadusele-1-saaksime-kratindusest-kindlamalt-raakida/

to remove obstacles to the free flow of data within the European Union (e.g., to remove data localization requirements for non-personal data) by, among other things, clarifying rules regarding processing of mixed datasets (containing both personal and non-personal data). Because this option has already been adopted into law applicable to all EU member states, it is outside the scope of the regulatory option assessment here.

- **European Agency for Robotics and Artificial Intelligence** proposed by the European Parliament - this option identified in our preliminary research was confirmed by scoping paper feedback as a no-go option, also having been rejected by the European Commission as it did not "consider it necessary to designate a new European Agency for robotics and artificial intelligence"[100]; the team briefly considered it but found it was not viable for analysis as lacking in depth.

The studied options can be classified functionally:

- **General rights/responsibilities** (e.g., electronic personhood, special rights for robots)
- **Regulation of specific portions/parts** (e.g., algorithmic accountability, automated decision-making)
- **Specific field/specific application regulation** (e.g., AI medical devices, online services)
- **Prohibition/restriction of activity** (e.g., moratorium, unfair/deceptive trade practices, anti-trust laws)
- **Mechanisms/tools for regulators** (e.g., regulatory sandboxes, registration of robots, common definitions, HRIAs)
- **Redress mechanisms** (e.g., redress by design, compensation fund)

A functionally oriented classification has the merit of combining different legal and policy dimensions transnationally in a highly global field, outlining common trends, shared interests and approaches.

4.2 Assessment of options: analysis of findings and results

The study team analysed the identified options in detail. **E**ach individual option was assessed during October to November 2019 using the criteria outlined in Methodology Section 2. For individual results, see Annex 4.

The study of the individual options faced a number of challenges. First, as pointed out above, we had to abandon or limit the study) of a few options due to complete lack of information/development of the proposals (e.g., a Convention on human rights in the robot age. In some cases, given that the options are of great interest to policymakers and are the subject of debate, we surmounted this, by doing some creative thinking and suggesting how the proposals might work and what their impacts might be (e.g., mandatory consumer protection impact assessment; robot rights proposal). In some cases we replaced them with more developed options. We attempted to contact stakeholders/experts (including proposers) to get their insights into some of the options, but in some cases they did not respond.

The sections below present the collated results of the assessment. For detailed or further information on each option, readers should consult Annex 4.

[100] See https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

4.2.1 Relevance/connection to AI and big data

Relevance/connection to AI and big data analytics

All the reviewed proposals refer to AI. In terms of the relevance and proximity of the proposals to AI, almost all the proposals also directly and explicitly relate to AI. The aims and objectives of the vast majority of these proposals are to directly regulate AI and suggest governance mechanisms for this technology. Solutions to the risks of AI are suggested, either as autonomous and independent, or as supporting measures to the existing legal and technological *status quo*.

There is a small number of proposals whose first aim is to regulate other relevant technologies, areas or legal fields and, in this context, they touch upon the regulation of AI. For example, the mandatory consumer protection impact assessment suggested by the AI HLEG, or the anti-trust regulations recommended by the US Senator Elizabeth Warren, aim to restore the balance in consumer protection and competition law respectively, where AI threatens to deregulate these areas. There is only one exception to the high degree of relevance of the revised proposals to AI. In particular, the examined duty of care for online harms by the UK Government discusses the use of AI as a tool of the concerned entities to monitor harmful user-generated content.

In terms of the specificity of the essence of the suggested measures, six[101] reviewed proposals regulate specific applications, purposes, areas or uses of AI, such as the CEPEJ *Ethical Charter*. In addition, a few proposals are addressed to specific stakeholders, such as the legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs), and the *Directive on Automated Decision-Making*, which call upon public authorities to take action. Seven proposals[102] specifically and restrictively refer to the governance of embodied AI, i.e., robots.

In terms of the substantive content, the reviewed proposals could be distinguished into proposals about: **a) who should be responsible for regulating AI** and similar technologies; **b) specific *ad hoc* measures to govern AI,** and **c) more generic measures** to regulate AI e.g., legislative reforms and frameworks and guiding principles.

The reviewed proposals could be classified in terms of the character of the proposed solutions to govern AI. Although all of the reviewed proposals aim to regulate, more or less directly, AI applications, the **suggested measures vary, including preventative, repressive and corrective solutions to AI-based systems.**

There are three main 'regulatory trends' as depicted in the reviewed proposals:

- There is a **commonly recognised need for regulation, soft or hard, of AI** and, ideally, at a supra-national level.
- A fair number of the proposals suggest the **creation of a regulatory agency** with registering and licensing, oversight, monitoring or enforcing powers.

[101] UN *Moratorium*, CEPEJ *Ethical Charter*, *DEEP FAKES Act*, *Directive on Automated Decision-Making*, US FDA regulation of adaptive AI/ML, register of algorithms.

[102] These are: Convention on human rights in the robot age; global legal AI and/or robotics observatory; EU-level special list of robot rights; adoption of common Union definitions of cyber physical systems, autonomous systems, smart autonomous robots; establishment of a comprehensive Union system of registration of advanced robots; general fund for all smart autonomous robots; US Federal Trade Commission to regulate robotics.

- A third common approach relates to the call for **reviewing the existing legal framework and either revising it to address the challenges and risks of AI or providing for specific legal acts or other instruments** (such as frameworks and codes of conduct) and tools to specifically govern AI.

4.2.2 Basis, nature and scope

1. Basis

For the purpose of this section, it should be noted that where the basis of the option is 'law' this has been understood broadly. It may require existing law to be adapted or interpreted in a different manner. It may also require enactments or amendments at a legislative or constitutional national level. A large number of the reviewed options also point to the solution of international legal agreements, namely at the Council of Europe or European Union level, or at a national (e.g., US Federal level). Other legal arrangements are also possible, with the proposal for regulatory sandboxes requiring the implementation of administrative processes to allow controlled testing.

In the reviewed proposals, the emerging common pattern indicates that:

- most options are based on existing law; and
- there is a need to provide for a new legal basis for the materialisation of some of the options.

New pieces of legislation

Where the option relies on legal bases and instruments, in most cases a new law is required, either at the national or supranational level. The suggested new pieces of legislation cover several legal fields and disciplines, ranging from human rights legislation to civil, criminal and public law. In other cases, specific legislation is recommended in existing legal disciplines, such as competition and anti-trust law (e.g., the anti-trust regulations recommended by the US Senator Elizabeth Warren).

On the contrary, where the proposal refers to ethical or non-binding guiding principles, there is limited legal effect or basis. Nonetheless, it is not precluded that these ethical principles could become a basis for legal action and integration, whereas, in some cases, the ethical principles draw on existing legal and human rights frameworks. For example, the Council of Europe's *European Ethical Charter on the use of Artificial Intelligence in judicial systems* has due regard to the fundamental rights guaranteed by the *European Convention on Human Rights* (ECHR) and the *Convention on the Protection of Personal Data* (*Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, ETS No. 108 as amended by the CETS amending protocol No. 223).

In some cases, a specific legal arrangement was suggested to address the challenge of uncertainty and potential high risks that AI may pose. This is the case of the moratorium on the development of lethal autonomous robotics, which is suggested as an early-stage measure to set the ground for international cooperation and agreement on the necessary solutions before the deployment or prohibition of AI uses.

What should be highlighted is that although the creation of a new piece of legislation is explicitly suggested or inferred (by the nature of the option) in the majority of the reviewed proposals, there is no reference to the content, scope, essence and structure of specific provisions and clauses. Furthermore, the interference and relationship with other legal disciplines and instruments is not analysed to prevent 'regulatory duplication', legal obscurity or inconsistencies.

Existing law as a legal basis for expansion and adaptation

Where the option relies on existing laws or frameworks, the proposal elaborates and expands on the legal provisions in place and recommends a 'legal widening' or 'adaptation' based on theories, comparative research and other frameworks to introduce new aspects in the existing legal landscape. Some reviewed options rely on existing governing mechanisms and frameworks by either amending the law or adopting a different interpretative approach to existing legislation. In this context, Kaminski and Malgieri[103] suggest the adoption of Algorithmic Impact Assessments (AIAs) under the GDPR. Although not specified, if this approach is endorsed but found inconsistent with the letter and/or spirit of the GDPR, a new legal act may be necessary to introduce the requirement for AIAs.

Similarly, the proposed redress-by-design mechanisms could be based on existing or new EU provisions in the same fashion as Article 25 of the GDPR on data protection, by design and default. The proposal for the US Federal Trade Commission to regulate robotics at the US level also relies on the current framework and mandate of the US Federal Trade Commission.

Existing law as a simulation case study

Among the reviewed options, specific reference is made to certain legal Acts as a case study. The *Accountability Act* of 2019, and the *DEEP FAKES Accountability Act*, (H.R. 3230) 116th Cong. (2019) proposed by Representative Yvette Clarke and referred to the Committee on the Judiciary, and the Committees on Energy and Commerce, and Homeland Security, belong to this category of regulatory options, where specific provisions are discussed.

Lack of detail on the basis

A few options **lack clarity and detail** on the appropriate and suitable legal bases, instruments and tools for materialising their recommendations. For example, the proposal for a general fund for all smart autonomous robots, or an individual fund for each and every robot category does not specify whether a legal act is required. However, it is reasonable to assume that the establishment of such instruments at a national or international level will require the intervention of the legislator via a Convention, Treaty, Regulation or Directive on compensation. Similarly, the options of mandatory consumer protection impact assessments, and the EU Task force of field specific regulators for AI/big data do not elaborate on the appropriate underlying or supporting legal bases.

2. Nature

The need for effective governance and regulation of AI is indicated by the fact that the **majority of the options are suggested as legally binding and enforceable mechanisms.** This means that should they apply, they should have a legal and mandatory status, such as the consumer protection impact assessment and the proposed Convention on human rights in the robot age. This is also aligned with the above findings that most of the reviewed options rely on existing or new legal bases or instruments.

A **small number of options are proposed as voluntary**, including the redress by design mechanisms for AI, and the non-binding recommendation for a legislative framework for independent and effective oversight over the human rights compliance of the development, deployment and use of AI systems

[103] Kaminski, Margot E. and Malgieri, Gianclaudio, "Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations", *U of Colorado Law Legal Studies Research Paper* No. 19-28, 2019. http://dx.doi.org/10.2139/ssrn.3456224

by public authorities and private entities. Although the distinction between mandatory and voluntary options remains relevant in this study, an interesting finding was that a **noticeable number of the proposals consider the most appropriate and inclusive governance mechanisms and tools, beyond this strict dichotomy of binding-voluntary.**

Indeed, some **options build on co-regulation or the combination of soft and hard law tools.** The proposal for the creation of the IAIO perfectly represents this approach, where the suggested organisation is defined in the proposal as a formal entity established by an international agreement governed by international law. Nonetheless, until its structures and mechanisms are formally completed and operationalised, it is suggested that its work should touch upon soft law instruments, namely standards and guidance as a reference point for national authorities. Similarly, the UK Centre for Data Ethics and Innovation (CDEI) is an advisory body to investigate and advise on how the UK could maximise the benefits of AI and data-driven technology; its work is closer to soft law rather than bringing legislative amendments or binding the public or private sector with enforceable recommendations.

Another case is the reference to both the voluntary and binding effects of the suggested options. For instance, the certification of algorithmic decision systems (ADS) could be on either a voluntary basis (as encouraged by the GDPR), or mandatory in certain areas such as justice and healthcare.

An interesting distinction to bear in mind is that the nature of the option may differ from the nature of the source including this option. This is the most frequent case, where the nature of the original proposal is not mandatory but a non-binding recommendation for consideration. Nonetheless, the nature of the suggested option, should this apply, is legally binding and mandatory.

3. Scope

The overall scope of the reviewed options is rather broad, aiming to cover various regulatory trends and options across the globe. Most of the reviewed options are **suggested or operationalised at a supranational or regional level,** either at the European Union or Council of Europe level. National and federal jurisdictions were also considered, including Canada, the United Kingdom, New Zealand, and options at a US federal level. The origin, application field and scope of the options are also indicative of the different approaches. For example, the three UK options relate to innovative legal solutions at a national level, i.e., the establishment of a Digital Authority and the CDEI, and the recognition of a duty of care. Despite the limited number of inputs from New Zealand, the single examined option on creating a register of algorithms was also useful for understanding and comparing the governance solutions to AI. Five options from the US were also considered, revealing different approaches to regulating sectoral uses of AI, including medical devices and deepfakes.

In some cases, the suggested proposal refers to national measures, such as the enactment of new legislation, but the proposal may have an international reach or basis, calling upon the national governments to legislate at a European or international level. In this context, **some proposals originate from supranational organisations or rely on international agreements, but they recommend actions and measures at a national level**. This is mainly the case of the proposals of the Council of Europe, such as the proposal of the Council of Europe Commissioner for Human Rights, which advises Members States to establish a framework for independent and effective oversight of AI applications.

In other cases, the proposal relates to the adoption of international measures, such as the recommendation of the Parliamentary Assembly of the Council of Europe, for the adoption of a new, international and legally binding Convention on human rights in the robot age, to **create common**

guiding principles to preserve human dignity in the way humans apply innovations in the field of the Internet of Things (IoT), including the Internet, robotics, AI, and virtual and augmented reality.

Another pattern worth noting is that the scope of some of the proposals is not limited to the governing law or jurisdiction of the bodies or authors of the proposals, but extends and generally covers AI uses and similar technologies. The three-level obligatory impact assessments for new technologies proposed by Paul Nemitz focuses on high-risk technologies and could become a reference point for different regulatory levels and territories. The same applies to other innovative and ground-breaking options. For instance, the UK CDEI operates in the UK and focuses on the UK market and legal order. Nonetheless, its impact is broader, with its innovative function and mandate alongside its recommendations addressing novel legal and technological questions beyond territorial or legal boundaries.

In terms of the material and substantive scope of the options, in a small number of the reviewed options, due regard has also been given to the specific application field of the AI-based technology. Further tailoring and considerations may apply in specific sectors, including the definition of specific evaluation criteria for the certification of algorithmic decision systems in certain areas such as justice and healthcare.

4.2.3 Purposes

In general, the reviewed proposals share the purpose and objective of ensuring that AI-driven systems are developed, designed and deployed in a manner that does not create risks for society and individuals. Whether on a hard or soft law level, the regulatory proposals purport to introduce general or specific mechanisms to better govern and regulate AI. They also outline the function, structure, tools, and aims of these mechanisms. Some specifically aim to address particular uses of AI, including deepfakes[104] and lethal autonomous robots.[105] In addition to the above, the proposals aim to shed light on the risks and challenges of AI, such as algorithmic opacity and potential data inaccuracies and bias, and the need to regulate its uses and applications.

Overall, the examined regulatory proposals could be considered to serve the purposes of **monitoring, preventing, licensing, restricting, certifying, assessing and controlling the uses of AI** or **setting standards and rules for the AI applications**. Regarding the audience for the examined proposals, there are several policymaking options addressed to the wider public for awareness, and policymakers for consideration and discussion. For example, the CDEI is tasked by the UK Government to connect policymakers, industry, civil society, and the public to develop the right governance regime for data-driven technologies.

Some regulatory options are standalone solutions to AI risks and challenges, whereas others have been suggested within a specific context to specify the application of this framework to AI uses. For instance, the CEPEJ *European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, aims to ensure that the use of AI tools and services in judicial systems improves the efficiency and quality of justice with due regard to the *European Convention on Human Rights* (ECHR) and the *Convention on the Protection of Personal Data*, whereas the suggested Algorithmic

[104] *Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019*, or the *DEEP FAKES Accountability Act*, (H.R. 3230) 116th Cong. (2019), and National Independent cross-sector advisory body (Centre for Data Ethics and Innovation).

[105] *Moratorium on the development of 'lethal autonomous robotics' (LARs)*, (UN report); *Moratorium on development of offensive LAWS* (AI HLEG).

Impact Assessments[106] aim to enhance algorithmic accountability and individuals' rights under the GDPR. Moreover, the Council of Europe Commissioner for Human Rights suggests the enactment of a legislative framework for independent and effective oversight over the human rights compliance of the development, deployment and use of AI systems by public authorities and private entities.

Depending on the nature and type of the regulatory option, each serves further specific purposes. In particular, some **options include the establishment and operation of regulatory bodies and agencies.** For instance, the Ad hoc Committee on Artificial Intelligence (CAHAI) of the Council of Europe aims to conduct a feasibility study and set the foundations of a legal framework and produce a progress report with proposals for further action and working methods by May 2020. SIENNA proposed a global legal AI and/or robotics observatory to systematically monitor and consider legislation, case law, emerging legal issues and inform future legislative work in AI and robotics. In a similar vein, Erdélyi and Goldsmith have proposed the establishment of an international AI regulatory agency to create a unified framework for the regulation of AI technologies and inform the development of AI policies around the world.

Finally, other options aim to provide ad-hoc solutions and recommendations, including:

- The establishment of a new statutory duty of care for online harms to restrict the use of AI in creating harmful content.
- Regulatory sandboxes to allow testing of regulatory schemes for new technology in a controlled environment, while enabling regulators and stakeholders to work together.
- Recognition of electronic personhood to establish accountability, liability and responsibility for decisions and actions taken by autonomous actors.
- Creation of a register of advanced robots within the European Union's internal market, or a register of algorithms in the public sector in New Zealand, to support traceability, monitoring and controls.
- Voluntary/mandatory certification of algorithmic decision systems to enhance trust in algorithmic decision systems and verify their compliance with commonly accepted rules.
- A general fund for all smart autonomous robots, or an individual fund for robots to guarantee compensation if the damage caused by a smart autonomous robot is not covered by insurance.

To conclude, although all the reviewed options aim to address the use, misuse, and abuse of AI, it is worth clarifying that these **options may serve slightly different or overlapping purposes**. For example, as indicated above, a few proposals aim to regulate AI at a more holistic level, either in terms of territorial scope, (i.e., Council of Europe, European Union, Federal systems or internationally) or material scope, i.e., AI in general. On the contrary, some options **target specific features and applications of AI or apply in specific legal fields**, e.g., competition and anti-trust law, or areas of deployment of AI, e.g. public sector. In addition, not all options require the same approach to achieve their stated purposes. The **stage of regulatory intervention also differs**, with some options aiming to regulate the development of AI, while others would regulate the AI applications (we recognise these two stages are tightly connected and regulating them jointly should be considered). Among the suggested governance mechanisms, the creation of regulatory bodies and the establishment of a legal framework are the most discussed. Finally, whereas the direct purpose of these options is to shed light on, and regulate the AI-based practices, it could be argued that their indirect purpose is to enhance trust in the use of AI under the suggested safeguards, and not to stifle AI.

[106] [106] Kaminski and Malgieri op. cit., 2019.

4.2.4 Gaps addressed

In general, all the reviewed proposals aim to address gaps in the regulation and governance of AI in specific fields or more generally. In brief, the identified gaps could be distinguished as outlined below:

Regulatory gaps, legal loopholes and lack of legal clarity

Some reviewed options[107] address the lack of commonly accepted standard definitions of autonomous systems, features and rules about the permitted uses of AI. Moreover, there is currently no legal framework to address the design of AI and push the public and private sector to conduct assessments on the systems' design. There is poor legal clarity and ill-designed consistency in regulating AI at national and supranational levels.[108] In this context, the proposal for a global legal AI and/or robotics observatory wishes to address gaps in legal knowledge and best practices around AI, whereas the proposal for the UK Digital Authority aims to address the failures of self-regulation. In addition, there is a wide lack of awareness and legal clarity from the side of companies and public authorities of whether and under what conditions AI is permitted

Lack of institutions, tools and mechanisms

A few reviewed options build on the identified lack of appropriate mechanisms, tools and bodies to ensure that the uses of AI are well-governed. For example, this is the case in the proposals for creating oversight bodies and regulatory agencies, such as the Centre for Data Ethics and Innovation (CDEI). In addition, the below options also address the lack of agreed measures, methodologies and standards to monitor and assess AI technologies:

- Register of algorithms used in government: Currently there is no systematic public tracking system for autonomous systems that have been deployed
- Algorithmic Impact Assessments under the GDPR
- Mandatory consumer protection impact assessments
- Regulatory sandboxes

Lack of technical and security standards

There is an increasing need for agreed protocols, security standards and specific performance objectives to enhance trust in AI and ensure that AI-based systems comply with specific rules.[109]

Lack of awareness

Whereas all reviewed options will eventually and gradually increase awareness of AI uses, some options specifically address the lack of public visibility of AI uses, including the proposal for the

[107] See, for example, Adoption of common Union definitions of cyber physical systems, autonomous systems, smart autonomous robots (*EU Parliament Civil Law Res 2017)*; *Algorithmic Accountability Act of 2019* (HR 2231, 116th Congress); *Directive on Automated Decision-Making*.

[108] SIENNA, *D4.2: Analysis of the legal and human rights requirements for AI and robotics in and outside the EU*, 2019; See van Veen, Christiaan, "Artificial Intelligence: What's Human Rights Got To Do With It?" *Points,* 14 May 2018. https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d

[109] See, for example, voluntary/mandatory certification of algorithmic decision systems (ADS) and US Food and Drug Administration regulation of adaptive AI/ML technology.

establishment of a comprehensive Union system of registration of advanced robots within the European Union's Market and the three-level obligatory impact assessments for new technologies.

Lack of rules to regulate the market and other fields

Some options aim to address gaps and challenges in relevant practice areas and legal disciplines due to the deployment of AI. For example, the uncontrolled use of AI has raised concerns about the uncertainty of the applicability of product liability law to decisions and actions taken by autonomous systems, the application of tort, civil, insurance, consumer protection and competition law in the context of AI technologies.[110]

It should be highlighted that most of the examined options address several legal, technical, regulatory and societal gaps. Overall, they reflect a lack of coordinated engagement of policymakers and other stakeholders at the global level.

4.2.5 Added value and stakeholders that would benefit

The added value

The added value of the examined proposals lies in the importance of the suggested measures as governance mechanisms to address the above-identified gaps, especially the lack of legal, regulatory and technical standards for AI across the globe. The examined proposals aim to bridge the gap between prohibiting and letting AI be uncontrolled, adopting a more flexible approach to AI. Moreover, most of the examined regional, national or context-specific proposals could become a reference point for governance and regulation in other emerging fields and countries. They could also set the ground for **international cooperation**, enhancing **legal clarity, trust** in the use of algorithms, and **accountability** for public authorities and private actors.

The added value of the examined proposals mainly relates to the suggestion of mechanisms and measures to **regulate non-regulated AI uses**, such as deep fakes and lethal robots (where such use falls outside existing law). Where the suggested options refer to creating a binding and integral framework, such as the *Binding Framework Convention* to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law (Council of Europe), this is also of salient importance, since such solutions could support consistent regulation and governance of AI, promoting investment, development and implementation of safe AI systems in the EU.

Another important element of the examined options is that their **focus on specific areas and applications** of AI[111] could significantly support policymakers and companies in considering and adapting these options to similar fields or whenever AI is used more generally. In addition, most of

[110] See indicatively, the proposals for FDA for algorithms; creating electronic personhood status for autonomous systems; general fund for all smart autonomous robots or individual fund for each and every robot category; using anti-trust regulations to break up big tech.

[111] This category includes, for example, *Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019* or the *DEEP FAKES Accountability Act*, (H.R. 3230) 116th Cong. (2019); establishment of a comprehensive Union system of registration of advanced robots within the Union's internal market where relevant and necessary for specific categories of robots and establishment of criteria for the classification of robots that would need to be registered (*EU Parliament Civil Law Res 2017/MK*)

the proposals point out the need for consistency, public engagement, standards, clear rules and governing principles.

Finally, some options relate to legal acts and legislative amendments, namely the *Directive on Automated Decision-Making and Algorithmic Accountability Act of 2019* (HR 2231, 116th Congress), or suggest innovative approaches/support mechanisms **to boost the existing legal framework**, such as Algorithmic Impact Assessments under the GDPR.

Stakeholders that would benefit

A large number of various stakeholders could be expected to benefit from the examined proposals. Overall, **individuals** will be most benefitted, especially if AI targets specific groups/communities of individuals, or disadvantages consumers purchasing AI-based systems. The suggested options aim to empower individuals, restrict the risky uses of AI, and enhance legal clarity, compliance with legal requirements, transparency and accountability.

Among the highly-advantaged stakeholders**, policymakers and regulators** enjoy a special position. The examined options could be used as simulation cases or knowledge bases for evidence-based and engaged policymaking and decision-making and further policy and impact assessments.

The reviewed options do not neglect the legitimate interests and rights of the industry, which will engage in AI applications under the safeguards of legal certainty and clarity.

Finally, specific stakeholders will benefit based on the nature and type of the option. For example, the *Moratorium on the development of 'lethal autonomous robotics' (LARs)/development of offensive LAWS*, considers the protection of 'rights' of victims, i.e., civilian populations who might be affected by the use of lethal autonomous robotics.

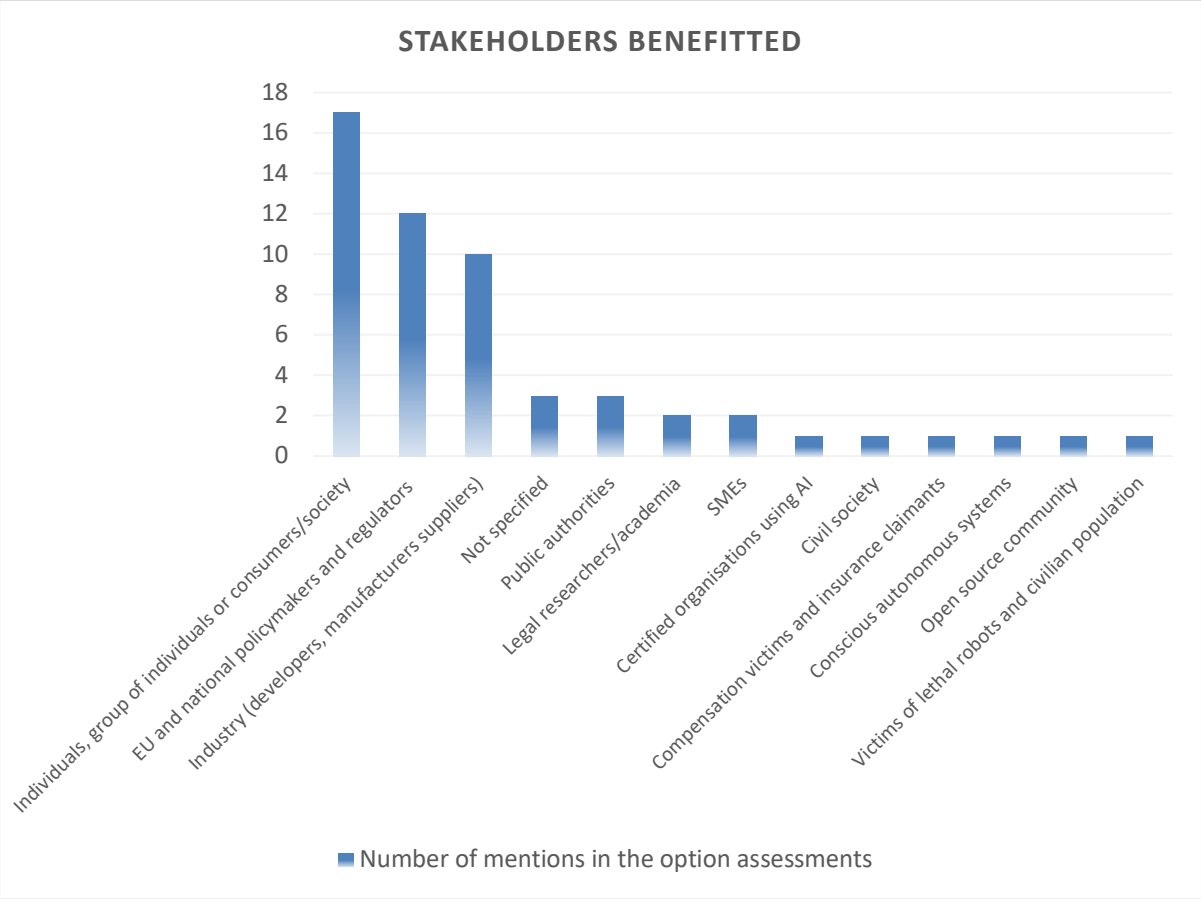

Fig 2: Stakeholders benefitted

4.2.6 Limitations, risks and challenges

There are **several limitations, risks and challenges** for the adoption and implementation of the reviewed options. A few options explicitly acknowledge the relevant limitations, risks and challenges, whereas other options do not. Moreover, a few of the suggestions have **no tested precedent at an EU level**, such as the register of algorithms. Therefore, consideration has been given to each option and its contextual, practical and legal implementation (often outside the framework of the examined options), to elaborate on their limitations, risks and challenges.

Limitations, risks and challenges mainly **depend on the nature and type** of the option (e.g., establishment of a regulatory body, soft law vs hard law options), the expected outcome (e.g., regulatory framework, technical standards), the material (e.g. AI in general or specific applications), and territorial scope (e.g. options with national or supranational effect and application field).

Limitations

The limitations of the reviewed proposals mainly refer to their scope and objective. A few options have a **limited application field** and apply to specific AI applications or jurisdictions. For example, the proposal for a register of algorithms used in government is limited to predictive algorithms and uses of AI by governmental departments in New Zealand, and does not apply to commercial uses of AI. In other cases, the **scope is rather broad** and the interplay with potentially overlapping legal frameworks should be considered. For example, the Canadian *Directive on Automated Decision-Making* is broader

than the requirements for automated decision-making under the GDPR, as it also applies to systems that assist in human-directed decisions.

Other limitations relate to the **lack of specific features, principles and functions** of the examined options. For example, although the publication of the results of the algorithmic assessments under the *Algorithmic Accountability Act of 2019* (HR 2231, 116th Congress) in the US is not provided, it could support its objective and enhance transparency.

The implementation of these options could be hindered due to other **technical, legal, or operational limitations**. In our study, these included:

- The establishment of oversight bodies with soft law powers and tools (e.g., Centre for Data Ethics and Innovation).
- Focus on specific jurisdictions and territories may hinder cooperation and consistency in a world where AI applications may have unlimited effects beyond territories.
- Lack of specific and tailored evaluation criteria (e.g., CEPEJ *European Ethical Charter on the use of Artificial Intelligence (AI) in judicial systems and their environment*).
- Fundamental notions should be defined better, considering the regulatory and practical effect of the options (e.g. this relates to the definition of deepfakes under the *DEEP FAKES Accountability Act*. Moreover, the suggested statutory duty of care for online harms (UK Government) lacks a clear delineation of legal but "harmful" content to be regulated).
- A more comprehensive and inclusive approach to the scope of options is required (e.g. establishment of a comprehensive Union system of registration of robots within the Union's internal market).
- Over-focus on the risks of AI and neglecting the relevant benefits and advantages.
- Over-focus on bias and discrimination and neglecting the examination of other fundamental rights and freedoms.
- Consideration should be given to various legal frameworks and requirements (e.g. data protection and intellectual property law).
- Where the options relate to bodies and agencies, their success may be limited by the capacity of their members (e.g., EU Taskforce of field specific regulators for AI/big data).
- Lack of transparency and clarity on the operation, powers, and relationships of agencies with other regulatory authorities (e.g., CDEI and IAIO). A similar limitation relates to the lack of management stability, weak collective control and oversight (e.g. International Artificial Intelligence Organization).
- Restricted understanding of the features and capabilities of AI.
- Constraints of resources (e.g., an FDA for algorithms)
- Lack of independence of publicly-funded bodies.
- Lack of consistency and consideration of the legal and political idiosyncrasies where the option has a supranational effect (e.g., *Moratorium on the development of 'lethal autonomous robotics' (LARs)* (UN report); *Moratorium on development of offensive LAWS* (AI HLEG)).
- Poor consideration of racial and gender bias and data privacy issues (e.g., US Food and Drug Administration regulation of adaptive AI/ML technology).
- Need for more detail on the practical and operational implementation of the option.
- Lack of consideration of both medium and long-term impacts.
- The political nature of the human rights framework poses some risks for the effectiveness of impacts assessments as a policy option (e.g. Legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities).

Risks

Risks relate to the occurrence of remote or probable threats and harms. Risks could relate to the rights and interests of individuals, stifling AI, or adversely affecting the existing regulatory framework or industry. A **major risk is to consider these options as panacea or replacement of existing frameworks**. For example, the report outlining the AI HLEG *Ethics Guidelines for Trustworthy AI* states that certification can "never replace responsibility. It should hence be complemented by accountability frameworks, including disclaimers as well as review and redress mechanisms".

A taxonomy of risks is challenging in the present study since the examined options require different approaches and relate to different measures. On the contrary, genuine risks have been identified while reviewing these proposals. The main risks are outlined below.

- **Privatization of regulation and scrutiny**, where private actors are involved (e.g. Voluntary/mandatory certification of algorithmic decision systems (ADS)).
- With specific regard to principled approaches, there is a **risk of providing false assurances** of fair, trustworthy and/or ethical AI (e.g. creating electronic personhood status for autonomous systems).
- Where the options require the publication of AI sensitive information, there may be **conflicts with intellectual property rights and prohibitions of releasing sensitive information** to the public (e.g. register of algorithms used in government and Algorithmic Impact Assessments under the GDPR).
- Where regulatory agencies are suggested, the mandate and powers should be clear, otherwise this may **duplicate the work of existing EU or national agencies** (e.g., EU Taskforce of field specific regulators for AI/big data).
- Regarding proposals for compensatory funds for AI incidents, the total amount of established claims may **exceed the aggregate amount of compensation available** (e.g., general fund for all smart autonomous robots, or individual fund for each and every robot category).
- **Negative impact on human rights** is also possible. New legislation regulating AI applications could threaten free speech (e.g., the *DEEP FAKES Accountability Act*, (H.R. 3230) and regulatory sandboxes). Similar to this, there are concerns for the freedom of expression if a statutory duty of care for online harms (UK Government) is established.
- If regulatory sandboxes are permitted, it is likely that this may provide participants with **unfair competitive advantages** both in regulatory advice and in being first to the market if appropriate safeguards do not apply. In addition, regulatory sandboxes could result in harm and liability issues in cases of failed testing, which could also damage the reputation of the regulator and public trust in the regulatory system.
- Where oversight bodies are suggested, **mission creep** is likely.
- Where interventions in specific legal areas are suggested, such as consumer protection and antitrust law, **adverse effects and deregulation** are likely.

Challenges

Challenges relate to internal or external factors that could prevent, fail or hinder the successful and effective deployment of the reviewed options. In general, a common challenge in implementing the suggested regulatory options is allocating resources, time and effort in further examining their appropriateness and efficiency, and designing their implementation, including policy assessments and stakeholder engagement. Another common challenge relates to the AI itself, i.e., AI development lacks common definitions, approaches, methods, aims, history, rules, principles, legal and accountability mechanisms. Moreover, there is no agreement or consensus on the definition of algorithms and AI, resulting in **confusion and ill-applied measures**, especially in voluntary systems. Definitions of AI technology need to be flexible enough to not hinder innovation (e.g., adoption of common Union

definitions of cyber physical systems, autonomous systems, smart autonomous robots (*EU Parliament Civil Law Res* 2017))

A main category of challenges for the effective adoption and materialisation of the examined options refers to the **need for a supporting framework and tools** to underpin the implementation of the proposals. Most of the options refer to *ad hoc* solutions and measures and not inclusive, thorough or holistic approaches to regulating AI. Therefore, these measures do not plan to replace existing measures and frameworks.

Regarding the proposals for international organisations, there is a challenge that Member States do not engage sufficiently/support such an instrument (e.g., *Binding Framework Convention* to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law (Council of Europe)).

Other challenges relate to:

- Resistance from actors to share sensitive information through impact assessment (Legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities).
- Operational burdens and resource-demanding options (such as impact assessment and regulatory agencies).
- Providing for and implementing safeguards for transparency and independence (e.g. where regulatory agencies are recommended).
- Need for consistency and established standards (e.g., methodologies, standards, risk scoring guidance for impact assessments).
- Lack of an accountability framework providing for sanctions in the case of failure to apply an option.
- Where the option relies on existing frameworks and suggests a novel approach to these, the main challenge relates to adopting the new interpretation. This requires acceptance by key stakeholders, including academia, human rights and civil organisations, and finally confirmed by European and national legislators or authorities. (Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations)
- Sustained funding and resources, ability to remain independent of the Government, support from policymakers and AI experts (EU Taskforce of field specific regulators for AI/big data, UK House of Lords Select Committee on Communications).
- Endorsement of new regulatory approaches (e.g., EU Taskforce of field specific regulators for AI/big data)
- Coherence in the definition of compensation award elements (General fund for all smart autonomous robots, or individual fund for each and every robot category)
- Focus on enforceable and binding hard law tools and solutions rather than on soft law.
- The relationship of oversight bodies with other supervisory authorities should be considered. In particular, due consideration should be given to ensuring that there is no conflict of interests or overlapping responsibilities among supervisory authorities (New statutory duty of care for online harms (UK Government) and Legislative framework for independent and effective oversight over the human rights compliance of the development, deployment and use of AI systems by public authorities and private entities).
- Lack of good implementation models of such mechanism (redress-by-design mechanisms for AI).
- Cooperation with national and European regulators and policymakers.

- Reliance on the political will of Member States to adhere to the option (e.g., Legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities).

4.2.7 Clarity, specificity, operationalisation and provisions for funding and review

Clarity and specificity for effective and efficient operationalisation

While some of the studied options were found to be clear and specific (e.g., *Directive on Automated Decision-Making*, *Algorithmic Accountability Act of 2019*[112], Register of algorithms used in government), many of the studied options are not sufficiently clear; in many cases they have been proposed and/or suggested very briefly[113], without detailed explanation, and often lack critical details. Some have been criticised as being too fuzzy.[114] In the case of the moratorium on LARS/LAWS, the underpinning frameworks (covering process, effects) have not been fully specified (e.g., The UN HRC's proposal had more detail than the other two documents on the subject), leaving critical questions unanswered.

In some cases, operationalisation has clearly not been considered, e.g., in the CEPEJ *European Ethical Charter on the use of Artificial Intelligence (AI) in judicial systems and their environment*, and the new statutory duty of care for online harms (UK Government). The European Parliament Resolution proposing the creation of electronic personhood status for autonomous systems is also not clear enough – it does not identify the parameters of the status, or which systems would be eligible for it (other than "at least the most sophisticated autonomous robots"). The proposed *Convention on human rights in the robot age* also does not provide any specifics. In the case of the CoE *Binding Framework Convention*,[115] while the option is sufficiently clear and specific, its effectiveness is limited as the mandate does not extend to drafting the legal framework itself; operationalisation cannot therefore be achieved. In the case of the US *DEEP FAKES Accountability Act* (HR 3230), its limitations have been outlined as obstacles to operationalisation.
Though some options might be clear as to their objectives/vision and intent, they lack detail on key elements: e.g., Kaminski and Malgieri's proposal for Algorithmic Impact Assessments (AIAs) under the GDPR needs more detail about the structure, order and planning of AIAs to ensure common standards, consistency and legal certainty. The legislative framework for independent and effective oversight over the human rights compliance by the Council of Europe Commissioner for Human Rights lacks

[112] Though the FTC would need to define some of the requirements, such as how often to require updated algorithmic assessments when an automated decision system changes.
[113] E.g., the legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities; redress by design mechanisms; voluntary/mandatory certification of algorithmic decision systems (ADS); EU Taskforce of field specific regulators for AI/big data authorities; redress by design mechanisms; voluntary/mandatory certification of algorithmic decision systems (ADS); EU Taskforce of field specific regulators for AI/big data
[114] This is evident in relation to the proposal to use anti-trust regulations to break up big tech and appoint regulators to reverse illegal and anti-competitive tech mergers. See: The Economist, "Dismembering big tech", *The Economist*, 24 Oct 2019. https://www.economist.com/business/2019/10/24/dismembering-big-tech
[115] *Binding Framework Convention* to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law (Council of Europe), including through a new ad hoc committee on AI (CAHAI)

details about the legal nature, structure, aims, powers, tools, cooperation, liability, composition, and accountability of this regulatory option. The AI HLEG proposal for mandatory consumer protection impact assessment lacks any form of detail on this - it is unclear who should be responsible for effecting and operationalising such assessments. The EU Parliament proposal to adopt common Union definitions of cyber physical systems, autonomous systems, and smart autonomous robots does not offer any definitions or subcategories, other than describing several mandatory and optional characteristics of a smart autonomous robot. The US FDA regulation of adaptive AI/ML technology may require additional statutory authority.

Where proposals for bodies are concerned, where a proposal has been fructified, e.g., the UK CDEI, it is (unsurprisingly) clearly specified. For the UK Digital Authority, details were not clear enough (with respect to its functions, instruction remit, relationships with other bodies). In the case of the International Artificial Intelligence Organization (IAIO), Erdélyi and Goldsmith have not defined the IAIO's precise purpose, membership, the issues to regulate, and the broad directions to follow (for want of international consensus). Tutt has outlined the powers of an FDA for algorithms in various capacities.

In a sense, there seems to be a greater likelihood of national-level proposals being more well-developed and good-to-go in terms of operationalisation, than international and/or EU-level ones (depending on the political will and buy-in). In some cases where details were lacking (e.g., regulatory sandboxes, EU-level special list of robot rights, establishment of a comprehensive Union system of registration of advanced robots), our individual assessments have anticipated some key elements and/or what needs to be done for effective operationalisation (within limits).

Sources of funding

The issue of funding is critical, especially in terms of setting up and maintaining effective operation of a regulatory agency/body and its sustainability; it is also relevant to legislative proposals, and other mechanisms and tools – it should be clearly determined from where new requirements to be imposed or measures to be implemented should be funded. It is not enough to have **funding** – it should also be **proper and adequate** for the purpose. But funding (especially public) is not unlimited and this is a critical part of why some regulatory proposals remain just that. Most of the studied options **have not defined sources of funding or have not explicitly addressed this** question (sometimes it might not be applicable, e.g., adoption of common Union definitions). The exceptions to this, where funding is addressed) include:

- *The Binding Framework Convention* to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law/CAHAI (Council of Europe) has a well-clarified source of funding, present and future, for intergovernmental work (i.e., meetings, reporting). It is unclear however, whether the funding would extend to anybody/agency/authority to be created.
- The proposal for Algorithmic Impact Assessments (AIA) under the GDPR suggests innovative ways to fund and support the AIA model and process - it requires companies or regulators to help fund the involvement of both, and provide technical expertise or the resources for obtaining technical expertise during the deployment of the AIA.
- New statutory duty of care for online harms (UK Government): The White Paper proposes that new fees, charges or a levy on in-scope entities will cover the costs of the regulator.
- Proposal for US FTC to regulate robotics: The FTC is already a federally funded agency. If the FTC takes over work that had been performed by other agencies, a re-allocation of budget resources would be needed.

- The UK CDEI (operational body), which is funded by the Department for Digital, Culture, Media and Sport with £2.5 million in 2019/20 and £5 million in 2020/21.
- Regulatory sandboxes are typically funded by the regulatory agency/body overseeing the sandbox.

Provisions for regular review and update

Regular review is needed to evaluate the effectiveness of the legislation, measure, agency or other regulatory mechanism. Where found lacking, amendments, changes and/or improvements need to be made to ensure objectives are being met, it is having the right impact and its adoption/enforcement is effective. Many of the studied proposals have considered this, but to varying degrees and in different manners. In some cases, none were elaborated in the proposal, i.e., there were no provisions/explicit instructions/lack of information about the review/update of the proposal itself.[116] These have probably been left to be defined in the legislative instrument or terms of reference/mandate/operational policy of the regulatory agency/advisory body when further defined. Examples of how some studied options have covered this need include:

- The CDEI has provisions for internal monitoring, tracking progress including via the use of metrics to track the full range of activities, publishing assessments via annual reports.
- AIAs under the GDPR propose that model AIA should be truly continuous: a process that produces outputs, but also includes ongoing assessment and performance evaluation, especially for those algorithms that change quickly over time.
- The register of algorithms has no specific provisions, but the information-keeping nature of this measure suggests that such registers should be reviewed and updated to reflect the status and categories of the algorithms applied.
- The CEPEJ *European Ethical Charter* suggests "the independent authorities mentioned in the Charter could be responsible to periodically assess the level of endorsement of the Charter's principles by all actors, and to propose improvements to adapt it to changing technologies and uses of such technologies."
- *The Directive on Automated Decision-Making* has an automatic review process planned every 6 months after the effective date.
- *The US DEEP FAKES Accountability Act* (H.R. 3230) contains provisions for the Attorney General, in coordination with other relevant Federal agencies, to submit a report to Congress every five years after enactment; the Attorney General shall publish a report containing, inter alia, (in order to increase the likelihood of such prosecutions), official guidance to Federal prosecutors regarding any potential legal concerns that may impede such prosecutions absent clarification.

Therefore we need good terms of reference for regular review and update - e.g., assessment of performance; assessment of the impacts (legal, political, economic, ethical) of the proposal; status

[116] These include, e.g., *Moratorium on LAWS/LARs*; adoption of common Union definitions; establishment of a comprehensive Union system of registration of advanced robots; proposals for the mandatory consumer protection impact assessment; legislative oversight framework; EU-level special list of robot rights; *Convention on human rights in the robot age*; EU Taskforce of field specific regulators for AI/big data; voluntary/mandatory certification of algorithmic decision systems; International Artificial Intelligence Organization; *Binding Framework Convention* including CAHAI; creating electronic personhood status for autonomous systems; general fund for all smart autonomous robots; mandatory consumer protection impact assessment; redress by design mechanisms; UK digital authority; proposal for the US Federal Trade Commission to regulate robotics; US Food and Drug Administration regulation of adaptive AI/ML technology; new statutory duty of care for online harms; the *Algorithmic Accountability Act* of 2019; an FDA for algorithms; using anti-trust regulations to break up big tech.

and uptake; effect of ending the measure (e.g., if it is a moratorium); whether new laws, complementary policies and practices might be needed; effectiveness of monitoring and enforcement (if applicable); and whether updates are needed due to new technological developments and/or changes in societal expectations and values. The timing of the review and update might vary but need to be clearly defined and the period specified, along with the procedural rules governing such a review and update.

4.2.8 Monitoring, oversight and enforcement

Monitoring, oversight and enforcement are addressed to different degrees in the reviewed proposals. Some proposed regulatory mechanisms are better defined (e.g., *US Algorithmic Accountability Act 2019*, *US DEEP FAKES Accountability Act*) than others. In some cases, it was not possible to identify clearly the monitoring, oversight and enforcement mechanisms due to lack of development of the proposals and/or insufficient detail available, or this remains yet to be determined/specified.[117] The presence of such mechanisms also goes to the heart of the type of regulatory proposal under scrutiny and their scope and mandate. E.g., In the case of the proposed Digital Authority, as it won't be a new 'regulator', it won't have independent monitoring, oversight and enforcement mechanisms. While the UK CDEI has a Board with internal oversight and monitoring, as far as external monitoring, oversight and enforcement are concerned, it has no such powers. Monitoring, oversight and enforcement mechanisms would also not be applicable in the case of, e.g., a global legal AI and/or robotics observatory.

As outlined above, a closer look at the individual options will show there are some gaps and room for improvement – e.g., in relation to the register for algorithms, though there is some information, more detail and provisions are required regarding the monitoring and enforcement powers of the regulatory agency responsible for the register.

4.2.9 Implementation burdens and challenges

Burdens on citizens

None of the reviewed options were reported as presenting any implementation burdens on citizens. In many cases this is not surprising, as the options studied do not directly call for action on their part, meaning there is at least no direct implementation burden. There might, however, be a burden in terms of participation (e.g., in the activities organised by the CDEI). In terms of the *DEEP FAKES Accountability Act*, there might be a chilling effect burden on the public/citizens, regarding the production of satirical or parodic political videos.[118]

Burdens on public administrations

[117] I.e., *Convention on human rights in the robot age*; EU-level special list of robot rights; establishment of a comprehensive Union system of registration of advanced robots; creating electronic personhood status for autonomous systems; adoption of common Union definitions; EU Taskforce of field specific regulators for AI/big data; voluntary/mandatory certification of algorithmic decision systems; mandatory consumer protection impact assessment; US Food and Drug Administration regulation of adaptive AI/ML technology; new statutory duty of care for online harms.

[118] Where the legally required elements could easily be stripped, to say nothing of the vagueness of to "humiliate or harass".

Many of the reviewed options call for active action from the government and its agencies, and carry compliance or operational burdens, e.g., amendment of existing law; enactment of new legislation and regulations; establishment of oversight bodies; carrying out of HRIAs on AI systems acquired, developed and/or deployed by those authorities; management and enforcement (e.g., in cases of registration systems for advanced robots, *DEEP FAKES Accountability Act*, selecting appropriate targets in case of anti-trust regulations); supervision and oversight (EU Taskforce of field-specific regulators).

This will require political will, investment of time and resources (expertise, and adequate financing – particularly relevant in the cases where new bodies are proposed, and the general fund). Depending on the regulatory option and the result sought to be achieved, the degree of the burden will vary.

Burdens on businesses, including SMEs

In many cases there will be a burden (compliance-related[119]) on businesses. These might vary depending on whether businesses need to mandatorily comply or chose to do this of their own volition. As pointed out by the HRIA legal framework assessment, "If the law imposes **excessive demands**, it will be a burden on the organisations that come under its scrutiny (and the **severity might be greater for SMEs** servicing the public sector and unprepared to mitigate adverse AI effects due to lack of will, policy or resources)".[120] There might also be **non-implementation reputational risk** burdens.

We are particularly interested in the burdens on SMEs, who are valuable players (undiscussed in the AI market), and at the same time the players that might struggle to cope with new requirements (in many cases they might not have a choice if a special consideration is not made for them). For SMEs, being under the control and monitoring of AI oversight bodies may be a **disincentive or hindrance in engaging with AI**, especially for start-ups, or may result in cover-ups of ethical issues (Legislative framework for independent and effective oversight) neither of which are ideal for the future regulation and responsible development of AI.

In all cases, burdens should be proportional to what is sought to be protected.

Implementation challenges

Other implementation challenges identified in the review of the options included:

- Creation of loopholes, i.e., the narrow scope of the moratorium on LARS/LAWs and its many exclusions and exceptions may lead to disappointing results.
- Difficulty of developing appropriate human rights indicators that have the required contextual specificity, tailored to the problems of the country concerned. (e.g., Legal HRIA Framework).
- Need to develop capacity to respond effectively due to the transboundary nature of AI and to constantly think how to interpret its mandates in light of emerging issues (IAIO).
- Institutional relations and cooperation (global legal AI/robotics observatory).
- Over-protectionism, and ambiguity and complexity of legislative requirements that form its basis (Voluntary/mandatory certification of algorithmic decision systems).
- Politics and national vetoes (redress by design mechanisms).

[119] Putting new measures in place to meet requirements of laws, certification, carrying out HRIAs or supporting this process, performing algorithmic or data protection impact assessments, cooperating and providing information to relevant authorities, carrying out training activities, and/or awareness-building.
[120] See Annex 4.4 of this report.

- reluctance to use antitrust laws to require market-dominant companies to assist their competitors (anti-trust regulations proposal).

4.2.10 Rights/interests affected

The findings vary as to the question of what rights or interests the reviewed options might neglect.

In some cases, these were specifically identifiable, though variable, in terms of whose rights/interests are affected, e.g., the LARs/LAWS moratorium neglects the rights/interests of states investing in building arsenals/deploying LARs/LAWS. In the case of the global legal AI and/or robotics observatory, the interests of actors/stakeholders who might not want their actions featured might be affected. The *Binding Framework Convention* is missing the direct democracy element (it also neglects specific human rights). The EU-level special list of robot rights doesn't address legal responsibility for autonomous decisions or actions by conscious autonomous systems which cause injury or damage to humans or property, other than noting that the proposed right would not apply for rogue robots. The mandatory consumer protection impact assessment neglects the interests of developers/manufacturers/suppliers (industry), in that it does not resolve conflicts between corporate gains and citizen rights, and policymakers and regulators, as it does not provide direction on implementation. The establishment of a comprehensive Union system of registration of advanced robots could neglect the rights/interests of data subjects. The general fund for smart autonomous robots could place a burden on the government in terms of coordination and management of the fund. The HRIA legal framework does not neglect any rights, as such, but will have some potentially adverse impacts on those subject to an HRIA. The CEPEJ *Ethical Charter* might neglect the interests of private stakeholders responsible for the design and deployment of artificial intelligence tools and services.

The register of algorithms used in government seems to neglect the interests and rights of public authorities and private companies, if confidential and sensitive information is requested from them about the uses of algorithms. A new statutory duty of care for online harms might neglect people's interests/rights of privacy and free expression. The *Directive on Automated Decision-Making* does not address use of AI/big data by provincial or local governments or by non-governmental entities. The *DEEP FAKES Accountability Act* would affect software manufacturers who will have their rights neglected at the expense of this putative greater good. Voluntary/mandatory certification of algorithmic decision systems (ADS) might neglect the interests of consumers, as costs of certification compliance might get passed down to them (even though it seeks to protect them from adverse impacts). The proposed *Algorithmic Accountability Act of 2019* does not address AI systems used by the public. The anti-trust regulations would adversely affect the big technological companies i.e., Amazon, Google, Facebook. The three-level obligatory impact assessments for new technologies neglects industry, especially SMEs (additional burdens).

In other cases[121], which parties' rights or interests would be neglected was not clear (i.e., either due to their being insufficiently elaborated, or need for further development of the option which would then have a bearing on this).

[121] e.g., IAIO; legislative framework for independent and effective oversight, adoption of common Union definitions; creating electronic personhood status for autonomous systems; *Convention on human rights in the robot age*; AIAs under the GDPR; EU Taskforce of field specific regulators for AI/big data; CDEI; Digital Authority; US Food and Drug Administration regulation of adaptive AI/ML technology; proposal for US Federal Trade Commission to regulate robotics; FDA for algorithms; redress by design mechanisms; regulatory sandboxes.

4.2.11 Impact on human rights

This section examines whether the reviewed options explicitly support or adversely affect human rights. As seen below, nearly half of the reviewed options explicitly support human rights[122]. Others might have this effect more indirectly[123], and some do not address this, and/or human rights falls outside their scope, or has not been sufficiently defined.

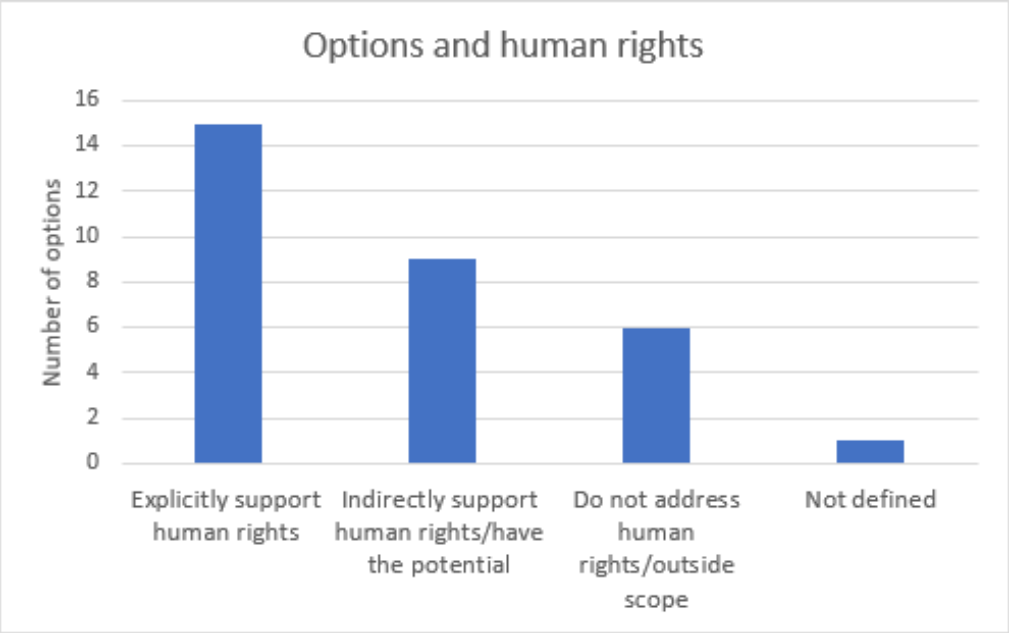

Fig 3: Reviewed options and human rights

The options might boost human rights in various ways, e.g., requiring assessments to identify and reduce risks of high-impact automated decisions (e.g., *Algorithmic Accountability Act of 2019*); protecting democracy and privacy by promoting healthy competition (anti-trust regulations); safeguard and enhance human rights, such as the right to privacy and freedom from discrimination, in the context of predictive algorithms (e.g., register of algorithms used in government); introducing compliance mechanisms to monitor, prevent and manage risks for human rights (legislative framework for independent and effective oversight; legal framework for human rights impact assessments); preventing loss of life plus the devaluation of it (e.g., LARs/LAWS moratorium).

Human rights covered by the reviewed options include: **data protection, dignity, equality, freedom from discrimination, privacy, and the right to life.**

[122] I.e., LARs/LAWS moratorium; CEPEJ *European Ethical Charter*; *Binding Framework Convention*; legal framework for HRIAs; legislative framework for independent and effective oversight; Algorithmic Impact Assessments under the GDPR; *Convention on human rights in the robot age*; mandatory consumer protection impact assessment; register of algorithms used in government; Centre for Data Ethics and Innovation; *Directive on Automated Decision-Making*; Digital Authority; *Algorithmic Accountability Act of 2019*; anti-trust regulations; three-level obligatory impact assessments for new technologies.

[123] I.e., IAIO; global legal AI and/or robotics observatory; redress-by-design mechanisms for AI; establishment of a comprehensive Union system of registration of advanced robots; EU Taskforce of field specific regulators for AI/big data; voluntary/mandatory certification of algorithmic decision systems (ADS); general fund for all smart autonomous robots; *DEEP FAKES Accountability Act*; new statutory duty of care for online harms.

4.2.12 Ethical principles and gender dimensions

Coverage of ethical principles in the reviewed options

Ethics and ethical principles have been broadly considered when assessing the reviewed proposals.[124] As part of the individual assessment of the regulatory proposals, we looked at how they addressed ethics and ethical principles. The identified ethical principles permeate the use of AI and big data across the public and private sector as well as the individual and collective sphere. In our review of the studied options, we came across ethics as both a principle governing the design and implementation of the 'options', and a target/aim of the options. The main prominent principles were fairness/fair treatment, transparency, accountability, and prohibition/minimisation of bias and discrimination. The following illustration shows the key ethical principles that featured repeatedly in the examined proposals.

Fig 4: Featured ethical principles – top scorers

In addition, other less discussed ethical principles were also mentioned in the reviewed proposals – these are important to consider and take into account:

[124] Noting that some of the mentioned principles would not strictly qualify as ethical principles in the traditional sense of the term.

Fig 5: Other ethical principles featuring in the studied proposals

A few of the studied options did not make any mention of, or had insufficient detail on ethics or ethical principles, e.g., global legal AI and/or robotics observatory, adoption of common Union definitions, creating electronic personhood, EU Taskforce of field specific regulators for AI/big data, IAIO, anti-trust regulation proposal. Some, such as the *Convention on human rights in the robot age*, and the legal framework for HRIAs, could be stated to have an implicit focus on these through the explicit focus on 'human rights'.

Gender dimensions

The vast majority of the examined proposals do not explicitly consider gender dimensions. However, the *Binding Framework Convention*[125] has a gender dimension in its guiding principles, and the Committee will appoint a Rapporteur on Gender Equality from amongst its members. Bodies such as the CDEI indirectly address gender dimensions in promoting equality and diversity and fairness of opportunity (as we expect will also be the case for the other proposals, since the law explicitly prohibits discrimination based on gender). The CDEI also states it takes into account risk of gender bias and discrimination where relevant. The proposal for AIAs under the GDPR might not explicitly

[125] *Binding Framework Convention to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law* (Council of Europe), including through a new ad hoc committee on AI (CAHAI).

consider gender dimensions but might embed such considerations in the narrative and methodology of this AIA. A sandbox could factor gender equality into the selection of participants.

The recommendation underpinning the proposed legal framework for HRIAs calls for implementing the *UN Guiding Principles on Business and Human Rights* and *Recommendation CM/Rec(2016)3* of the Committee of Ministers to Member States on human rights and business, in a non-discriminatory manner with due regard to gender-related risks. The US *Algorithmic Accountability Act of 2019* does not explicitly address gender, but the proposal is based on identifying and reducing risks of biased and discriminatory decisions, which includes gender bias. The US *DEEP FAKES Accountability Act* Bill provides that on the effective date of the Act, the Attorney General shall publish a report containing, inter alia, a description of the impact of intimate and sexual deep fakes on women and marginalized communities.

4.2.13 Feasibility and sustainability

This section provides some insights into the feasibility, sustainability and future-proofing of the studied options (using data derived from responses to question 17 of the option study).

Feasible and sustainable options

The following were identified as feasible and sustainable:

- International Artificial Intelligence Organization (if it draws the right kind of international support from policymakers).
- Global legal AI and/or robotics observatory (feasible, depending on policy and good financial support).
- CEPEJ *European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment* (has received a lot of publicity and featured in training courses and masterclasses and other dissemination and awareness activities).
- Establishment of a comprehensive Union system of registration of advanced robots within the Union's internal market.
- Adoption of common Union definitions of cyber physical systems, autonomous systems, smart autonomous robots (as part of a broader EU law on autonomous systems).
- General fund for all smart autonomous robots, or individual fund for each and every robot.
- Algorithmic Impact Assessments under the GDPR (the purpose and functionalities of this AIA model is to remain sustainable and law-, policy- and technology- responsive).
- *US Algorithmic Accountability Act of 2019* (HR 2231) (though will likely face political opposition from industry).
- Proposal for US Federal Trade Commission to regulate robotics (FTC already regulates a wide variety of businesses and if supported by policy – adapt its regulatory approach to new technologies and new issues).
- Register of algorithms used in government (if it remains under review and is kept updated and can draw adequate funding).
- US Food and Drug Administration regulation of adaptive AI/ML technology (adapts a current regulatory approach to the more flexible requirements of AI/ML medical systems).
- *Directive on Automated Decision-Making* (though will require vendors to provide more algorithmic accountability than is currently required).
- Centre for Data Ethics and Innovation (already recently established and includes embedded mechanisms to ensure that it remains sustainable and operational and that its work remains relevant).

- Regulatory sandboxes (already in use in over 20 countries[126] to test regulation of new technologies).[127]

Non-feasible options

Options that were identified as non-feasible, or have drawn criticism and are potentially most likely to be affected by future developments, include the proposal for anti-trust regulations (slated for the non-feasibility of tech breakups and unwindings, and the potential for weakening anti-trust enforcement if this fails), and the proposal for the US *DEEP FAKES Accountability Act* (criticised as not feasible or recommendable).

Dependencies

In some cases a judgment on feasibility or sustainability was not made by the researcher , but conditionalities for feasibility and sustainability were outlined. E.g., in the case of the *Moratorium on the development of 'lethal autonomous robotics' (LARs)/offensive LAWS*, it would depend on the type of moratorium implemented (as they are susceptible to policy changes (e.g., where new policy determine these are counter-productive to innovation, or economic prosperity).

The feasibility, sustainability and future-proof character of the legislative framework for independent and effective oversight over human rights compliance would require an assessment of the specific impact, budget and policy assessments in each Member State. Redress-by-design mechanisms for AI are future-proof to the extent that they can easily adapt/align with societal values as they change. The feasibility and sustainability of the EU Taskforce of field specific regulators for AI/big data depend on internal and external buy-in, and EU political will to create such a taskforce/and if created to keep it going. The feasibility and sustainability of the proposal for voluntary/mandatory certification of algorithmic decision systems (ADS) would depend on sustained efforts/support from governments/public sector to incentivise its creation and then effective use. The proposed (new) FDA for algorithms is susceptible to policy changes (e.g., deregulation) and the restriction of its powers by changes to policy/legislation. The sustainability of the Digital Authority would depend on the policy and funding model adopted and its usefulness in regulating the digital world. In the case of the three-level obligatory impact assessments for new technologies, it would depend on the political will to put this into action and institutional resistance and/or buy-in, and their connections with other obligatory requirements such as data protection impact assessments under the GDPR.

For some of the studied options, this question was found to be unanswerable or remains to be determined as there is not enough data, or the proposals have not been sufficiently defined to make an adequate assessment, e.g., *Binding Framework Convention to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law* (Council of Europe) including through a new ad hoc committee on AI (CAHAI); EU-level special list of robot rights; mandatory consumer protection impact assessment; creating electronic personhood status for autonomous systems; *Convention on human rights in the robot age*; legal framework for human rights impact assessments (HRIAs); new statutory duty of care for online harms.

Factors

[126] Jenik, Ivo and Kate Lauer, "Regulatory Sandboxes and Financial Inclusion", CGAP Working Paper, October 2017. https://www.cgap.org/sites/default/files/Working-Paper-Regulatory-Sandboxes-Oct-2017.pdf

[127] One stakeholder board member pointed out that given AI and big data are such novel fields, this "trial-by-error approach" provides for agility.

The assessment also brought to light the various **factors upon which feasibility and sustainability** would depend and that need to be considered in the adoption and/or use of the studied options. These include:

- Ability to adapt to reflect technological developments/developments in AI/big data/ICT
- Buy-in/Support (trust)/opposition from other stakeholders such as industry
- Ability to meet changing societal needs over time
- Ability to accommodate changes in societal expectations
- Use of informal governance techniques to support its growth and ability to exist
- Attraction of funds and human resources
- Ability to deliver good outputs
- Sound management of acquired funding
- Ability to respond to reflect the legal idiosyncrasies of each Member State and respond to the policy and technological needs and priorities
- Regular reviews and updates
- Market incentives and forces
- Competing priorities and conflicts (e.g., of the bodies housing the new body/forming it)
- Strength of underpinning (technical or regulatory) framework
- Whether it is able to make a positive impact and complement other measures
- Success in drawing a line between what is in scope and out of its scope.

4.2.14 Impact on innovation

This section presents the findings of how the studied options will potentially adversely affect the ability of businesses and others to innovate. This assessment relates to the options are they have been elaborated or in application at the time this research was carried out (noting that they are currently at various levels of maturity).

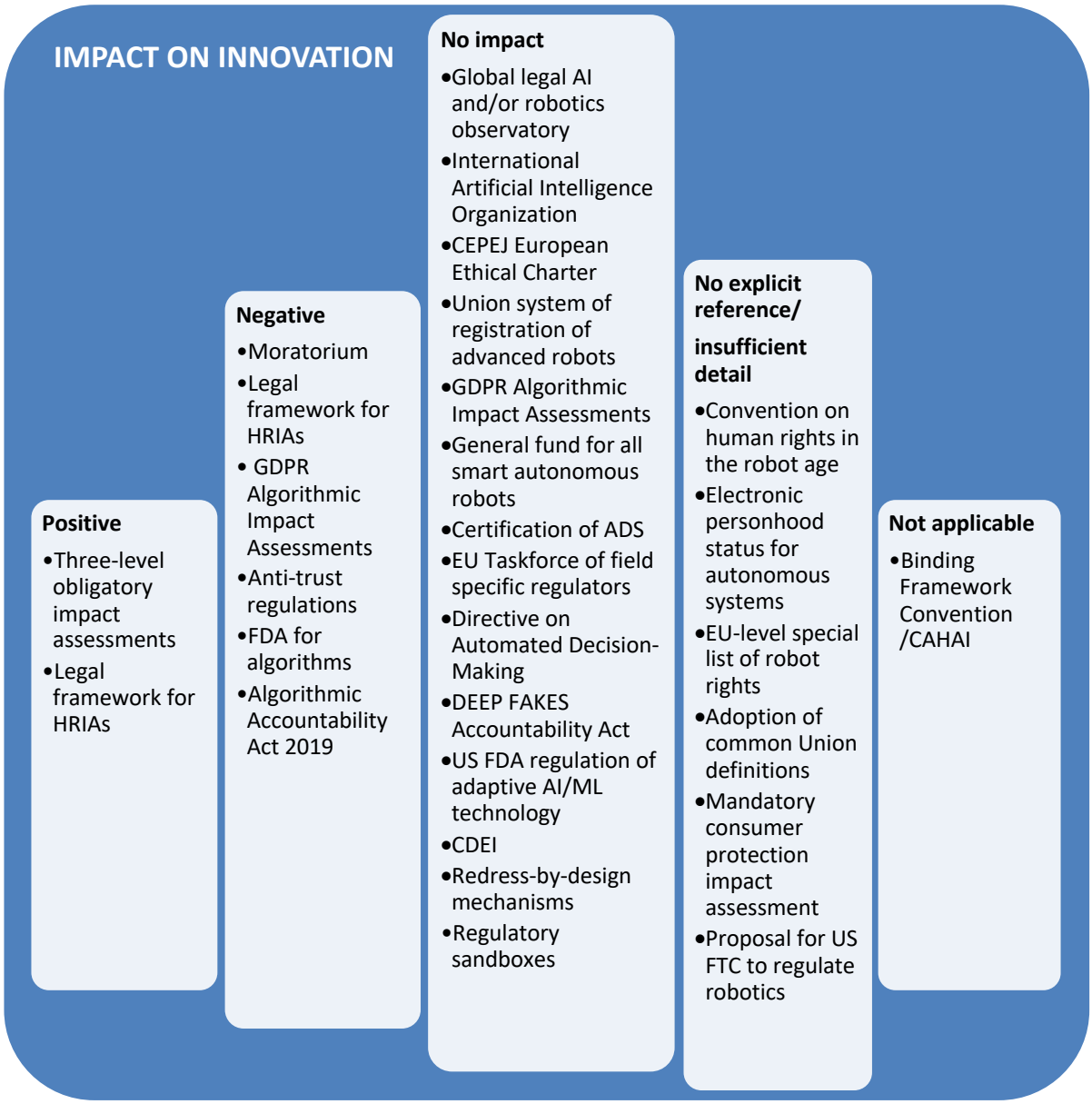

Fig 6: Impact on innovation

In addition, it was anticipated there would be an impact on innovations depending on the fulfilment of certain conditions, e.g., the ability of the legislative framework for independent and effective oversight over human rights to impact the ability of businesses to innovate would depend on the statutory powers and missions of the bodies. In the case of the UK Digital Authority, impacts would be felt if it became over-prescriptive. If the register of algorithms used in government was implemented for businesses, it could stifle innovation, creativity and financial prosperity and could have adverse effects.

4.2.15 Suitability/fit with EU legal framework

This section briefly presents information on the suitability/fit of the reviewed options with the EU legal framework, including the powers and competences of the EU to implement such options in accordance with the EU acquis.

International-level options

At the international level, the proposed options fit well with the EU legal framework and show promise for implementation at the EU-level (e.g., the LARs/LAWS moratorium could be actualised via an EU Parliament Resolution or Council Decision). Implementation would depend on the EU AI regulatory strategy and whether any established applicable criteria are met (e.g., IAIO might have to meet criteria for informal governance arrangements). The *Binding Framework Convention* could be embedded into secondary legislation. The legislative framework for oversight bodies could be adopted at the EU or national level (if the EU does not choose to act). Options such as the legal framework for HRIAs could help EU Member States comply and meet their positive and procedural obligations under the *European Convention on Human Rights*. The CEPEJ European *Ethical Charter* is line with the fundamental rights guaranteed in the *European Convention on Human Rights* (ECHR) and the Council of Europe *Convention on the Protection of Personal Data*.

EU-level options

The EU-level options generally fit with the EU legal framework (existing[128] or new[129]). E.g., the EU-level special list of robot rights might be similar to Council Directive 98/58/EC *Concerning the Protection of Animals Kept for Farming Purposes* - it could define rights for non-human (but sentient) creatures used by humans. The proposal to establish a comprehensive Union system of registration of advanced robots is consistent with existing registration systems in the EU. The EU Taskforce of field specific regulators for AI/big data could be modelled on existing task forces.[130]

The proposal to create electronic personhood status for autonomous systems is considered unsuitable for the EU-level as the competence to determine who is a natural person and legal personhood are a national law competence, for the time being.

National level

The national level options are designed for implementation at that level.[131] Some of the reviewed options (seven) are from outside the EU; they are significant nonetheless and present models for future regulatory actions (legislation, new bodies[132], or other mechanisms) at the national level.

Some of the options generally fit with the EU legal framework (by extended application) or could be introduced via new legislation if deemed appropriate for Union-level action, e.g., redress by design mechanisms (though there might not be the political appetite for this), or they could be introduced

[128] Mandatory consumer protection impact assessment fits within the goals of the EU *Consumer Protection Directive*, which seeks to achieve a high level of consumer protection across the EU; Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations; voluntary/mandatory certification of algorithmic decision systems (ADS) could be based, e.g., on Article 42 of the *General Data Protection Regulation* or other sectoral legislation.

[129] General fund for all smart autonomous robots fits well within the EU legal framework. A Directive (akin to Directive 2009/103/EC) could be the way forward if deemed appropriate.

[130] E.g., Financial Action Task Force (FATF) [first closed, then open-ended]; Taskforce on Article 50 negotiations with the United Kingdom; Advanced manufacturing Task Force; Smart Grids Task Force;

[131] E.g., *DEEP FAKES Accountability Act 2019*. While a related communication has been issued by the European Commission, the competence to legislate with respect to deep fakes is better placed at the national EU Member State level (though the cross-border dimensions might make a European approach necessary for effective and coordinated action and to protect the EU, its citizens, its policies and its Institutions, as outlined in the Communication). Also, the US Federal Trade Commission to regulate robotics - most unfair trade practice legislation is at the Member State level; US Food and Drug Administration regulation of adaptive AI/ML technology.

[132] The FDA for algorithms.

into an existing Regulation via revision process. (e.g., rebuilding the GDPR for AI). The new statutory duty of care for online harms (UK) is compatible with the EU's e-Commerce Directive, which requires online service providers to act on illegal user-generated content (UGC) once they have been notified of or become aware of its existence (though there is also the potential for the new regulation to conflict with users' rights to privacy and free expression).

Some of the reviewed options could be stated to fit within the framework of the GDPR, e.g., if the EU chose to add additional protections for automated decision-making (Canada *Directive on Automated Decision-Making*). The US *Algorithmic Accountability Act of 2019* also might be suitable for adaptation within this framework, but as written, but it is not fully consistent with the GDPR requirements.

A national independent cross-sector advisory body such as the CDEI could promote the goals of EU legislation and provide a model for other countries. The Digital Authority too envisages liaison with European and international bodies responsible for internet regulation. The register of algorithms used in government is not likely to be incompatible with the EU legal framework but will require a legal basis at the EU-level for the establishment of an overseeing European or national agency to maintain the register.

Cross-over options

With regard to the anti-trust regulations to break up big tech, the EU might not have direct jurisdiction on the matter of breaking up big tech companies. However, some other actions could be envisaged through competition law.[133]

Three-level obligatory impact assessments for new technologies are a good fit with the EU legal framework, in as much as it would help Member States comply with the need to apply measures to protect the rule of law, democracy and human rights against AI-based infringements.

The regulatory sandbox mechanism can fit within an existing regulatory mechanism at any level: EU, Member State, or local.

4.2.16 Chances of success

Based on the research and assessments of the options (question 21) (and in a few cases, along with the stakeholders consulted[134]), this section presents the results on the question of how likely the options are to succeed. We note this assessment might be highly subjective and reflects the views of the research team based on the information available to them on the options at the time of research and their considered judgment of the option.[135] There is potential for these scores to be re-visited as the options further develop or they are affected, for example, by legislative and/or policy changes, technological advancements, industry or public support/rejection.

Key: 5 - Extremely likely; 4 - likely; 3 - neutral; 2- unlikely; 1 - extremely unlikely.

[133] E.g., disempowering through fines or mandating that some activities must be blocked as illegal. The EU and/or Member States could find business activity unlawful and pause the company.
[134] See Annex 1 – seven options received feedback.
[135] No specific metrics were applied.

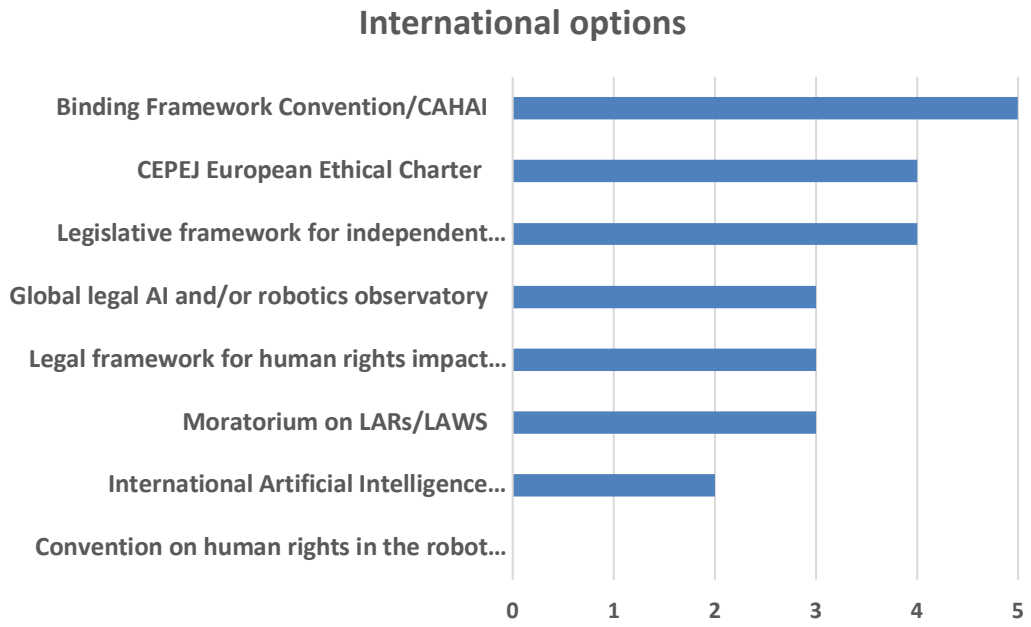

Fig 7: International options: scores

For the international options, the three options that look most promising are: the *Binding Framework Convention*, the CEPEJ European *Ethical Charter* and the Legislative Framework for independent and effective oversight.

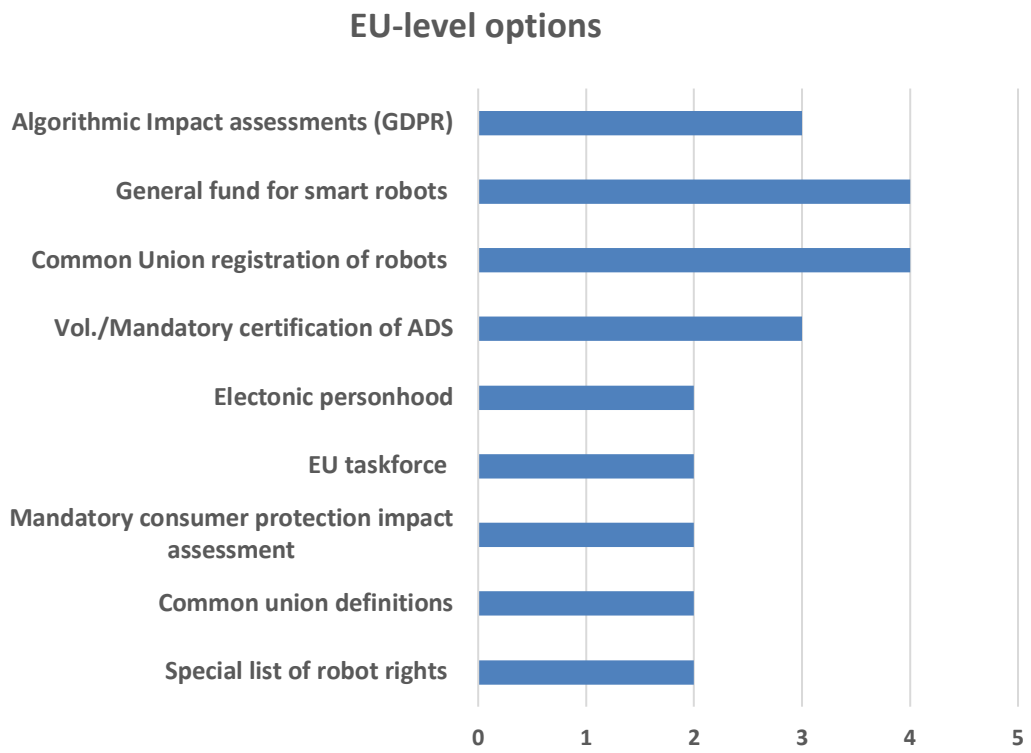

Fig 8: EU-level options: scores

At the EU-level, the general fund for smart robots and the Common Union registration of robots fared extremely well; with AIA's under the GDPR and voluntary/mandatory certification of ADS not far behind.

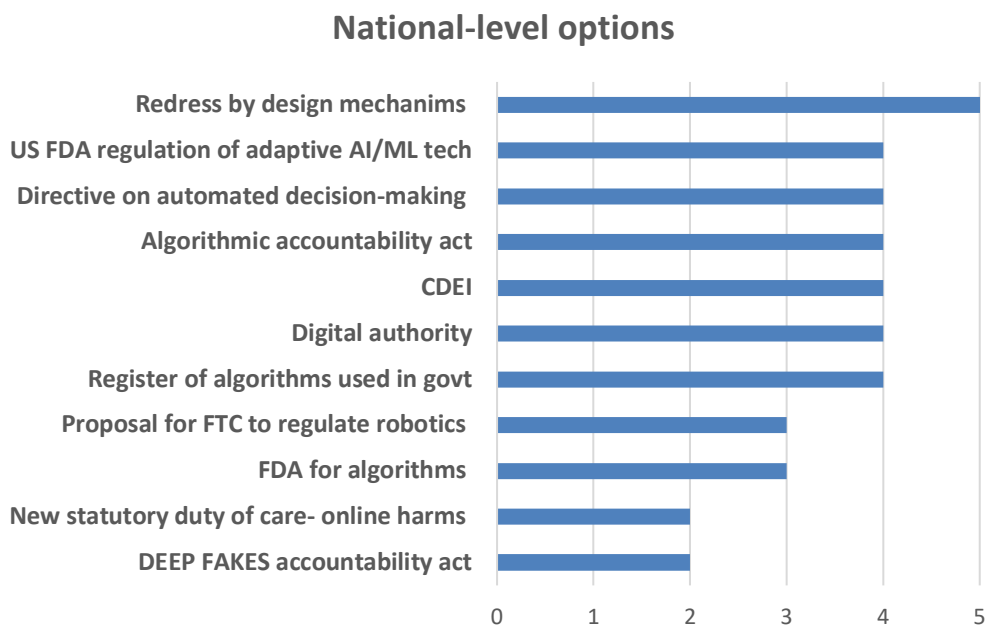

Fig 9: National options: scores

At the national level, the most promising was redress by design, followed by proposed 'specific' legislation. Bodies such as the CDEI and Digital Authority also look promising, as does the proposal for a register of algorithms used in government.

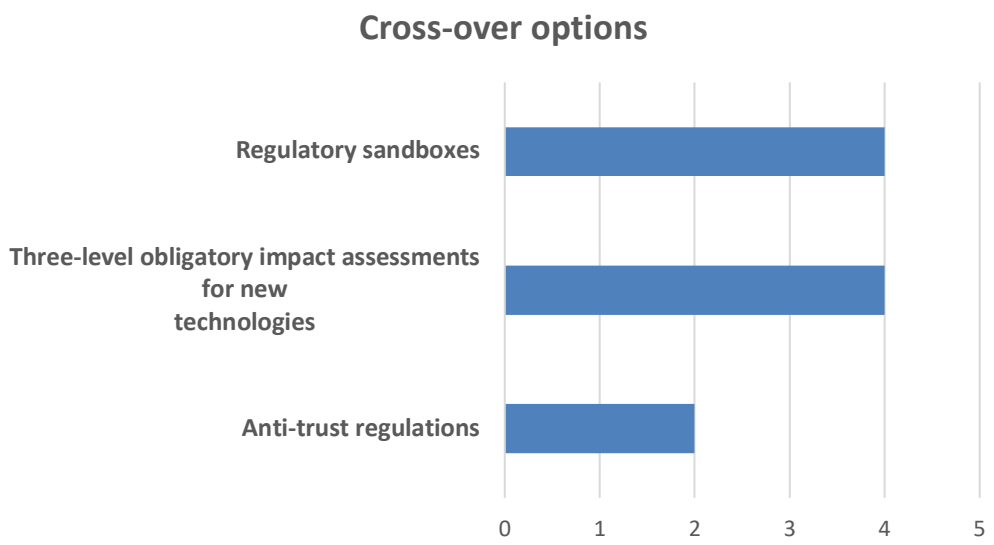

Fig 10: Cross-over options: scores

The general/cross-over options ranged from unlikely to likely.

4.2.17 Overall conclusion (factors critical to adoption and/or success)

The assessment of the options also looked at factors critical to their adoption and/or success. This is covered in detail in responses to question 22 in the individual assessments in Annex 4. Below we present some key findings.

International level options

For a moratorium on LARs/LAWS to succeed, it would be best placed as a **global** measure.

The success of a legal framework for HRIAs for the public sector will depend on **how well the framework and procedures are set out**, whether there is a specification of when they should be carried out, what **incentives** are offered for their use/penalties set for non-compliance, what comes within their scope, what are the key requirements, when it should be carried out, who should be involved and what the process should be.

The *Binding Framework Convention* would require **consensus** in this field, before being in a position to draw up a legal framework, which, unless **adopted as an equivalent at the EU-level**, will lack legal enforcement in countries.

Establishing an AI oversight body necessitates **legislative amendments, impact, policy and financial assessments** and advanced planning.

With respect to the IAIO, as AI is a very sensitive issue area, achieving **political consensus** to transform the IAIO or an equivalent organization into a more formal entity will likely be a persistent problem.

A global legal observatory will need to be able to c**apture high quality data, present good analysis and interpretation** of the information collected (if this is within scope) and be able to disseminate and report its results well.

EU-level options

An EU-level special list of robot rights currently faces opposition from several quarters, and a proposal to create qualified rights for conscious autonomous systems is highly unlikely to be viable until there is a **broader acceptance** of the idea of machine consciousness, or autonomous systems are deployed more widely (especially in non-industrial settings).

The adoption of common Union definitions of cyber physical systems, autonomous systems, and smart autonomous robots will need to **consider existing definitions in use** by standards-setting organisations, **focus on functions and capabilities** rather than specific mechanisms, and limit references to specific examples of existing technology.

Any discussion of a special legal status for autonomous systems needs to **clearly differentiate between legal agenthood in contracts and business law** (e.g. special status for ascertaining accountability, liability and responsibility), and the **broader human or constitutional rights** for autonomous systems, as outlined by Prof. Ugo Pagallo.[136]

Implementation of a registration system of advanced robots would need a **responsible coordinating entity,** clear definition of the types of systems that must be registered, mechanisms to review and update the definition as technology evolves, defined information to be provided with the registration, rules for how the registered information will be used and disclosed, and **consequences for non-compliance**.

[136] Pagallo, Ugo, "Vital, Sophia, and Co.—The Quest for the Legal Personhood of Robots", *Information*, 9(9), 2018, pp. 230. https://www.mdpi.com/2078-2489/9/9/230

One key factor critical to the adoption of the general fund for smart autonomous robots is having a **good legal framework** as its basis.

Factors critical to the adoption of mandatory consumer protection impact assessments include **testing** of such a proposal as a non-mandatory tool and buy-in by regulators.

Factors critical to the adoption of the EU Taskforce of field specific regulators for AI/big data include the **political will to adopt** and specify its responsibilities (other regulatory agencies might not be willing to relinquish their control/or cooperate); whether it can truly be harnessed to develop/support the adoption of high-quality regulation in AI/big data without fuelling further a race to the bottom. Factors for its success include whether the task force is able to successfully carry out its designated functions by **not being bogged down in red tape.**

The adoption and implementation of AIAs under the GDPR are contingent on the **approval and confirmation of European and national authorities** that this approach fits within the GDPR or falls under a new legislative requirement.

Factors critical to the success of voluntary/mandatory certification of algorithmic decision systems (ADS) include ensuring there is **no potential for misuse** of the certification scheme (e.g., misrepresentation, fraudulent representation of certification, free riding, conflicts of interest e.g., certification of subscribers from whose subscriptions the certifier profits).

National level

One factor critical to the success of the US *DEEP FAKES Accountability Act* will be whether the law is able to surmount the criticism that it only **addresses the** symptom and not the **cause** of the problem.

Industry opposition to the US *Algorithmic Accountability Act of 2019* would be lessened if the FTC provides: **guidelines** to help determine whether an application is high-risk, and tools and guidelines to help with the algorithmic assessment.

The success of the Canadian *Directive on Automated Decision-Making* lies in providing **easy-to-use tools** for the risk assessment and algorithmic impact assessment.

One factor that might contribute to the adoption and success of redress by design mechanisms for AI is **a 'crisis' which will make people more aware and force action.** Consumer associations' and civil society organisations' support could boost the adoption of such mechanisms.

Establishing an obligation to keep a register of algorithms requires the **identification and establishment of the responsible agency** for this, the extent of this obligation (e.g., private or public sector), and the specific materialisation of this obligation (e.g. what types of information should be recorded about the uses of algorithms).

The effectiveness of the Digital Authority will depend on **proper funding, ability to coordinate and instruct different regulators, ability to remain politically impartial** and independent of the Government, and democratic scrutiny.

An FDA for algorithms would need a depth of technical know-how, and a rich **diversity of expertise** to grasp the breadth of society; it would also need **distinct trigger points** on when to review and at what level of scrutiny, as pointed out by Groth, Nitzberg and Russell.[137]

With respect to the CDEI, it has been pointed out that a bespoke system may not be that effective if it is not part of a **uniform regulatory approach** to AI on a supra-national level.

The proposal for a US Federal Trade Commission to regulate robotics is primarily a political choice and will require **political will** to be adopted and implemented.

Cross-over options

The factors critical to the success of anti-trust regulations are whether other legislative and regulative tools are **able to** address and/or **redress the concerns of data power imbalances** and whether further harms to consumer welfare result.

Factors critical to the adoption/success of three-level obligatory impact assessments include a **strong governance framework, stakeholder buy-in,** and **transparency** that facilitates some form of external oversight and review; this is in addition to their clear placement and connection with existing legislation and other forms of impact assessment.

For regulatory sandboxes, factors critical to success include **thoughtful design** of the sandbox parameters, **transparency** in the design, operation and outcomes, and **close communication and cooperation** with stakeholders.

4.3 How AI challenges regulation and EU aspirations for better law-making

AI challenges regulation, and these challenges should be duly recognised and addressed when regulating for AI. At the very core, the definition and conceptualisation of AI itself is an issue (as highlighted in the regulatory options study)[138]. As one of our stakeholder board members pointed out, "Can one really claim to draw a bright line around the concept of AI, so that anyone could agree what's in and what's out?". [139] AI challenges the different forms of regulation, like other technologies do, as highlighted below:

Forms of regulation	How AI challenges it
Ex ante (before an event occurs)	Scherer: "difficult and impracticable because AI research and development may be discreet (requiring little physical infrastructure), discrete (different components of an AI system may be designed without conscious coordination), diffuse (dozens of individuals in widely dispersed geographic

[137] Groth, Olaf J., Mark J. Nitzberg, Stuart J. Russell, "AI Algorithms Need FDA-Style Drug Trials", *WIRED opinion,* 15 August 2019. https://www.wired.com/story/ai-algorithms-need-drug-trials/#

[138] See section 4.2.6 of this report.

[139] Stakeholder Board Member correspondence, 13 December 2019. Further, it was questioned, "are there generalizations that could apply to all members of the category? If not, how do we formulate or choose different ethical inspirations for different subsets of cases?". This needs further discussion.

Forms of regulation	How AI challenges it
	locations can participate in an AI project), and opaque (outside observers may not be able to detect potentially harmful features of an AI system)." [140]
Ex post (regulation enacted after harm caused following an event)	Scherer: "The autonomous nature of AI creates issues of foreseeability and control that might render ex post regulation ineffective, particularly if an AI system poses a catastrophic risk." [141]
Principles-based (broad guiding principles instead of precise rules in pursuit of desired regulatory outcomes) (PBR)	Leenders: "this approach relies on trust, responsibility and good faith between the regulator and regulated". But "contemporary AI industry is not marked by high levels of trust between innovators and the public. In this environment, the literature asserts that PBR would increase regulatory uncertainty while allowing for sub-optimal regulatory compliance." [142]
Risk-based regulation (RBR) (focus on targeting activities that pose the greatest threats)	Deciding what risks deserves priority; the discretionary selection of risks might be an issue. The focus on some 'great' risks might lead others to be ignored or concealed.[143]
Precautionary-based regulation (focus on pre-emptive remedies that aim to predict the future, and future hypothetical problems)	Thierer: "preemptive bans or highly restrictive regulatory prescriptions can limit innovations that yield new and better ways of doing things." [144] More specifically as Thierer highlights, "There may very well be some serious issues raised by robotics and AI that we cannot ignore, and which may even require a little preemptive, precautionary policy. And the same goes for general computing and the Internet. But that is not a good reason to just create new bureaucracies in the hope that some set of mythical technocratic philosopher kings will ride in to save the day with their supposed greater "expertise" about these matters".[145]

Table 4: How AI challenges different forms of regulation

[140] Scherer, Matthew U. "Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies." *Harv. JL & Tech.* 29, 2015, p. 353.
[141] Ibid.
[142] Leenders, Gijs, "The Regulation of Artificial Intelligence — A Case Study of the Partnership on AI", *Becoming Human: Artificial Intelligence Magazine*, 13 April 2019. https://becominghuman.ai/the-regulation-of-artificial-intelligence-a-case-study-of-the-partnership-on-ai-c1c22526c19f
[143] As clarified by a Stakeholder Board member, when it has been determined that a regulation is needed because the risk must be avoided, prevented, reduced, mitigated, transferred, etc), the regulator has a choice between (a) strict and technical measures (which require a prior detailed, if possible quantified, assessment of the risk, and usually takes the form of prescribed actions or even the prohibition of certain technology). This type of regulation could be performance-based (focused on the goal of improving the current state, without the prescription or prohibition of certain technology, in order to encourage technological progress) and (b) the adoption and respect of certain principles, without the regulation saying much about how to do it. to a large extent the EU GDPR is principle-based, with ex-post 'control' by courts and liability systems.

[144] Thierer, Adam, "Problems with Precautionary Principle-Minded Tech Regulation & a Federal Robotics Commission", *The Technology Liberation Front,* 22 September 2014. *https://techliberation.com/2014/09/22/problems-with-precautionary-principle-minded-tech-regulation-a-federal-robotics-commission/*
[145] Ibid.

Referring to the *Interinstitutional Agreement between the European Parliament, the Council of the European Union and the European Commission on Better Law-Making*,[146] we can distil the following aspirations for Union legislation:

- high-quality
- ensuring that such legislation focuses on areas where it has the greatest added value for European citizens
- is as efficient and effective as possible in delivering the common policy objectives of the Union
- is as simple, comprehensible and as clear as possible
- is consistent
- avoids over-regulation and administrative burdens for citizens, administrations and businesses, especially small and medium-sized enterprises (SMEs)
- is designed with a view to facilitating its transposition and practical application and to strengthening the competitiveness and sustainability of the Union economy
- to legislate only where and to the extent necessary
- is supported by a transparent process
- allows citizens, administrations and businesses to easily understand their rights and obligations
- includes appropriate reporting, monitoring and evaluation requirements
- is practical to implement
- is supported by impact assessments (covers the existence, scale and consequences of a problem and the question of whether or not Union action is needed)
- includes public and stakeholder consultation.

All the above aspirations need bearing in mind in attempts to regulate AI, though we recognise there are limitations e.g., "what is efficient cannot ipso facto be held just, and so, acceptable", as pointed out by our Stakeholder Board Member.[147] In our assessment of the regulatory options, we evaluated the options against some of these aspirations,[148] e.g., whether the options are sufficiently clear, specific and able to be effectively and efficiently operationalised (connected to the better law-making aspirations of simplicity, comprehensibility and clarity). We also discussed the implementation burdens (e.g., administrative or other burdens) the options might create for citizens, public administrations, businesses and particularly SMEs. Further, the assessment also shed light on whether the options had explicit monitoring, oversight and enforcement mechanisms, gaps, provisions for review and update. However, given the limited scope of this study, we underline the need for thorough and detailed impact assessments (using this study as input where desirable and including in all cases adequate levels of public and stakeholder consultation).

In particular, the aspirations for better law-making call us to avoid overregulation and to legislate only where and to the extent necessary, and need to be seriously discussed given the multi-level calls and championing of the regulation of AI and big data. The appetite for (hard-hitting) legal regulation is steadily growing (internationally, regionally - especially at the EU-level, and in some countries) with respect to AI (though caution is also advised), and the regulatory state of play will change significantly over the next five years, especially at the EU-level and in the US (which as evident in this study is already fast-advancing with AI-specific legislative proposals). The danger with this is it might cause effects that are unintended in a field that is very dynamic and fast-developing.

[146] *OJ L* 123, 12.5.2016, p. 1–14.
[147] Stakeholder Board member input, 13 December 2019.
[148] The options studied were at different levels (international, EU-level and national) so we did not use all the criteria, also some of these were not used after internal discussions, testing and scoping paper feedback.

As seen in Section 3, there are already **concerns about over-regulation** from various quarters (policymakers, think tanks, media, and unsurprisingly industry, which actively lobbies for minimal regulation). Excessive (and unjustified) regulation could 'kill' in some cases (e.g., where outright bans/prohibitions on AI applications are used) or seriously maim the golden AI innovation goose (something that was highlighted in relation to the anti-trust regulatory option[149]). It might also lead to AI innovators seeking out AI-safe havens (in countries with dubious ethics and human rights credentials), or reduce the potential benefits that could be gained from its development and use. This could stem from or be fuelled further by knee-jerk political responses (e.g., in response to AI-harms frenzies whipped up irresponsibly by the media), lack of proper impact assessment of measures sought to be implemented, haphazard stakeholder consultation, and, most dangerously, a 'regulate first ask questions later' culture (which might be justified with respect to certain high-risk applications like lethal autonomous weapons systems, but definitely not in low-risk cases or where technical solutions, or setting of standards might be better placed to address concerns).

A 2018 Royal Society report states the following, "Exaggerated expectations and fears about AI, together with an over-emphasis on humanoid representations, can affect public confidence and perceptions. They may contribute to **misinformed debate, with potentially significant consequences for AI research, funding, regulation and reception**." The report further outlines that bad regulation is a potential consequence of the disconnect from the "reality of the technology". This caution should be well-received when seeking to regulate AI. Further, the report states:

> False expectations can mean that a sector is allowed to grow without further intervention by governments, such as providing supportive regulation and market structures. As a result, a sector might grow slowly, reducing potential benefit. Or, it might grow fast, but in ways that are not aligned with social values, or in ways that lead to a bubble that will cause harm when it bursts. False fears, meanwhile, can lead to either over-regulation that suffocates growth and innovation, or to spending significant time and other resources on regulating something that will not require such regulation.[150]

Petit outlines the example of a deficient AI airliner pilot that causes an accident and shows how "the potential for knee-jerk regulation of AIs and robotics is easy to foresee", but suggests that "instead of prohibiting, regulation should seek to improve machine-human cooperation in ways that enhance safety."[151] This is often under-addressed in discussions on regulation of AI.

A number of other **challenges, specific to AI-based systems**, need to be taken into account when considering regulatory initiatives targeting AI. The issue of bias in training data and trained models is often complicated and confusing. As bias may lead to undesirable, or even illegal, discrimination, it appears a natural regulatory goal to insist on reducing bias. It is, however, important to understand that while we definitely do not want to see AI models amplifying bias in their training data, artificially removing bias may produce models that simply do not reflect the reality, resulting in AI systems with poor predictive capabilities. Another challenge that requires consideration is behavioural fluidity of AI-based systems, especially those which are trained regularly or near-continuously. Unlike traditional cyber systems, assumptions about such AI systems, based on their testing and validation results, may be highly unreliable, and special care is required when using them or integrating with other technologies. Dependence of AI algorithms and models on potentially untrustworthy data and

[149] See Watney, Caleb, "A Framework for Increasing Competition and Diffusion in Artificial Intelligence", *R Street,* 21 March 2019. https://www.rstreet.org/2019/03/21/a-framework-for-increasing-competition-and-diffusion-in-artificial-intelligence/

[150] The Royal Society, op.cit, 2018.

[151] Petit, N., "Law and regulation of artificial intelligence and robots: Conceptual framework and normative implications", Working Paper, 2017, p. 30 https://dx.doi.org/10.2139/ssrn.2931339

challenges related to their testing and validation make it hard to detect and analyze flaws and vulnerabilities in AI-based systems and often significantly broaden their attack surface. To ensure effectiveness and viability of AI-related regulation, proper attention should be paid to the peculiarities of AI systems, including the challenges and issues mentioned above.

When to regulate

It is also important to consider WHEN to regulate AI and big data; this will contribute to the effectiveness and meaningfulness of AI regulation. Table 5 below illustrates how this might work at different stages of the AI application/system lifecycle:

AI lifecycle stage	What might be regulated	How	Pros	Cons
Before research and development R&D	The exploration of R&D ideas into the tech itself, e.g., AI systems making life-or-death decisions	E.g., Prohibition/Ban/Moratorium	Takes into account dangerous and harmful threats to society. Helps guard against undesirable technologies, uncontrolled effects and wasteful investment.	Not very feasible – too early. Stunts innovation Restricts freedom of speech and expression, freedom of the arts and sciences, intellectual creativity and property.
Research, design and development	E.g., Design of machine learning systems, automated warfare systems	E.g., Tech/field specific legislation; mandatory standard; regulatory oversight body inspections; mandatory impact assessments; patent restrictions; registration requirements	Compliance with legal and human rights requirements. Creates incentives for private actors to internalize the costs of their behaviour. Fosters responsible research and innovation.	Restricts research and development. Increases costs of R&D (compliance costs). Added regulatory burdens. Could create competitive disadvantages for SMEs lacking resources.
Production	E.g., AI systems, products	E.g., Legislation (consumer safety); mandatory standard/conformity assessment. Mandatory registration and classification. Mandatory algorithmic impact assessment	Protection of consumer welfare. Assuring high standards. Facilitating transparency and accountability. Legal certainty. Compliance with legal standards. Enhances public trust in AI.	Compliance and resource costs. If not combined with previous controls, these measures could find the AI application/system incompatible with the legal framework or unlawful.
Piloting and/or testing	E.g., introduction of testing pilot for automated vehicles	Pilot testing legislation; safety regulations	Helps determine whether risks and dangers are adequately considered. Controls risks and adverse effects.	If issues do arise, and these are un-addressable, the product/system might be considered a failure and unable to be put on marker or might have to

AI lifecycle stage	What might be regulated	How	Pros	Cons
				revert to redevelopment.
Commissioning/ procurement	AI, data analytics solutions and services	E.g. Directives, regulations, policies, guidance relating to the procurement of AI systems, products, services for the public sector e.g., UK *Draft Guidelines for AI procurement* (2019)	Stringent regulation will shape adoption. Guard against undesirable technology use and implementation.	Will inhibit timely adoption. Risk of inconsistencies if the public and private sector do not adhere to the same criteria.
Implementation and use	Use of AI, data analytics solutions and services	E.g., existing law (data protection, privacy); new tech/field-specific laws and regulations.	Regulates at point of use.	Risk of undesirable aspects has already been added in. Potential failure to regulate if the existing framework may not be applicable, enforceable or effective.[152] Amplification of risks and impacts. Enforcement hurdles due to the entrenchment of social/technical norms. May need to implement costly redress mechanisms and lead to costly litigation.

Table 5: When to regulate

Table 5 represents a model for effective, targeted, market-responsive and tailored regulation to specific AI applications and AI in general. In addition to the requirement for diverse controls and measures (e.g., soft and hard law solutions), a multi-layered approach to the time-point of regulatory intervention is also required. In particular, the proven inadequate framework to address AI challenges and the alleged risk of uncontrollable, unknown and opaque risks and consequences of AI require ongoing and 'incremental' regulation. A regulatory framework permeating all the stages of AI deployment from the early intellectual conceptualisation of the AI potential to its actual implementation could e.g.,:

- Foster a culture of precaution (market participants) and safe use (individuals).
- Increase general awareness of the risks and impacts of AI.
- Monitor and control AI applications.
- Institutionalise/normalise regulatory control and oversight (from top to bottom - from governments--> administration--> professional bodies--> individuals) contrary to 'sporadic' ad hoc control/law.

[152] E.g., different material scope, lack of provisions for legal claims, different purposes and objectives.

- Establish safeguards and mechanisms for emerging risks and abuses ('risk alarms').
- Enable the implementation of privacy by design, data protection by design and default, ethics by design, human rights impact assessments, algorithmic impact assessments.
- Make sure that legal loopholes are addressed in a timely manner and at the right level. Otherwise, where the law is outdated, there is a risk that a disproportionate burden is imposed on market participants to act lawfully in an opaque environment and the judiciary to judge without legal tools and training. Moreover, this could prevent the risk of chilling effects, ill-applied analogy or uncontrolled AI development.

Regulatory prudence

Based on the EU aspirations for better law-making, what we recommend is greater wisdom to be applied with regard to regulatory decision-making in the AI context. We call for **regulatory serenity** to accept the things that cannot change (e.g., that AI applications and uses will advance and morph at a pace that is faster than the law and change the way we live and interact), are beneficial to society, and acknowledge and find a way to address residual risks; the courage to change the things that can (e.g., certain uses/applications of AI should be prohibited; secure, ethical, human rights-respectful and responsible AI should be facilitated and incentivised; mandating humans in the loop, ethics and/or human rights by design), and wisdom to know the difference (especially where a different and more effective solution than the letter and sole application of the law might be called for, e.g., use of technical solutions to ensure cybersecurity of AI). Any regulatory measures adopted need to be proportionate, practical and effective.

One promising recently mooted approach by NESTA is 'anticipatory regulation'[153]; an approach to regulation that provides a set of behaviours and tools – i.e., a way of working – that is intended to help regulators identify, build and test solutions to emerging challenges. It has six principles: inclusive and collaborative; future facing; proactive; iterative; outcomes-based; experimental. This approach, as Armstrong et al. outline, is "more forward-facing than either advisory or adaptive approaches, meaning regulators have to deal with more uncertainty, less evidence and a greater number of possible risks".[154] It might be useful to evaluate and test the value of such an approach in regulating AI. Options such as regulatory sandboxes fit into this approach.

[153] Armstrong, Harry, Chris Gorst and Jen Rae, "Renewing regulation 'Anticipatory regulation' in an age of disruption", March 2019. https://media.nesta.org.uk/documents/Renewing_regulation_v3.pdf
[154] Armstrong, Gorst and Rae, op. cit., 2019.

5. Conclusion: key considerations for regulating AI and big data

Section 3 revealed how the regulation of AI and big data is an arena fraught with divergence and disagreements (with good reason); some prefer a heavier touch (which will better support human rights, and reduce development on unacceptable AI/wrongful use), others a lighter touch (fearing the effects of hard regulation on R&D). Proposals for regulations do address ethical concerns and human rights (though human rights are not always the direct and/or explicit objective, material scope, or aim of the proposals, but are indirectly affected, e.g. through anti-trust legislation). However, where there is great variation to the specificity of such proposals, it has been highlighted that it is important to recognize that some challenges require prioritization, in the sense that different values or goals sometimes directly or indirectly contradict each other, and in such cases we require detailed analysis.

Section 4 looked at various regulatory options – i.e., proposals for regulating AI and big data and analysed 31 of these at the national, European and/or international level. Amongst the various categories analysed, the report also identified key factors critical to the adoption/success of the proposals. The report also considered how AI challenges regulation and how timing it right is important (though challenging in itself).

Based on the research in Sections 3 and 4, we further present some key considerations to address challenges in regulating AI and big data.

Striking a balance between enabling beneficial AI and risk mitigation

The value and benefits of AI technologies (social and/or economic) is not in doubt. In the vision of the AI HLEG *Ethics Guidelines for Trustworthy AI,* beneficial AI is that which benefits human beings including future generations; such AI is sustainable, environmentally and socially responsible, taking into account its impact on society.

Given the capacity for AI to pose great risks and harms (e.g., LARs/LAWS) and have serious impacts on human rights and society, there is an urgent need to mitigate such risk impacts head on, using the best possible means (whether technical, standards, or ethical or legal means, either exclusively or in combination depending on the context and the state of the art in addressing the risks early-on and mitigating any negative impacts).

Striking a balance between enabling beneficial technologies and risk mitigation is complex and might not always be possible. AI innovators need to be free to innovate, but at the same time total freedom might lead to irresponsible innovation. This requires policymakers and legislators not only to understand the differential nature of AI and big data risks (e.g., some AI will introduce new risks, some will amplify existing risks, and some will in effect help to reduce/address existing risks). It also requires an **understanding of how AI actors will respond to the regulatory actions** and incentives. This will lead to a more refined understanding of whether a hard regulatory, middle path or soft regulatory approach would benefit the AI technology/application/sector. More importantly, the **possibility of regulatory failure should also be considered** – amplification of risks due to reckless or casual and unconsidered adoption of laws to regulate AI or even the adoption of bad AI law (e.g., a law facilitating AI-based discrimination and/or surveillance). Further, "overregulated societies will miss out on the massive growth technology brings", and we should consider a "harmonious evolution of legislation

alongside technology."[155] Striking a balance is important where possible, for the wider public in terms of societal benefits and/or risks, impacts on future generations and the consequences of AI regulation on other jurisdictions and markets.

Smart mixing for good results

As highlighted by one of our Stakeholder Board members, none of the examined proposals provide the 'silver bullet' to solve all of the ethical and/or human rights challenges of AI. Which (combination of) regulatory options address ethical and/or human rights challenges of AI most effectively may, for example, depend on (the structure of) markets and requirements, and the incentives for market participants to comply with ethical and human rights standards. If market participants take a very proactive stance in this arena (either because of market incentives or not), co- and self-regulation may provide better incentives for these companies to enhance their practices vis-à-vis traditional binding regulatory options[156] (which often adopt standards these companies might already exceed). Conversely, reactive companies may only respond to traditional binding instruments. Furthermore, even for the EU it is a challenge to control and supervise the development and usage of AI outside the EU (with effects within the EU), if the countries of origin have adopted less thorough or even no standards. Only international (more voluntary) regulatory instruments such as certification are able to control this international arena to a certain extent.

Beyond this, traditional regulatory instruments are often less well-equipped to respond to fast technical developments and are often less flexible. Thus, the challenge is not so much to find one effective regulatory instrument, but a smart mix of instruments (i.e., technical, standards, law and ethical) in consultation with stakeholders. This approach seems to offer the flexibility needed to address the challenges of AI. It allows the industry to design self-regulatory tools, or actors to work together with co-regulatory mechanisms or where needed the use of legislation to provide further legal clarity, proscribe harmful AI technologies or provide redress for harms.

For example, in the case of the *predictive policing* scenario,[157] SHERPA envisaged a mix of regulatory and general measures at work: legislation supporting transparent and/or explainable AI; developing algorithms that reduce bias; training of police officers and database operators as to the limitations of data analysis; a national authority overseeing the police use of algorithms; redress mechanisms for harms caused. The SHERPA *AI information warfare* scenario envisaged the use of existing and new laws (though difficulties in agreement were foreseen), citizen education programmes, public tracking and notices of foreign disinformation campaigns, and co-ordinated strategy for countering attacks and cyber defensive counter-measures.[158] In either case, protection of ethical principles and human rights would call not for a single type approach, but a mix of self-binding/voluntary, interventionist, and facilitative regulatory measures. Regulation will also need to be agile.[159]

---

[155] Independent Expert Report, *100 Radical Innovation Breakthroughs for the future*, European Commission, May 2019. https://ec.europa.eu/info/sites/info/files/research_and_innovation/knowledge_publications_tools_and_data/documents/ec_rtd_radical-innovation-breakthrough_052019.pdf

[156] This would allow companies to have a bit more flexibility how to address these questions – and might be more suited to the fast developing environment. [Comment by SHERPA stakeholder board member].

[157] SHERPA, "Scenario: predictive policing in 2025", 2019. https://www.project-sherpa.eu/scenarios/predictive-policing-complete/

[158] SHERPA, Scenario: Information Warfare in 2025, 2019. https://www.project-sherpa.eu/scenarios/warfare-complete/

[159] Wallach and Marchant state that AI and robotics, "present a serious challenge to traditional models of government regulation. These technologies are advancing so quickly that in many sectors, traditional regulation cannot keep up, given the cumbersome procedural and bureaucratic procedures and safeguards that modern legislative and rulemaking processes require. Consequently, regulatory systems will predictively fail to put in

Not just ethics and human rights: we need super-security for high-risk/high-impact AI

One particular point that needs highlighting is the inadequate regulatory attention to AI security. AI security is critical, particularly in terms of reliability and resilience of AI systems to attacks. Some of the studied options do have some focus on security, e.g., *Algorithmic Accountability Act of 2019*; CEPEJ *Ethical Charter*. But on the whole the options studied do not adequately focus on security of AI or seek to enhance this more fundamentally. Given the high risks and stakes (non-obvious/hidden security vulnerabilities or malicious manipulation of AI to cause serious harm) there is **a need to actively discuss and work on regulatory options that support super-secure AI where most needed[160] and put it at the forefront alongside ethics and human rights discussions**.

When it comes to security, especially where there will be **high likelihood and high severity of risk/impact on rights and freedoms of individuals, and especially the vulnerable**, AI applications and systems should be **treated and regulated in a similar or even more enhanced way** to physical risky objects, such as medical devices and planes. Security should be understood broadly in terms of the security features and safeguards of AI as well as its impact on individuals and society. In this context, it is also necessary to ensure that AI applications adhere to security requirements and more widely, standards that protect sustainability and viable solutions for the environment and biodiversity. Indeed, the deployment of AI applications "should be conditional on whether they serve the goal of reducing CO2 emissions and halting the loss of biodiversity".[161]

Security could be seen as a **standalone requirement and regulatory focus** (a regulatory goal itself). But it could be considered as the means to achieve the safeguarding of ethics and human rights law. This might be achieved in many different ways (a combination of additional mandatory technical security measures, security standards and requirements; guidance from ENISA, a specific legal act setting out technical security standards or over-arching security principles for AI or principles for specific AI apps/systems).

What next?

A policy brief taking into account the results of this Study is also planned. SHERPA will use the results of the report to support its Delphi study on ethics and human rights. The report will also feed into planned focus groups in 2020 where stakeholders will discuss regulatory options further. After this, SHERPA recommendations for action by various stakeholders will be finalised.

---

place appropriately tailored regulatory measures by the time new applications of fast-moving technologies begin to affect society. Perhaps even worse, if a regulatory system does somehow manage to rush into place new regulations for an emerging technology, they will likely be obsolete by the time the ink dries on the enactment." Wallach, Wendell, and Gary Marchant, "Toward the Agile and Comprehensive International Governance of AI and Robotics [point of view]," *Proceedings of the IEEE* 107.3, 2019, pp. 505-508.

[160] We recognise super-security might be far reaching in some cases.

[161] IAPP, Privacy 2030: A New Vision for Europe, November 2019, p.22. https://iapp.org/resources/article/privacy-2030/

References

1. "Self-regulation of AI is 'dangerous': CognitiveScale CEO", *Reuters* (VIDEO), 23 January 2019, https://mobile.reuters.com/video/2019/01/23/self-regulation-of-ai-is-dangerous-cogni?videoId=506752793&videoChannel=118156
2. ACM US Public Policy Council, ACM Europe Council Policy Committee, "Statement on Algorithmic Transparency and Accountability", 2017. https://www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf
3. Alarie, Benjamin, Anthony Niblett, and Albert H. Yoon, "Regulation by Machine", 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, http://www.mlandthelaw.org/papers/alarie.pdf
4. Armstrong, Harry, Chris Gorst and Jen Rae, "Renewing regulation 'Anticipatory regulation' in an age of disruption", March 2019. https://media.nesta.org.uk/documents/Renewing_regulation_v3.pdf
5. Arrieta-Ibarra, Imanol, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier and E. Glen Weyl, "Should We Treat Data as Labor? Moving beyond 'Free'", *AEA Papers and Proceedings*, Vol. 108, 2018, pp. 38-42. https://ssrn.com/abstract=3093683
6. Article 19, "About us". https://www.article19.org/about-us/
7. Artificial Intelligence for Development, "A roadmap for artificial intelligence for development in Africa", 8 May 2019. https://ai4d.ai/blog-africa-roadmap/
8. Balkin, J. M., "Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation", *UC Davis Law Review*, Vol. 51, Issue 3, 2018, pp. 1151-1210.
9. BDVA, "Data-driven artificial intelligence for European economic competitiveness and societal progress", BDVA Position Statement, November 2018, p. 8, http://www.bdva.eu/sites/default/files/AI-Position-Statement-BDVA-Final-12112018.pdf
10. Besaw, C. & J. Filitz, "AI & Global Governance: AI in Africa is a Double-Edged Sword", United Nations University Centre for Policy Research. https://cpr.unu.edu/ai-in-africa-is-a-double-edged-sword.html
11. Black, J., "Critical Reflections on Regulation", 27 *Australian Journal of Legal Philosophy,* 2002, 1
12. Black, Julia, and Robert Baldwin, "Really responsive risk-based regulation," *Law & policy,* 32.2, 2010, pp. 181-213
13. Brownsword R. & M. Goodwin, *Law in Context: Law and the Technologies of the Twenty-First Century: Text and Materials,* Cambridge University Press, 2012.
14. Brummer, Christ and Yesha Yadav, "Fintech and the Innovation Trilemma", *The Georgetown Law Journal*, Vol. 107, Issue 2., 2019, pp. 235-307. https://georgetownlawjournal.org/articles/298/fintech-and-the-innovation-trilemma/pdf
15. Cao, S., "It's Serious This Time: Multibillion-Dollar Fines Could Hit Facebook & Google, UK Warns" *Observer*, 28 February 2019. https://observer.com/2019/02/facebook-google-face-multi-billion-dollar-fine-uk-content-regulator/
16. Carrel, P., "Germany must close digital technology gap, Merkel ally says", *Reuters* (TECHNOLOGY NEWS), 8 November 2018. https://www.reuters.com/article/us-germany-tech/germany-must-close-digital-technology-gap-merkel-ally-says-idUSKBN1ND1W4
17. Casey, Bryan, Ashkon Farhangi, and Roland Vogl, "Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise", *Berkeley Technology Law Journal*, Vol. 34, Issue 1, 2019, pp. 143-188
18. Castro, D., "Europe will be left behind if it focuses on ethics and not keeping pace in AI development", *Euronews* (Opinion), 7 August 2019. https://www.euronews.com/2019/08/07/europe-will-be-left-behind-if-it-focuses-on-ethics-and-not-keeping-pace-in-ai-development
19. Chivot, E. and Daniel Castro, "The EU's 'softball' approach to Artificial Intelligence will lose to China's 'hardball'", *Euronews* (Opinion), 5 February 2019. https://www.euronews.com/2019/02/05/the-eu-s-softball-approach-to-artificial-intelligence-will-lose-to-china-s-hardball-view
20. Chung, Jason and Amanda Zink, "Hey Watson – Can I Sue You for Malpractice? Examining the Liability of Artificial Intelligence in Medicine", *Asia Pacific Journal of Health Law & Ethics,* Vol. 11, Issue 2, 2018, pp. 51-80
21. Clarke, Roger. "Regulatory alternatives for AI," *Computer Law & Security Review*, 2019.

22. Coglianese, Cary and David Lehr, "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era", *Faculty Scholarship at Penn Law*, 2017, 1734. http://scholarship.law.upenn.edu/faculty_scholarship/1734
23. Consultative Committee of the Convention for the Protection Of Individuals with regard to Automatic Processing Of Personal Data (Convention 108), Guidelines On Artificial Intelligence and Data Protection**,** Strasbourg, 25 January 2019 T-PD(2019)01.
24. Council of Europe, Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes, Adopted by the Committee of Ministers on 13 February 2019 at the 1337th meeting of the Ministers' Deputies. https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b
25. Council of Europe, Study On The Human Rights Dimensions Of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications as Finalised On 6 October 2017, MSI-NET(2016)06 rev3 FINAL.
26. Cumbley, Richard, and Peter Church, "Is "big data" creepy?", *Computer Law & Security Review* 29.5, 2013, pp. 601-609
27. Dawson, D, E Schleiger, J Horton, J McLaughlin, C Robinson, G Quezada, J Scowcroft and S Hajkowicz, Artificial Intelligence: Australia's Ethics Framework. Data61 CSIRO, Australia, 2019.
28. Delcker, J. "Germany's €3B plan to become an AI powerhouse", *Politico* (Article), November 2018. https://www.politico.eu/article/germanys-plan-to-become-an-ai-powerhouse/
29. Delcker, J., "AI experts call to curb mass surveillance", *Politico* (Article), 24 June 2019. https://www.politico.eu/article/eu-experts-want-curtailing-of-ai-enabled-mass-monitoring-of-citizens/
30. Delcker, J., "Europe's silver bullet in global AI battle: Ethics", *Politico* (Article), 17 March 2019. https://www.politico.eu/article/europe-silver-bullet-global-ai-battle-ethics/
31. Delcker, J., "Finland's grand AI experiment", *Politico* (Article), 2 January 2019. https://www.politico.eu/article/finland-one-percent-ai-artificial-intelligence-courses-learning-training/
32. Delcker, J.,"Europe's AI ethics chief: No rules yet, please", *Politico* (Article), 30 October 2018. https://www.politico.eu/article/pekka-ala-pietila-artificial-intelligence-europe-shouldnt-rush-to-regulate-ai-says-top-ethics-adviser/
33. Determann, L., "No One Owns Data", *Hastings Law Journal*, Vol. 70, Issue 1, 2019, pp. 1-44, http://www.hastingslawjournal.org/wp-content/uploads/70.1-Determann.pdf
34. Edelman, Edelman AI Survey Results Report, 2019. https://www.edelman.com/sites/g/files/aatuss191/files/2019-03/2019_Edelman_AI_Survey_Whitepaper.pdf
35. Etzioni, Oren, "Point: Should AI Technology Be Regulated?: Yes, and Here's How", *Communications of the ACM,* December 2018, Vol. 61 No. 12, pp. 30-32
36. European Commission For The Efficiency Of Justice (CEPEJ), Charter on the use of Artificial Intelligence in judicial systems and their environment. Adopted at the 31st plenary meeting of the CEPEJ, Strasbourg, 3-4 December 2018. https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c
37. European Commission, High-Level Expert Group on Artificial Intelligence. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
38. European Parliament, "Legislative Train Schedule: Connected Digital Single Market", 2018. https://www.europarl.europa.eu/legislative-train/theme-connected-digital-single-market/file-artificial-intelligence-for-europe
39. European Parliament, Civil Law Rules on Robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.pdf
40. European Parliament, Follow up to the European Parliament resolution of 16 February 2017 on civil law rules on robotics. http://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2017/11-20/A8-0005-2017_EN.pdf
41. Future of Life Institute, "A Principled AI Discussion in Asilomar". https://futureoflife.org/2017/01/17/principled-ai-discussion-asilomar/
42. Future of Life Institute, "Asilomar AI PRINCIPLES". https://futureoflife.org/ai-principles/
43. G20, "G20 Ministerial Statement on Trade and Digital Economy", 8-9 June 2019. https://www.mofa.go.jp/files/000486596.pdf

44. Gal, Michal S. and Niva Elkin-Koren, "Algorithmic Consumers", *Harvard Journal of Law & Technology*, Vol. 30, Issue 2, 2017, pp. 309-353
45. Gayle, D. "UK, US and Russia among those opposing killer robot ban" *The Guardian*, March 29, 2019. https://www.theguardian.com/science/2019/mar/29/uk-us-russia-opposing-killer-robot-ban-un-ai
46. Gobierno de México, "Estrategia de Inteligencia Artificial MX 2018". https://www.gob.mx/mexicodigital/articulos/estrategia-de-inteligencia-artificial-mx-2018
47. Groth, Olaf J., Mark J. Nitzberg, Stuart J. Russell, "AI Algorithms Need FDA-Style Drug Trials", *WIRED opinion,* 15 August 2019. https://www.wired.com/story/ai-algorithms-need-drug-trials/#
48. Hendler, J. A., "Comments to Food and Drug Administration on AI-Augmented Software as a Medical Device Discussion Paper", ACM US Technology Policy Committee, 3 June 2019, https://www.acm.org/binaries/content/assets/public-policy/ustpc-comments-fda-software-based-device-safety-060319.pdf
49. Hern, A., "'Partnership on AI' formed by Google, Facebook, Amazon, IBM and Microsoft", *The Guardian*, 28 September 2016. https://www.theguardian.com/technology/2016/sep/28/google-facebook-amazon-ibm-microsoft-partnership-on-ai-tech-firms
50. High-Level Expert Group on AI, "Ethics guidelines for trustworthy AI", European Commission, 2018. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419
51. IAPP, "Privacy 2030: A New Vision for Europe", November 2019. https://iapp.org/resources/article/privacy-2030/
52. IEEE P2802 - Standard for the Performance and Safety Evaluation of Artificial Intelligence Based Medical Device: Terminology. https://standards.ieee.org/project/2802.html
53. IEEE, "Artificial Intelligence", IEEE Position Statement, Approved by the IEEE Board of Directors, 24 June 2019, https://globalpolicy.ieee.org/wp-content/uploads/2019/06/IEEE18029.pdf
54. IEEE, P7006 - Standard for Personal Data Artificial Intelligence (AI) Agent (2018). https://standards.ieee.org/project/7006.html
55. Independent Expert Report, "100 Radical Innovation Breakthroughs for the future, European Commission, May 2019. https://ec.europa.eu/info/sites/info/files/research_and_innovation/knowledge_publications_tools_and_data/documents/ec_rtd_radical-innovation-breakthrough_052019.pdf
56. ITU, "AI for Good Global Summit". https://aiforgood.itu.int/
57. ITU, "Artificial Intelligence". https://www.itu.int/en/ITU-T/AI/Pages/default.aspx
58. Kaminski, Margot E. and Malgieri, Gianclaudio, "Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations", *U of Colorado Law Legal Studies Research Paper* No. 19-28., 2019. http://dx.doi.org/10.2139/ssrn.3456224
59. Kaminski, Margot E., "Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability", *Southern California Law Review*, Vol. 92, No. 6, 2019
60. Kenyan Wall Street, "Kenya Govt unveils 11 Member Blockchain & AI Taskforce headed by Bitange Ndemo", *The Kenyan Wall Street*, 28 February 2018. https://kenyanwallstreet.com/kenya-govt-unveils-11-member-blockchain-ai-taskforce-headed-by-bitange-ndemo/
61. Koschwitz, L., "The copyright reform bug that risks derailing Europe's AI ambitions", *Politico* (Sponsored Content), 5 September 2018. https://www.politico.eu/sponsored-content/the-copyright-reform-bug-that-risks-derailing-europes-ai-ambitions/
62. Kuner, Christopher, Dan Jerker B. Svantesson, Fred H. Cate, Orla Lynskey, and Christopher Millard, "Machine learning with personal data: is data protection law smart enough to meet the challenge?", *International Data Privacy Law*, Vol. 7, Issue 1, 2017, pp. 1-2,. https://doi.org/10.1093/idpl/ipx003
63. Larus, James, Chris Hankin, Siri Granum Carson, Markus Christen, Silvia Grafa, Oliver Grau, Claude Kirchner, Bran Knowles, Andrew McGettrick, Damian Andrew Tamburri, and Hannes Werthner, "When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making", *Informatics Europe & EUACM*, 2018. https://www.acm.org/binaries/content/assets/public-policy/ie-euacm-adm-report-2018.pdf
64. Leenders, Gijs, "The Regulation of Artificial Intelligence — A Case Study of the Partnership on AI", *Becoming Human: Artificial Intelligence Magazine*, 13 April 2019. https://becominghuman.ai/the-regulation-of-artificial-intelligence-a-case-study-of-the-partnership-on-ai-c1c22526c19f
65. Leon, H., "Top Secret Military-Grade Surveillance Drones Might Be Coming To Your Neighborhood", *Observer,* 28 June 2019. https://observer.com/2019/06/gorgon-stare-aerial-surveillance-drones/

66. Mak, R., “Breakingviews - Review: Why an AI apocalypse could happen”, *Reuters* (BREAKINGVIEWS), 14 June 2019. https://www.reuters.com/article/us-tech-artificial-intelligence-breaking/breakingviews-review-why-an-ai-apocalypse-could-happen-idUSKCN1TF1FH
67. Martinho-Truswell, Emma, Hannah Miller, Isak Nti Asare, André Petheram, Richard Stirling, Constanza Gómez Mont, and Cristina Martinez, “Towards an AI strategy in Mexico: Harnessing the AI Revolution”, 2018. https://docs.wixstatic.com/ugd/7be025_e726c582191c49d2b8b6517a590151f6.pdf
68. Mayer-Schonberger, Viktor, and Yann Padova, "Regime change: enabling big data through Europe's new data protection regulation", *Colum. Sci. & Tech. L. Rev.* 17, 2015, p. 315
69. Morning Consult, “National Tracking Poll #170401”, 30 March - 1 April 2017. https://morningconsult.com/wp-content/uploads/2017/04/170401_crosstabs_Brands_v3_AG.pdf, pp. 118-123
70. Motoyama, S., “Inside the United Nations’ Effort To Regulate Autonomous Killer Robots: Meet the UN diplomat heading up the coming ‘killer robot’ conference”, *The Verge*, 27 August 2018. https://www.theverge.com/2018/8/27/17786080/united-nations-un-autonomous-killer-robots-regulation-conference
71. New Generation of Artificial Intelligence Development Plan, State Council Document No. 35, translation available at: https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf
72. NIST, “U.S. Leadership plan in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools, 2019.” https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf
73. Nuffield Council on Bioethics, “Artificial intelligence (AI) in healthcare and research”, Briefing Note, 2018, http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf
74. OECD, “Recommendation of the Council on Artificial Intelligence”. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
75. OECD, “What are the OECD Principles on AI?”. https://www.oecd.org/going-digital/ai/principles/
76. Omarova, S. T., “New Tech v. New Deal: Fintech as a Systemic Phenomenon”, *Yale Journal on Regulation*, Vol. 36, Issue 2, 2019, pp. 735-793. http://yalejreg.com/articlepdfs/36-JREG-735-Omarova.pdf
77. Pagallo, Ugo, “Vital, Sophia, and Co.—The Quest for the Legal Personhood of Robots”, *Information*, 9(9), pp. 230 (2018). https://www.mdpi.com/2078-2489/9/9/230
78. Partnership on AI, “Tenets”. https://www.partnershiponai.org/tenets/
79. Petit, N., “Law and regulation of artificial intelligence and robots: Conceptual framework and normative implications”, Working Paper, 2017, p. 30 https://dx.doi.org/10.2139/ssrn.2931339
80. Price II, W. N., “Artificial Intelligence in Health Care: Applications and Legal Implications”, *The SciTech Lawyer*, Vol. 14, Issue 1, 2017, pp. 10-13. https://repository.law.umich.edu/articles/1932
81. Privacy International and Article 19, “Privacy and Freedom of Expression in the Age of Artificial Intelligence”, April 2018. https://privacyinternational.org/sites/default/files/2018-04/Privacy%20and%20Freedom%20of%20Expression%20%20In%20the%20Age%20of%20Artificial%20Intelligence.pdf
82. Privacy International, “About Privacy International”. https://privacyinternational.org/about
83. Scherer, Matthew U. "Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies," *Harv. JL & Tech.* 29, 2015, p. 353.
84. Sela, A., “Can Computers Be Fair? How Automated and Human-Powered Online Dispute Resolution Affect Procedural Justice in Mediation and Arbitration”, *Ohio State Journal on Dispute Resolution*, Vol. 33, Issue 1, 2018, pp. 91-148.
85. SHERPA, “Scenario: predictive policing in 2025”, 2019. https://www.project-sherpa.eu/scenarios/predictive-policing-complete/
86. SHERPA, Scenario: Information Warfare in 2025, 2019. https://www.project-sherpa.eu/scenarios/warfare-complete/
87. SIENNA, *D4.2: Analysis of the legal and human rights requirements for AI and robotics in and outside the EU*, 2019.
88. Sokol, D. Daniel, and Roisin Comerford,"Antitrust and Regulating Big Data," *Geo. Mason L. Rev.* 23, 2015, p. 1129.
89. Taddeo, Mariarosaria and Floridi, Luciano, “Regulate Artificial Intelligence to Avert Cyber Arms Race”, *Nature* 556, 296-298, 2018, doi: 10.1038/d41586-018-04602-6.

90. The Council of Europe Commissioner for Human Rights, "Unboxing Artificial Intelligence: 10 steps to protect Human Rights, p. 6. https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64
91. The Economist, "Dismembering big tech", *The Economist*, 24 Oct 2019. https://www.economist.com/business/2019/10/24/dismembering-big-tech
92. The Federal Government, "Artificial Intelligence Strategy", November 2018. https://www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie_engl.pdf
93. The Law Library of Congress, Global Legal Research Directorate, "Regulation of Artificial Intelligence in Selected Jurisdictions", January 2019. https://www.loc.gov/law/help/artificial-intelligence/regulation-artificial-intelligence.pdf, pp. 119-132
94. The President, Maintaining American Leadership in Artificial Intelligence, Executive Order 13859 of 11 February 2019. Federal Register 84(31), pp. 3967-3972. https://www.govinfo.gov/content/pkg/FR-2019-02-14/pdf/2019-02544.pdf
95. The Public Voice, "Universal Guidelines for Artificial Intelligence", 23 October 2018. https://thepublicvoice.org/ai-universal-guidelines/
96. The Select Committee On Artificial Intelligence Of The National Science & Technology Council, The National Artificial Intelligence Research And Development Strategic Plan: 2019 Update. A Report, 2019. https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf
97. Thierer, Adam, "Problems with Precautionary Principle-Minded Tech Regulation & a Federal Robotics Commission", *The Technology Liberation Front,* 22 September 2014. *https://techliberation.com/2014/09/22/problems-with-precautionary-principle-minded-tech-regulation-a-federal-robotics-commission/*
98. Thierer, Adam, Andrea Castillo O'Sullivan, and Raymond Russell, "Artificial Intelligence and Public Policy", Mercatus Research, Mercatus Center at George Mason University, 2017. https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf
99. Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry (2018–2020), translation available at: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-government-outlines-ai-ambitions-through-2020/
100. UN, "Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression". https://undocs.org/A/73/348
101. UN, "Sustainable Development Goals". https://sustainabledevelopment.un.org/?menu=1300
102. UNICRI, "UNICRI Centre for Artificial Intelligence and Robotics". http://www.unicri.it/in_focus/on/UNICRI_Centre_Artificial_Robotics
103. van Veen, Christiaan, "Artificial Intelligence: What's Human Rights Got To Do With It?" *Points,* 14 May 2018. https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d
104. Veale, Michael and Lilian Edwards, "Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling", *Computer Law & Security Review*, Vol. 34, Issue 2, 2018, pp. 398-404. https://doi.org/10.1016/j.clsr.2017.12.002
105. Villaronga, Eduard Fosch, Peter Kieseberg, and Tiffany Li. "Humans forget, machines remember: Artificial intelligence and the Right to Be Forgotten", *Computer Law & Security Review*, Vol. 34, Issue 2, 2017, pp. 304-313. https://doi.org/10.1016/j.clsr.2017.08.007
106. Wachter, Sandra, and Brent Mittelstadt, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI", *Columbia Business Law Review*, Vol. 2019, No. 2, May 2019. https://journals.library.columbia.edu/index.php/CBLR/article/view/3424
107. Wachter, Sandra, Brent Mittelstadt, and Chris Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR", *Harvard Journal of Law & Technology*, Vol. 31, Issue 2, 2018, pp. 841-887
108. Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation", *International Data Privacy Law,* Vol. 7, Issue 2, May 2017, pp. 76-99
109. Wallach, Wendell, and Gary Marchant, "Toward the Agile and Comprehensive International Governance of AI and Robotics [point of view]," *Proceedings of the IEEE* ,107.3, 2019, pp. 505-508
110. Watney, Caleb, "A Framework for Increasing Competition and Diffusion in Artificial Intelligence", *R Street,* 21 March 2019. https://www.rstreet.org/2019/03/21/a-framework-for-increasing-competition-and-diffusion-in-artificial-intelligence/

111. World Economic Forum, "Public Concern Around Use of Artificial Intelligence is Widespread, Poll Finds", World Economic Forum, 1 July 2019. https://www.weforum.org/press/2019/07/public-concern-around-use-of-artificial-intelligence-is-widespread-poll-finds
112. Zhang, Baobao and Allan Dafoe "Artificial Intelligence: American Attitudes and Trends. Center for the Governance of AI", Future of Humanity Institute, University of Oxford, 2019. https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/

Annexes

1. List of stakeholders consulted

Name	Organisation	Mode/type of input (month/year)
Thamar Zijlstra Marlou Bijlsma	NEN	Discussion of task (June 2019)
Andreas Andreou	AEQUITAS	Discussion of task (June 2019)
Chiara Giovannini, ANEC Félicien VALLET, Commission Nationale de L'informatique des Libertes Krista Varantola, University of Tampere Lisa Sammer, SAP Maja Brkan, Maastricht University Maria de Kleijn, SVP Analytical Services Marie-Valentine Florin, L'Ecole Polytechnique Fédérale De Lausanne Martijn Scheltema, Erasmus University Rotterdam Mihail Kritikos, Institute of European Studies, Vrije Universiteit Brussel Yoan Miche, Nokia Bell Labs	SHERPA Stakeholder Advisory Board	Feedback on scoping paper on criteria and identification of options (Sept. 2019)
Vincent Bryce	DMU/Nottingham University	Feedback on scoping paper on criteria and identification of options (Sept. 2019)
Albena Kuyumdzhieva	European Commission	Feedback on scoping paper on criteria and identification of options (Sept. 2019)
Andrew Murray	London School of Economics and Political Science	Inputs/feedback on option (Nov 2019)
Giuseppe Stefano Quintarelli	Copernicani	Inputs/feedback on option (Nov 2019)
Olivia Erdélyi	School of Law, University of Canterbury	Inputs/feedback on option (Nov 2019)

Name	Organisation	Mode/type of input (month/year)
Félicien Vallet	Commission Nationale de L'informatique des Libertes	Feedback on option (Nov 2019)
Adam Holland	Berkman Klein Center	Feedback on option (Nov 2019)
Javier Valls Prieto	University of Granada	Inputs/feedback on option (Oct 2019)
Zuzanna Warso	Trilateral Research	Feedback on option (Oct 2019)
Virginia Dignum	Umeå University, Department of Computer Science; SHERPA SAB	Feedback on draft report (Dec 2019)
James Rule	Center for the Study of Law and Society, University of California; SHERPA SAB	Feedback on draft report (Dec 2019)
Carl Wiper	Information Commissioner's Office (ICO); SHERPA SAB	Feedback on draft report (Dec 2019)
Marie-Valentine Florin	EPFL - L'Ecole polytechnique fédérale de Lausanne; SHERPA SAB	Feedback on draft report (Dec 2019)
Maria de Kleijn-Lloyd	SVP Analytical Services, ELSEVIER; HERPA SAB	Feedback on draft report (Dec 2019)

2. Scoping paper

Link: https://www.project-sherpa.eu/regulatory-options-for-smart-information-systems-sis-ai-and-big-data-analytics/

3. Final list of options studied

INTERNATIONAL LEVEL

Title	Proposer(s)	Analysed by
1. Moratorium on the development of 'lethal autonomous robotics' (LARs) (UN); Moratorium on development of offensive LAWS (AI HLEG)	UN HRC Special Rapporteur report 9 April 2013, A/HRC/23/47; European Parliament Resolution 2018; High-Level Expert Group on AI (AI HLEG) 2019	TRI
2. Binding Framework Convention to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law	High Level Conference on AI, Helsinki, Feb 2019 (Finnish Presidency, French Presidency of CoE, Committee of Ministers and CoE)	UCLANCY
3. Legislative framework for independent and effective oversight over the human rights compliance of the development, deployment and use of AI systems by public authorities and private entities	Council of Europe	TRI
4. Legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities	Council of Europe Commissioner for Human Rights	TRI
5. Convention on human rights in the robot age	Parliamentary Assembly of the Council of Europe	UCLANCY
6. CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment	EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ)	TRI
7. International Artificial Intelligence Organization	Olivia J. Erdélyi and Judy Goldsmith	TRI
8. Global legal AI and/or robotics observatory	SIENNA H2020, D4.2	TRI

EU-LEVEL

Title	Proposer(s)	Analysed by
9. EU-level special list of robot rights	SHERPA project (Deliverable D1.5)	UCLANCY
10. Adoption of common Union definitions of cyber physical systems, autonomous systems, smart autonomous robots	European Parliament	UCLANCY
11. Creating electronic personhood status for autonomous systems	European Parliament	UCLANCY

Title	Proposer(s)	Analysed by
12. Establishment of a comprehensive Union system of registration of advanced robots within the Union's internal market where relevant and necessary for specific categories of robots and establishment of criteria for the classification of robots that would need to be registered	European Parliament	UCLANCY
13. General fund for all smart autonomous robots or individual fund for each and every robot category	European Parliament	TRI
14. Mandatory consumer protection impact assessment	AI HLEG	TRI
15. EU Taskforce of field specific regulators for AI/big data	SHERPA SAB member	TRI
16. Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations	Margot E. Kaminski and Gianclaudio Malgieri	TRI
17. Voluntary/mandatory certification of algorithmic decision systems (ADS)	Certification of AI systems at EU level (AI HLEG, Policy and Investment recommendations, 2019); Voluntary/mandatory certification of algorithmic decision systems (ADS), (STOA study Understanding algorithmic decision-making: Opportunities and challenges, 2019);	TRI

NATIONAL LEVEL

Title	Proposer(s)	Analysed by
18. Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or the "DEEP FAKES Accountability Act", (H.R. 3230) 116th Cong. (2019)	Rep. Yvette D. Clarke	TRI
19. Algorithmic Accountability Act of 2019 (HR 2231, 116th Congress)	Rep. Yvette D. Clark	UCLANCY
20. Directive on Automated Decision-Making	Government of Canada	UCLANCY

Title	Proposer(s)	Analysed by
21. US Food and Drug Administration regulation of adaptive AI/ML technology	US Food and Drug Administration	UCLANCY
22. New statutory duty of care for online harms (UK Government)	UK Government	UCLANCY
23. Redress by design mechanisms for AI	High-Level Expert Group on Artificial Intelligence (AI HLEG)	TRI
24. Register of algorithms used in government	New Zealand Law Foundation and University of Otago	TRI
25. Digital Authority	UK House of Lords Select Committee on Communications	TRI
26. Independent cross-sector advisory body (Centre for Data Ethics and Innovation)	UK Government	TRI
27. FDA for algorithms	Andrew Tutt	TRI
28. US Federal Trade Commission to regulate robotics	Various, Woodrow Hartzog	UCLANCY

CROSS-OVER

Title	Proposer(s)	Analysed by
29. Using anti-trust regulations to break up big tech and appoint regulators to reverse illegal and anti-competitive tech mergers	Elizabeth Warren, US Senator	TRI
30. Three-level obligatory impact assessments for new technologies	Paul Nemitz	TRI
31. Regulatory sandboxes	European Commission, European Parliament, EC AI HLEG	UCLANCY

4. Individual assessments of proposed options

4.1. Moratorium on the development of 'lethal autonomous robotics'/offensive LAWS

Option: Moratorium on the development of 'lethal autonomous robotics' (LARs) (UN report); Moratorium on development of offensive LAWS (AI HLEG) Proposers: Various. UN HRC Special Rapporteur report 9 April 2013, A/HRC/23/47; European Parliament Resolution 2018; High-Level Expert Group on AI (AI HLEG) 2019; References/links to relevant document: https://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf http://www.europarl.europa.eu/doceo/document/B-8-2018-0362_EN.html; https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 29 Oct 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	Connected/applicable to 'lethal autonomous weapon systems' or LAWS refers to weapon systems without meaningful human control over the critical functions of selecting and attacking individual targets or lethal autonomous robotics (LARs) or automated lethal weapons. The **UN HRC report** calls for placement of a national moratorium on LARs. The **EU Parliament Resolution** recognises that a "growing number of states have called for a preventative prohibition on LAWS and a moratorium on the use and production of such autonomous systems". The **AI HLEG** calls for monitoring and restricting the development of automated lethal weapons, including cyber-attack tools that can have lethal consequences if deployed adoption of a moratorium on the development of offensive LAWS.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: International and or national law, depending on framework that is used as its basis. Nature: As explained by Yin, "Moratorium, as a postponement or suspension of an activity, is widely used as a middle ground between YES and NO in the international legal arena, which reflects the value of compromise and cooperation in international intercourse. Moratorium in international legal setting is considered an option where countries are unable to perform their obligations for a reason- able time period, or an extraordinary situation requires countries to take exceptional measures or countries deem it necessary or indispensable for achieving some policy goals. The special values of moratoria shed light on difficult and complex issues to be addressed by States." Yin, Wenqiang, "Moratorium in international law" *Chinese Journal of International Law* 11.2, 2012, pp. 321-340. https://pdfs.semanticscholar.org/5305/42e917fa74c3dbd90c93caa9e354b872542b.pdf Scope: In the **UN HRC report**, the Special Rapporteur recommend, "Human Rights Council should call on all States to declare and implement national moratoria on at least the testing, production, assembly, transfer, acquisition, deployment and use of LARs until such time as an

Option: Moratorium on the development of 'lethal autonomous robotics' (LARs) (UN report); Moratorium on development of offensive LAWS (AI HLEG) Proposers: Various. UN HRC Special Rapporteur report 9 April 2013, A/HRC/23/47; European Parliament Resolution 2018; High-Level Expert Group on AI (AI HLEG) 2019; References/links to relevant document: https://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf http://www.europarl.europa.eu/doceo/document/B-8-2018-0362_EN.html; https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 29 Oct 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	internationally agreed upon framework on the future of LARs has been established". The recommendation is for a national moratorium on LARs as an "immediate step". It commends the US 2012 Department of Defense Directive that bans the development and fielding of LARs unless certain procedures are followed and suggests this "may open up opportunities for mobilizing international support for national moratoria". The focus in the **EU Parliament Motion** is on " moratorium on the use and production of such autonomous systems" (though its intent could be read to as being preventative). The **AI HLEG Policy recommendations** moratorium is specific to "development of offensive LAWS". The first document is more detailed and wider in scope than the latter two, which seem to focus more specifically.
3. Purpose/objective/what need does the option fulfil?	To restrict the development and/or stop the deployment of lethal autonomous weapon systems. To ensure humans are kept in control of weapons (EU Parliament Resolution). The UN HRC report outlines: "Moratoria are needed to prevent steps from being taken that may be difficult to reverse later, while an inclusive process to decide how to approach this issue should occur simultaneously at the domestic, intra-State, and international levels." https://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf
4. What gap does it address?	The UN HRC report outlines, "Moratoria are needed to prevent steps from being taken that may be difficult to reverse later, while an inclusive process to decide how to approach this issue should occur simultaneously at the domestic, intra-State, and international levels." See https://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf
5. What added value does it have?	A moratorium would potentially signal the dangers associated with LARs and/or LAWs. It might allay concerns of people and communities that are affected by the use and implementation of such systems. As Yin explains, it is particularly useful tool in cases where countries are faced with international complexities to address that cannot be reconciled in a reasonable term.
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed*	Limitations and risks: Such a moratorium might not take into account the differences of the status quo in the countries and their political ambitions on LARS/LAWS and therefore might be considered unjust by some or adversely affect some parties. Such a moratorium might be hard to

Option: Moratorium on the development of 'lethal autonomous robotics' (LARs) (UN report); Moratorium on development of offensive LAWS (AI HLEG) Proposers: Various. UN HRC Special Rapporteur report 9 April 2013, A/HRC/23/47; European Parliament Resolution 2018; High-Level Expert Group on AI (AI HLEG) 2019; References/links to relevant document: https://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf http://www.europarl.europa.eu/doceo/document/B-8-2018-0362_EN.html; https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 29 Oct 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).	accomplish because of the existing AI race to the bottom competition between different countries. As Yin outlines, "it is undoubtedly a challenge to put a moratorium in place, which always needs coordination of political wills of the related countries. However, where countries could not find other better solutions to intractable issues, moratorium might be the most practical one." Yin, Wenqiang, "Moratorium in international law" *Chinese Journal of International Law* 11.2, 2012, pp. 321-340. https://pdfs.semanticscholar.org/5305/42e917fa74c3dbd90c93caa9e354b872542b.pdf
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	The underlying framework in all three cases has not yet been specified. There are more details in the UN HRC report than the other two documents. Open questions that remain include what the process would be, exceptions, various effects of the moratorium (i.e., freezing the status quo effect, reversing effect). Yin, Wenqiang, "Moratorium in international law" *Chinese Journal of International Law* 11.2, 2012, pp. 321-340. https://pdfs.semanticscholar.org/5305/42e917fa74c3dbd90c93caa9e354b872542b.pdf
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	These have not been discussed in relation to the proposals. We could anticipate that depending on the level, there would be agreement and mechanisms implemented for monitoring and evaluating the moratorium without which such a measure would be ineffective.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	As Yin explains, if a moratorium with a "zero standard could seem too rigid and strict, or disproportionate to the required situations for some parties." Yin, Wenqiang, "Moratorium in international law" *Chinese Journal of International Law* 11.2, 2012, pp. 321-340. https://pdfs.semanticscholar.org/5305/42e917fa74c3dbd90c93caa9e354b872542b.pdf
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil*	Victims and casualties of LARs/LAWS. Civilian population who might be affected.

Option: Moratorium on the development of 'lethal autonomous robotics' (LARs) (UN report); Moratorium on development of offensive LAWS (AI HLEG) Proposers: Various. UN HRC Special Rapporteur report 9 April 2013, A/HRC/23/47; European Parliament Resolution 2018; High-Level Expert Group on AI (AI HLEG) 2019; References/links to relevant document: https://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf http://www.europarl.europa.eu/doceo/document/B-8-2018-0362_EN.html; https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 29 Oct 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
society; individuals, others (please specify)]	
11. Whose rights and/or interests does this option neglect?	State building arsenals/deploying LARs/LAWs.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	A moratorium would support human rights and societal values especially the protection of life (preventing loss of life plus the devaluation of it), international stability and security. In the UN HRC report, the Special Rapporteur draws attention to the supremacy and non-derogability of the right to life under both treaty and customary international law.
13. How does it address ethics and ethical principles? Which ones?	It addresses the following ethical principles: reducing harm, accountability, moral responsibility and meaningful human control.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not specified in any of the three documents.
16. What provisions are there for regular review and update?	Not specified. But one could anticipate it would be kept under review. Terms of reference for the review might include: assessment of the impacts (legal, political, economic, ethical) on LARs/LAWS, assessment of the impacts of allowing the moratorium to end and of extending it, what news laws, complementary policies and practices might be needed if the moratorium is ended (whether these exist/have been put in place so that moratorium could be ended).

Option: Moratorium on the development of 'lethal autonomous robotics' (LARs) (UN report); Moratorium on development of offensive LAWS (AI HLEG) Proposers: Various. UN HRC Special Rapporteur report 9 April 2013, A/HRC/23/47; European Parliament Resolution 2018; High-Level Expert Group on AI (AI HLEG) 2019; References/links to relevant document: https://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf http://www.europarl.europa.eu/doceo/document/B-8-2018-0362_EN.html; https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 29 Oct 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	This would depend on the type of moratorium implemented. Moratoria are susceptible to policy changes (e.g., where new policy determine these are counter-productive to innovation, economic prosperity).
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Such measures might affect the future development of LARs/LAWS technologies in terms of reducing/curtailing demand and curtailing their economic growth.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The EU Parliament could pass a resolution supporting a moratorium particularly having regard international and EU policy support on the topic. There could be a Council Decision establishing the position to be taken by the European Union.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	Depending on how expressed, the moratorium's exceptions for some activities related LARs/LAWs might create loopholes, i.e., the narrow scope of the moratorium and its many exclusions and exceptions may lead to disappointing results.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	3
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	As Jenks outlines, "Moratoriums have the advantage of scalability, later becoming a ban or alternatively, if technology mitigates the LAWS concerns, the moratorium could be lifted". See Jenks, Chris, False Rubicons, Moral Panic & Conceptual Cul-De-Sacs: Critiquing & Reframing the Call to Ban Lethal Autonomous Weapons (2016). Pepperdine Law Review, Vol. XLIV, No. 1, 2016; SMU Dedman School of Law Legal Studies Research Paper No. 243. https://ssrn.com/abstract=2736407 Moratoria could encourage States to ban development, use of or delay access to them LARS/LAWs, it may deter or delay research on these. At

Option: Moratorium on the development of 'lethal autonomous robotics' (LARs) (UN report); Moratorium on development of offensive LAWS (AI HLEG) Proposers: Various. UN HRC Special Rapporteur report 9 April 2013, A/HRC/23/47; European Parliament Resolution 2018; High-Level Expert Group on AI (AI HLEG) 2019; References/links to relevant document: https://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf http://www.europarl.europa.eu/doceo/document/B-8-2018-0362_EN.html; https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 29 Oct 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	the international level, this would work best at the United Nations General Assembly level (akin to the Moratorium on the death penalty). Such a view is supported by the Future of Life Institute which recommends that "For obvious reasons, any moratorium should be global and sponsored by a UN-led commission". https://futureoflife.org/ai-policy-challenges-and-recommendations/
References consulted	Jenks, Chris, False Rubicons, Moral Panic & Conceptual Cul-De-Sacs: Critiquing & Reframing the Call to Ban Lethal Autonomous Weapons (2016). Pepperdine Law Review, Vol. XLIV, No. 1, 2016; SMU Dedman School of Law Legal Studies Research Paper No. 243. https://ssrn.com/abstract=2736407 US Department of Defense Directive, "Autonomy in Weapons Systems", Number 3000.09 of 21 November 2012, Glossary Part II. Yin, Wenqiang, "Moratorium in international law" *Chinese Journal of International Law* 11.2, 2012, pp. 321-340. https://pdfs.semanticscholar.org/5305/42e917fa74c3dbd90c93caa9e354b872542b.pdf

4.2. Binding Framework Convention for AI

Option: Binding Framework Convention to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law (Council of Europe) including through a new ad hoc committee on AI (CAHAI) Proposer: High Level Conference on AI, Helsinki, Feb 2019 (Finnish Presidency, French Presidency of CoE Committee of Ministers and CoE) Reference/link to relevant document: **https://www.coe.int/en/web/human-rights-rule-of-law/council-of-europe-s-contribution** Assessed by: UCLANCY Date of assessment: 11/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it	The binding Framework Convention is currently an action of the CoE, part of an umbrella framework with multiple actions and instruments, both current and forthcoming, and all set out at

Option: Binding Framework Convention to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law (Council of Europe) including through a new ad hoc committee on AI (CAHAI) Proposer: High Level Conference on AI, Helsinki, Feb 2019 (Finnish Presidency, French Presidency of CoE Committee of Ministers and CoE) Reference/link to relevant document: **https://www.coe.int/en/web/human-rights-rule-of-law/council-of-europe-s-contribution** Assessed by: UCLANCY Date of assessment: 11/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	https://www.coe.int/en/web/human-rights-rule-of-law/council-of-europe-s-contribution The Council of Europe is now working to assess and anticipate the impact of digital technologies and to develop legal and practical instruments to ensure that these technologies remain respectful of human rights and the principles of democracy and the rule of law. **So there is an expectation that AI and big data respect and build in human rights and general principles of law**. With respect to the impact of AI on justice (justice is dematerialised; for example online dispute resolution systems 'dehumanises' justice), one of the instruments created is the Charter on the use of AI in judicial systems and their environment (European Commission for the Efficiency of Justice, CEPEJ, December, 2018). Forthcoming is the development of an international legal instrument to establish common standards for the criminal law aspects of automated technologies, in particular automated vehicles. On 13 February 2019, the Committee of Ministers adopted a Declaration constituting the first global instrument to formalise the substantial risks of the capacity to manipulate algorithmic processes. The Council of Europe co-organised, together with the Finnish Chairmanship of the Committee of Ministers, the Conference "Mastering the rules of the game - the impact of the development of artificial intelligence on human rights, democracy and the rule of law" in Helsinki on 26 and 27 February 2019. Its Conclusions (some of which concern the impact of AI on justice) have influenced the Committee of Ministers' subsequent discussions and actions. At its 129th Session held in Helsinki on 16-17 May 2019, the Committee of Ministers instructed its Deputies to examine, on the basis of multi-stakeholder consultations**, the feasibility and potential elements of a legal framework for the development, design and application of AI, based on Council of Europe standards in the field of human rights, democracy and the rule of law**. To this end, the terms of reference of a new ad hoc committee on AI (CAHAI) were adopted in September 2019. A draft Recommendation of the Committee of Ministers on the impact of algorithmic systems on human rights is also being prepared by the Committee of experts on Human Rights Dimensions of automated data processing and different forms of artificial intelligence (CDMSI/MSI-AUT).
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and	**AD HOC COMMITTEE ON AI (CAHAI)** https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016809737a1

Option: Binding Framework Convention to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law (Council of Europe) including through a new ad hoc committee on AI (CAHAI) Proposer: High Level Conference on AI, Helsinki, Feb 2019 (Finnish Presidency, French Presidency of CoE Committee of Ministers and CoE) Reference/link to relevant document: **https://www.coe.int/en/web/human-rights-rule-of-law/council-of-europe-s-contribution** Assessed by: UCLANCY Date of assessment: 11/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
scope (e.g., national or international, topic/domain/tech specific/general))?	*Set up by the Committee of Ministers under Article 17 of the Statute of the Council of Europe and in accordance with Resolution CM/Res(2011)24 on intergovernmental committees and subordinate bodies, their terms of reference and working methods and approved in Sept 2019.* ToR cover period until December 2021. Broad international scope under the CoE, under the 'Strengthening the Rule of Law' Sector, 'Information society and internet governance' programme. The idea is to introduce/enhance technology in human rights and general principles of law. This is a soft instrument placed under the authority of the Committee of Ministers of the CoE.
3. Purpose/objective/what need does the option fulfil?	The CAHAI has both main and specific tasks. Specific tasks include the conducting of a feasibility study and setting the foundations of a legal framework, produce a progress report with proposals for further action and working methods by May 2020.
4. What gap does it address?	It proposes to fill in a gap of a legal framework for the development, design and application of AI in view of CoE's standards on human rights, democracy and the rule of law.
5. What added value does it have?	It would have high added value to the extent that there is no comprehensive legal instrument on human rights and general principles of law that would regulate AI at the international/European level.
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*	The CAHAI would not produce the draft legislation but rather propose its legal foundations. Its mandate is therefore restricted. Like with any instrument of international law, there is a risk that member states do not engage sufficiently/support such an instrument. The main challenge is therefore how to compel member states to adopt a common legal instrument, abide to it and implement it at the international/national level. Even if this is a possibility, how about non-member states? There is a risk that they do not regulate AI at all or in a way that is not compatible with the specific provisions of the proposed legal framework.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	The option is sufficiently clear and specific. Its effectiveness is limited however as the mandate does not extend to drafting the legal framework itself. Operationalisation cannot therefore be achieved.

Option: Binding Framework Convention to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law (Council of Europe) including through a new ad hoc committee on AI (CAHAI) Proposer: High Level Conference on AI, Helsinki, Feb 2019 (Finnish Presidency, French Presidency of CoE Committee of Ministers and CoE) Reference/link to relevant document: **https://www.coe.int/en/web/human-rights-rule-of-law/council-of-europe-s-contribution** Assessed by: UCLANCY Date of assessment: 11/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Members of the CAHAI are representatives of governments of recognised expertise in the field of digital governance and legal implications of the functioning of different forms of AI relevant to the CoE mandate. These have the right to vote; any other representative sent by member states does not have the right to vote. Observers can be present. The rules of procedure of the Committee are governed by Resolution CM/Res(2011)24 on intergovernmental committees and subordinate bodies, their terms of reference and working methods, which is a general framework. The CAHAI will meet at State meetings. Planning, monitoring and evaluation and working methods are set out in the Resolution (2011). No specific enforcement mechanism is designed for this purpose. In the legal framework, it would be important to consider specific enforcement mechanisms although the CoE work on the basis of consensus.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	This is not directly relevant to citizens, public administrations or even businesses but mainly addressed to Governments. One could imagine that this work could impact the soft application or interpretation of general principles of law or constitutional rights at the national level.
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Beyond Government of member States who are direct participants to the work, participants may extend to CoE Institutions but also to EU representation such as the Fundamental Rights Agency, Observer States to the CoE and other IOs (OSCE, OECD, WHO, UNESCO and UN agencies). Such participants would not have a right to vote. Observers may include the European Network of National Human Rights Institutions, CoE partner internet companies, civil society organisations, other private sector and academic actors of relevance to the work of the Ad hoc Committee. The scope of stakeholders involved is therefore quite broad.
11. Whose rights and/or interests does this option neglect?	Citizens could be involved directly in the process through a public consultation effort. Direct democracy is missing from this initiative and specific human rights are not mentioned.

Option: Binding Framework Convention to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law (Council of Europe) including through a new ad hoc committee on AI (CAHAI) Proposer: High Level Conference on AI, Helsinki, Feb 2019 (Finnish Presidency, French Presidency of CoE Committee of Ministers and CoE) Reference/link to relevant document: **https://www.coe.int/en/web/human-rights-rule-of-law/council-of-europe-s-contribution** Assessed by: UCLANCY Date of assessment: 11/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	It explicitly supports human rights in their entirety, bridging the gap between human rights and technologies/AI.
13. How does it address ethics and ethical principles? Which ones?	The Ad hoc Committee in its work will take into account CoE standards relevant to the design, development and application of digital technologies, in the fields of human rights, democracy and the rule of law, in particular on the basis of existing legal instruments. It will also take into account relevant existing universal and regional international legal instruments, work undertaken by other CoE bodies as well as ongoing work in other international and regional organisations. Finally it will take due account of a gender perspective, building cohesive societies and promoting and protecting rights of persons with disabilities in the performance of its tasks. Ethical principles may not be expressly embedded but are addressed through democratic consideration.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	In addition to having gender dimension in its guiding principles, the Committee will appoint a Rapporteur on Gender Equality from amongst its members.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	It has a well-clarified source of funding, present and future, for intergovernmental work (i.e. meetings, reporting). It is unclear however whether the funding would extend to anybody/agency/authority to be created.
16. What provisions are there for regular review and update?	It is an ad-hoc committee, so no provision for regular review or update.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g.,	N/A

Option: Binding Framework Convention to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law (Council of Europe) including through a new ad hoc committee on AI (CAHAI) Proposer: High Level Conference on AI, Helsinki, Feb 2019 (Finnish Presidency, French Presidency of CoE Committee of Ministers and CoE) Reference/link to relevant document: **https://www.coe.int/en/web/human-rights-rule-of-law/council-of-europe-s-contribution** Assessed by: UCLANCY Date of assessment: 11/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
technological, policy changes, social demands?	
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	N/A
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	It is a good match with EU legal framework as it proposes to address the same basic principles. Depending on how AI is approached in the EU (as a shared competence between the EU and its member states), we could see a neater application of the binding framework principles, embedded into EU secondary legislation.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	Lack of enforcement and legal teeth are challenges. The proposal does not go far enough.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	5 but it is only a binding framework; much more work needed beyond.
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	All states of the CoE will have to find a consensus in this much needed field, before being in a position to draw up a legal framework, which, unless adopted as an equivalent at the EU level, will lack legal enforcement in countries.
References consulted	https://www.coe.int/en/web/secretary-general/-/high-level-conference-artificial-intelligence https://rm.coe.int/speech-of-thorbj-rn-jagland-secretary-general-council-of-europee-in-he/1680934fc2 https://rm.coe.int/conference-report-28march-final-1-/168093bc52 https://www.coe.int/en/web/human-rights-rule-of-law/council-of-europe-s-contribution https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016809737a1

Option: Binding Framework Convention to ensure that AI is designed, developed and applied in line with European standards on human rights, democracy and the rule of law (Council of Europe) including through a new ad hoc committee on AI (CAHAI) Proposer: High Level Conference on AI, Helsinki, Feb 2019 (Finnish Presidency, French Presidency of CoE Committee of Ministers and CoE) Reference/link to relevant document: **https://www.coe.int/en/web/human-rights-rule-of-law/council-of-europe-s-contribution** Assessed by: UCLANCY Date of assessment: 11/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	https://rm.coe.int/final-declaration-of-the-french-presidency-conference-of-mj-coe-15-oct/168098383f https://rm.coe.int/16808ac918 https://www.coe.int/en/web/data-protection/convention108-and-protocol

4.3. Legislative framework for oversight over the human rights compliance of AI systems

Option: Legislative framework for independent and effective oversight over the human rights compliance of the development, deployment and use of AI systems by public authorities and private entities Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	Assessment
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such	The Council of Europe Council of Europe Commissioner for Human Rights calls Members States to: • legislate for the establishment of a framework for independent and effective oversight over the human rights compliance of AI systems, drawing on existing oversight bodies including National Human Rights Structures where possible; • take steps to ensure all relevant oversight bodies have access to sufficient expertise, have received appropriate training on AI systems

Option: Legislative framework for independent and effective oversight over the human rights compliance of the development, deployment and use of AI systems by public authorities and private entities Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	Assessment
as transparency, robustness and security measures?) Give an application example)	and their implications for human rights, and have received adequate funding and other resources in order to carry out their functions effectively; • ensure that the functions of the relevant oversight bodies are adequate for the purpose of investigating and monitoring all actors, whether public or private, that may be responsible for AI system human rights violations (including those that occur during their development, testing and use). The Council of Europe urges for the application of independence and transparency safeguards, under which the oversight bodies and their staff enjoy institutional, operational, financial and personal independence. Oversight bodies must be also supported with resources and tools, including access to training and testing datasets, AI inputs/outputs, models/algorithms, operational guidance and human rights due diligence.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	The nature of this recommendation is not legally binding but advises Members States to establish a framework for independent and effective oversight of AI applications. This mainly includes the establishment and operation of independent and efficient administrative, judicial, quasi-judicial and/or parliamentary oversight bodies. So, the change this proposal aims to bring should be reflected in a regulatory and policy level. Member States should enact legislation to establish such oversight bodies or amend existing legislation to mandate existing bodies (this could include Data Protection Authorities and Research Ethics Committees) to undertake the oversight of AI applications. More specifically, to ensure complete financial, functional and operational independence, Member States could provide constitutional status to these oversight bodies, where possible.
3. Purpose/objective/what need does the option fulfil?	This proposal suggests the creation of a legislative framework for independent and effective oversight over human rights compliance of the development, deployment and use of AI systems by public authorities and private entities. In particular, this proposal aims to monitor, prevent and mitigate the negative impact of AI systems on human rights. It also aims to create independent oversight and compliance mechanisms over the AI applications in the public and private sector.
4. What gap does it address?	The Framework/establishment of oversight bodies aims to address the legal gaps in the governance of AI and is part of the 10-point Recommendation on AI and human rights issued by the Commissioner for Human Rights.
5. What added value does it have?	This option provides a valuable focal point for international dialogue and collaboration on AI public policy issues. Moreover, it suggests an inclusive governance framework for the establishment of oversight bodies. In addition, it covers the use of AI both by public and private sector.

Option: Legislative framework for independent and effective oversight over the human rights compliance of the development, deployment and use of AI systems by public authorities and private entities Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	Assessment
	The recommendation for establishing AI oversight bodies is not a novel recommendation (see e.g., European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics available at http://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html?redirect, suggestions by Big Brother Watch available at http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69661.html and recommendation of the Observatory for Responsible Research and Innovation in ICT available at http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69593.html. However, the added value of the recommendation lies in suggesting a thorough legislative and governance framework that is not restricted to self-regulation or co-regulation, ethics advisory or monitoring, or auditing responsibilities.
6. What are the limitations, risks and challenges?	Limitations: This proposal lacks detail and does not consider the legal requirements and idiosyncrasies in the different Member States, since independence may not be feasible or applicable in all Member States. In addition, oversight bodies and the exercise of their functions require that there is an enforceable governance framework for AI in national jurisdictions. Nonetheless, it is rather unlikely that most Member States have already legislated for the use and applications of AI in an adequate and complete manner. Risks: The main risk here is the creation of oversight bodies with restricted powers or limited resources. It is likely that under-funded and inefficient bodies may act as a typical mechanism for oversight, without actually being able to monitor and intervene in the public and private uses of AI. The main challenge of the recommendation relates to its actual enforcement. Member states separately should be convinced about the requirement for such oversight bodies, check their budget and proceed to policy assessments before establishing oversight bodies. Moreover, resource constraints and poor funding may prevent oversight bodies from being fully autonomous, independent and functional. In this context, some oversight bodies may not have the resources to recruit experts, conduct audits and provide scientific advice. Finally, the relationship of such oversight bodies with other supervisory authorities should be considered. In particular, due consideration should be given to ensuring that there is no conflict of interests or overlapping

Option: Legislative framework for independent and effective oversight over the human rights compliance of the development, deployment and use of AI systems by public authorities and private entities Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	Assessment
	responsibilities among supervisory authorities. For example, there should be a clear delineation of tasks and responsibilities between AI oversight bodies and Data Protection Authorities.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	The regulatory option is clear and key elements have been identified. However, there is a need for more detail about the legal nature, structure, aims, powers, tools, cooperation, liability, composition, and accountability of this regulatory option. Although the national margin for discretion and appreciation in legislation should be respected, there is a need for consistency and harmonisation in this area. Otherwise, different standards will apply decreasing harmonisation, efficiency and cooperation.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	This option is an oversight mechanism itself. For more detail please see above.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	The establishment of oversight bodies requires coordination and policy assessments. Establishing such bodies will take time and this is a more future-orientated solution with Member States having to amend national law or enact new legislation. For SMEs, being under the control and monitoring of AI oversight bodies may be a disincentive or hindrance in engaging with AI, especially for start-ups or may result in cover-ups of ethical issues.
10. Which stakeholders would benefit most from the use of this option?	This is not clarified in the original document. However, the anticipated benefits depend on the missions, tasks and powers of such oversight bodies and whether they are actually independent. This proposal provides that oversight bodies should have the power to intervene in circumstances where they identify (a risk of) human rights violations occurring. They should also regularly report to parliament and publish reports about their activities. Moreover, it is stated that they should proactively investigate and monitor the human rights compliance of AI systems, receive and handle complaints from affected individuals, carry out periodic reviews of AI system capabilities and technological developments. This system could increase transparency, accountability and compliance with law. Businesses will be supported with guidance and competition will be fair under the safeguard of a body overseeing the uses of AI in the market.

Option: Legislative framework for independent and effective oversight over the human rights compliance of the development, deployment and use of AI systems by public authorities and private entities Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	Assessment
	Moreover, the work of such bodies could be a point of reference for experts in academia, policymakers, the wider public and the judiciary. There is no doubt that the effective operation of oversight bodies will be of benefit to society if they act as a safeguard against risks for human rights because of AI applications. In addition to this, the operation of oversight bodies could have additional benefits for all the above actors if combined with other powers, such as licensing of AI applications, providing advice on security standards, imposing fines and acting as an Ombudsperson for complaints of the public.
11. Whose rights and/or interests does this option neglect?	Not elaborated.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	It actively and explicitly supports human rights by introducing compliance mechanisms to monitor, prevent and manage risks for human rights.
13. How does it address ethics and ethical principles? Which ones?	Not elaborated. However, this recommendation also refers to THE FIVE PRINCIPLES OF THE ETHICAL CHARTER ON THE USE OF ARTIFICIAL INTELLIGENCE IN JUDICIAL SYSTEMS AND THEIR ENVIRONMENT and to empowering National Human Rights Structures to perform a role in providing independent and effective oversight over the human rights compliance of AI systems. Traditionally, such structures also consider ethical principles in performing their roles, including fairness, minimisation of bias and discrimination, autonomy, self-governance, rule of law and democratic governance.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No, it doesn't refer to diversity and equality aspects. The composition of this body may depend on national policy and legal requirements.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a	No. The recommendation states that oversight bodies should be provided with adequate funding and other resources in order to carry out their functions effectively. It also explains that this suggested framework may include mechanisms that consist of a combination of administrative, judicial, quasi-

Option: Legislative framework for independent and effective oversight over the human rights compliance of the development, deployment and use of AI systems by public authorities and private entities Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	Assessment
body/agency/authority? Outline.	judicial and/or parliamentary oversight bodies effectively cooperating with each other. The funding and support of such bodies will originate from public resources and the state budget. It is not clarified whether other sources of funding could be used, such as funding from the European Union and public-private partnerships. The issue of finding is crucial, though, to ensure the independence of the body. The potential sources of funding are limited.
16. What provisions are there for regular review and update?	Provisions for the regular review and update of the Framework itself have not been specified. It is stated that the bodies should carry out periodic reviews of AI system capabilities and technological developments. In addition, the work of such bodies should be ongoing until their mission is abolished or paused.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	To ensure the feasibility, sustainability and future-proof character of this recommendation, specific impact, budget and policy assessments are required in each Member State. The size, structure, legal nature and powers of these bodies should reflect the legal idiosyncrasies of each Member State and respond to the policy and technological needs and priorities.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Businesses will need to provide all the information necessary for effective oversight of AI systems upon request and regularly report to the oversight bodies. This means that businesses will have to ensure that they are aware of such obligations and have the necessary resources and expertise to comply with this legislative framework. Whether the operation of oversight bodies will adversely impact the ability for businesses to innovate depends on the statutory powers and missions of the bodies. For example, if the deployment of AI applications is subject to the licensing and permission of the oversight body, it is likely that businesses will have to consider the effort, time, and financial resources to be put into this system.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	Oversight bodies are a commonly adopted institution under the EU legal framework. If the EU does not exercise its competence in enacting legal acts to establish AI oversight bodies (e.g., the case of Data Protection Authorities and the outline of their powers and responsibilities under Regulation (EU) 2016/679), Member States could also do this.

Option: Legislative framework for independent and effective oversight over the human rights compliance of the development, deployment and use of AI systems by public authorities and private entities Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	Assessment
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	As explained above.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	4
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Establishing an AI oversight body necessitates legislative amendments, impact, policy and financial assessments and advanced planning.
References consulted	As indicated above. In addition: Carrier, Ryan, "Implementing Guidelines for Governance, Oversight of AI, and Automation", Communications of the ACM, (62) 5, 2019, pp. 12-13 Etzioni, Amitai and O. Etzioni, "Designing AI systems that obey our laws and values", 2016 59(9), *Communications of the ACM*, pp. 29-31 Hall, Wendy and J. Pesenti, "Growing the artificial intelligence industry in the UK", 2017. https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk The British Academy and the Royal Society, "Data management and use: Governance in the 21st century - a British Academy and Royal Society project", 2018. https://royalsociety.org/topics-policy/projects/data-governance/ Women Leading in AI, *Principles for Responsible AI*, 2019. https://womenleadinginai.org/report2019

4.4. Legal framework for public authorities to carry out HRIAs

Option: Legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities. Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 15 Oct 2019 Stakeholder(s)/experts consulted in option assessment: Zuzanna Warso, Trilateral Research Ltd.	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example	AI systems acquired, developed and/or deployed by those authorities. In the document outlining this proposal the Council of Europe states, "AI is used as an umbrella term to refer generally to a set of sciences, theories and techniques dedicated to improving the ability of machines to do things requiring intelligence. An AI system is a machine-based system that makes recommendations, predictions or decisions for a given set of objectives." The document outlining this proposal calls for transparency requirements for oversight into AI systems; independent oversight bodies should proactively investigate and monitor the human rights compliance of AI systems; discrimination risks must be prevented and mitigated with special attention for groups that have an increased risk of their rights being disproportionately impacted by AI; and that The development, training, testing and use of AI systems that rely on the processing of personal data must fully secure a person's right to respect for private and family life under Article 8 of the European Convention on Human Rights, including the "right to a form of informational self-determination" in relation to their data. See https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	The proposal suggests: Member states should establish a legal framework that sets out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities. As part of the HRIA legal framework, public authorities should be required to conduct a self-assessment of existing and proposed AI systems. This self-assessment should evaluate the potential impact of the AI system on human rights taking into account the nature, context, scope, and purpose of the system. Where a public authority has not yet procured or developed a proposed AI system, this assessment must be carried out prior to the acquisition and/or development of that system. The HRIAs must also include a meaningful external review of AI systems, either by an independent oversight body or an external researcher/auditor with relevant expertise, in order to help discover, measure and/or map human rights impacts and risks over time. Public bodies should consider involving National Human Rights Structures (NHRSs) in carrying out this meaningful external review. Self-assessments and external reviews should not be limited to an evaluation of the models or algorithms behind the AI system, but should include an evaluation of how decision-makers might collect or influence the inputs and interpret the outputs of such a system. It should also include an assessment of whether an AI system remains under meaningful human control throughout the AI system's lifecycle. Member states may delineate the types of AI system that are subject to HRIAs under the law, but such delineations must be comprehensive

Option: Legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities. Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 15 Oct 2019 Stakeholder(s)/experts consulted in option assessment: Zuzanna Warso, Trilateral Research Ltd.	
Criteria/touch point	**Assessment**
	enough to cover all AI systems that have the potential to interfere with an individual's human rights at any stage of the AI system lifecycle. The Commissioner has issued the 10-point Recommendation on AI and human rights in which this proposal is outlined in "accordance with the mandate of the Commissioner for Human Rights to promote the awareness of and effective observance and full enjoyment of human rights in Council of Europe member states as well as to provide advice and information on the protection of human rights (Articles 3 and 8 of Resolution (99) 50 of the Committee of Ministers". p. 5 https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64
3. Purpose/objective/what need does the option fulfil?	A legal framework that sets out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities would help identify if an AI system poses risks to human rights and would help identify measures, safeguards, and mechanisms envisaged for preventing or mitigating that risk. The Commissioner suggests putting such a framework in place would help implement and operationalise HRIAs in a similar vein as other forms of impact assessment conducted by public authorities, such as Regulatory Impact Assessments and Data Protection Impact Assessments.
4. What gap does it address?	This proposal would help in holding AI actors to account for AI-related human rights violations. It would also address the issue of the "lack of a regulatory framework that can address the societal issues raised by these data- intensive technologies". (See Mantelero, Alessandro, "AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment", *Computer Law & Security Review*, 2018, 34(4), pp. 754-772)
5. What added value does it have?	It will potentially help reduce negative human rights impacts of algorithms used in public sector and enhance their benefits to society by providing stronger safeguards. It would foster due diligence and early addressing of any human rights issues of public sector AI systems as it calls for the HRIA to set out, inter alia, the measures, safeguards, and mechanisms envisaged for preventing or mitigating that risk.
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are*	Limitations: political will of Member States to set out a legal framework for HRIAs and/or open their public sector AI systems to scrutiny (self or otherwise). Risks: Some risks identified in relation to general HRIAs are also applicable here. E.g., • the political nature of the human rights framework poses some risks for the effectiveness of HRIAs as a policy and advocacy tool:

Option: Legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities. Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 15 Oct 2019 Stakeholder(s)/experts consulted in option assessment: Zuzanna Warso, Trilateral Research Ltd.	
Criteria/touch point	**Assessment**
potential or possible harms; challenges are difficulties it might face or be presented with).	http://siteresources.worldbank.org/PROJECTS/Resources/40940-1331068268558/HRIA_Web.pdf • carrying out HRIAs of good quality can be an onerous endeavour in terms of time, financial resources, data collection and types of expertise required: http://siteresources.worldbank.org/PROJECTS/Resources/40940-1331068268558/HRIA_Web.pdf • Resistance from actors reluctant to publicize sensitive information or damaging findings uncovered through HRIAs: http://siteresources.worldbank.org/PROJECTS/Resources/40940-1331068268558/HRIA_Web.pdf Challenges: As identified by some, the lack of clear definitions and standards regarding business obligations for human rights so far limits the objectivity and comparability of such tools: See https://www.business-humanrights.org/sites/default/files/reports-and-materials/Impact-assessments-CSR-Europe-June-2010.pdf
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	The Council of Europe Commissioner for Human Rights has only very briefly outlined the proposal in: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	The proposal states: The HRIAs must also include a meaningful external review of AI systems, either by an independent oversight body or an external researcher/auditor with relevant expertise, in order to help discover, measure and/or map human rights impacts and risks over time. Public bodies should consider involving National Human Rights Structures (NHRSs) in carrying out this meaningful external review. Self-assessments and external reviews should not be limited to an evaluation of the models or algorithms behind the AI system, but should include an evaluation of how decision-makers might collect or influence the inputs and interpret the outputs of such a system. It should also include an assessment of whether an AI system remains under meaningful human control throughout the AI system's lifecycle.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Yes, it will create implementation burdens for Member States and public authorities involved. W.r.t, public authorities, it will require them to conduct a self-assessment of existing and proposed AI systems evaluating the potential impact of the AI system on human rights taking into account the nature, context, scope, and purpose of the system. Where a public authority has not yet procured or developed a proposed AI system, this assessment must be carried out prior to the acquisition and/or development of that system. Thus, there is an obvious direct cost of

Option: Legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities. Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 15 Oct 2019 Stakeholder(s)/experts consulted in option assessment: Zuzanna Warso, Trilateral Research Ltd.	
Criteria/touch point	**Assessment**
	implementing HRIAs and compliance with them in public sector. If the law imposes excessive demands, it will be a burden on the organisations that come under its scrutiny (and the severity might be greater for SMEs servicing the public sector and unprepared to mitigate adverse AI effects due to lack of will, policy or resources)
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Individuals whose human rights are at risk of being violated by public sector AI systems; groups that have an increased risk of their rights being disproportionately impacted; profiled individuals belonging to specific groups. It would also benefit the public by making public authorities more accountable in their development and use of AI systems.
11. Whose rights and/or interests does this option neglect?	This option does not neglect any rights as such but will have some potentially adverse impacts on those subject to an HRIA. For example, the requirement for transparency might conflict with business interests in not publicising sensitive, proprietary or confidential information.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	The proposal states "the HRIA must set out the measures, safeguards, and mechanisms envisaged for preventing or mitigating that risk. In circumstances where such a risk has been identified in relation to an AI system that has already been deployed by a public authority, its use should be immediately suspended until the abovementioned measures, safeguards and mechanisms have been adopted." https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64
13. How does it address ethics and ethical principles? Which ones?	Indirectly via its focus on human rights.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	The Recommendation states: "Member states should effectively implement the UN Guiding Principles on Business and Human Rights and the Recommendation CM/Rec(2016)3 of the Committee of Ministers to member states on human rights and business. They should do so in a non-discriminatory manner with due regard to gender-related risks."
15. Does it have a well-clarified source of funding, present and future, especially where the option is a	Not elaborated.

Option: Legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities. Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 15 Oct 2019 Stakeholder(s)/experts consulted in option assessment: Zuzanna Warso, Trilateral Research Ltd.	
Criteria/touch point	**Assessment**
body/agency/authority? Outline.	
16. What provisions are there for regular review and update?	Not elaborated.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	To be determined. HRIAs are versatile tools but the law might face difficulties in being able to draw a line between what is in scope and out of its scope.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Yes. It will (positively) affect the design, development and use of existing and proposed AI systems in the public sector. Where effective, it will foster responsible innovation. It might also restrict the ability of businesses to innovate unrestrictedly with no consideration for human rights.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	This proposal is a good fit with the EU legal framework in as much as it would help Member States comply with the need to apply measures to protect human rights against AI-based violations and meet their positive and procedural obligations under the European Convention on Human Rights. As the Council outlines, "Member states should specifically ensure that their legislation creates conditions that are conducive to the respect for human rights by AI actors and do not create barriers to effective accountability and remedy for AI- related human rights violations." https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	There is also the "difficulty of developing appropriate human rights indicators that have the required contextual specificity, which is tailored to the problems of the country concerned." (Velluti, S., "The Promotion and Integration of Human Rights in EU External Trade Relations", *Utrecht Journal of International and European Law*, 32(83), 2016, pp.41–68)
21. Based on this study, how likely is this option to succeed ? (1 – Extremely	3

Option: Legal framework in Member States setting out a procedure for public authorities to carry out human rights impact assessments (HRIAs) on AI systems acquired, developed and/or deployed by those authorities. Proposer: Council of Europe Commissioner for Human Rights Reference/link to relevant document: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 Assessed by: TRI Date of assessment: 15 Oct 2019 Stakeholder(s)/experts consulted in option assessment: Zuzanna Warso, Trilateral Research Ltd.	
Criteria/touch point	**Assessment**
unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	The success of a legal framework for HRIAs for public sector will depend on how well the framework and procedures are set out, whether there is a specification of when they should be carried out, what incentives are offered for their use/penalties set for non-compliance, what comes within their scope, what are the key requirements, when it should be carried out, who should be involved and what the process should be. Generally there is No 'one size fits all' model for conducting HRIAs, and there are no guarantees that HRIAs will be robust, meaningful and change policy outcomes. Harrison, James. "Human Rights Impact Assessments of Trade Agreements: Reflections on Practice and Principles for Future Assessments. Background Paper for the Expert Seminar on Human Rights Impact Assessments of Trade and Investment Agreements, 23-24 June 2010 Geneva. https://warwick.ac.uk/fac/soc/law/.../hrequalityimpact/vienna_trade_hria_paper.doc
References consulted	Indicated above.

4.5. Convention on human rights in the robot age

Option: Convention on human rights in the robot age Proposer: Parliamentary Assembly of the Council of Europe Reference/link to relevant document: https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=23726&lang=en Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it	The Rathenau Institute report[1] recommended that the Parliamentary Assembly of the Council of Europe ("PACE") *recommend developing* a "convention on safeguarding human rights in the robot age to create common guiding principles to preserve human dignity in the way humans apply innovations in the field of the Internet of Things, including the Internet,

Option: Convention on human rights in the robot age Proposer: Parliamentary Assembly of the Council of Europe Reference/link to relevant document: https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=23726&lang=en Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	robotics, AI, and virtual and augmented reality". The report did not include a specific proposal for the convention. Neither the resulting Report by the Committee on Culture, Science, Education and Media[2] nor the Recommendation adopted by PACE[3] included a recommendation for a new binding convention. The adopted PACE Recommendation included only one proposal for a regulatory change (see below). The remaining proposals were for non-binding guidelines or updates to non-binding strategies. Recommendation: . Finalise the modernisation of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108) in order to have new provisions making it possible to put rapidly in place more appropriate protection. One year after the adoption by PACE of the Recommendation, a Protocol was adopted on 18 May 2018[4] to modernise the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data ("Convention 108"), which Protocol is currently awaiting approval by all Convention 108 signatories.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	The Rathenau Institute report recommends the development of a new Convention on human rights in the robot age to "create common guiding principles to preserve human dignity in the way humans apply innovations in the field of the Internet of Things, including the Internet, robotics, AI, and virtual and augmented reality". The Convention would be international and binding on signatories.
3. Purpose/objective /what need does the option fulfil?	The scope of the proposed Convention is vague, but the report recommends that the Council of Europe form opinions on a variety of topics as a first step toward setting an agenda to develop a Convention on robot ethics. The topics include: • psychological experiments involving humans taking place on the Internet and whether the firms that are doing these psychological experiments on the Internet should follow the ethics codes that currently apply when doing psychological experiments • whether and how persuasion software can be developed that respects people's agency • how information and communication technologies (ICTs) can be designed in such a way that they comply with the right to respect for family life • guidelines on engineering techniques and methods that permit AI and robotics to fully respect the individual's dignity and rights, allowing vulnerable groups to fully and effectively participate in society and live their lives in dignity

Option: Convention on human rights in the robot age Proposer: Parliamentary Assembly of the Council of Europe Reference/link to relevant document: https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=23726&lang=en Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	• guidance on ownership matters in the robot age (interference by virtual objects with tangible property, continued control, access and use of connected devices, and ownership of data) • how to apportion liability with regard to robots • the role of information gatekeepers as news editors and possibly as public watchdogs • how algorithmic accountability or fairness can be facilitated and how the developers of algorithms can be enabled to devise automated decisions that respect human rights and will not (unintentionally) discriminate against individuals • a framework of minimum norms to be taken into account when a court uses AI • to what extent in the context of the robot age the right to respect for privacy implies the right to not be measured, analysed or coached • to what extent in the context of the robot age the right to respect for family life should also include the right to meaningful human contact.
4. What gap does it address?	The report highlights a variety of human rights challenges and potential issues that may arise from development of intelligent artefacts and/or increased use of connected devices with respect to: • privacy • human dignity • ownership • safety, responsibility and liability • freedom of expression • prohibition of discrimination • access to justice and the right to a fair trial The report also suggests that the Council consider two new human rights: • right not to be measured, analysed or coached • right to meaningful human contact
5. What added value does it have?	(Insufficient detail)
6. What are the limitations, risks and challenges?	(Insufficient detail)
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	No. The report suggests that the Council explore and reach a position on a number of topics but doesn't suggest specific proposals.

Option: Convention on human rights in the robot age Proposer: Parliamentary Assembly of the Council of Europe Reference/link to relevant document: https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=23726&lang=en Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	(Insufficient detail)
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	(Insufficient detail)
10. Which stakeholders would benefit most from the use of this option?	(Insufficient detail)
11. Whose rights and/or interests does this option neglect?	(Insufficient detail)
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	The report explicitly discusses human rights challenges that emerging technology may provide but doesn't provide specific proposals to address the challenges.
13. How does it address ethics and ethical principles? Which ones?	(Insufficient detail)
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	(Insufficient detail)

Option: Convention on human rights in the robot age Proposer: Parliamentary Assembly of the Council of Europe Reference/link to relevant document: https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=23726&lang=en Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	(Insufficient detail)
16. What provisions are there for regular review and update?	(Insufficient detail)
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	(Insufficient detail)
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	(Insufficient detail)
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	(Insufficient detail)
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	(Insufficient detail)

Option: Convention on human rights in the robot age Proposer: Parliamentary Assembly of the Council of Europe Reference/link to relevant document: https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=23726&lang=en Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	(Insufficient detail)
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	(Insufficient detail)
References consulted	1. Van Est, R. & J.B.A. Gerritsen, with the assistance of L. Kool, Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality – Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE), The Hague: Rathenau Instituut 2017. https://www.rathenau.nl/sites/default/files/2018-02/Human%20Rights%20in%20the%20Robot%20Age-Rathenau%20Instituut-2017.pdf 2. Committee on Culture, Science, Education and Media, Technological convergence, artificial intelligence and human rights report, Doc. 14288, 10 Apr 2017. https://assembly.coe.int/nw/xml/XRef/Xref-DocDetails-EN.asp?FileID=23531&lang=EN 3. Council of Europe – Parliamentary Assembly, Recommendation 2102 (2017), adopted 28 April 2017. https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=23726&lang=en 4. Council of Europe Committee of Ministers, Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108), CM(2018)2-final, 18 May 2018. https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=090000168089ff4e

4.6. CEPEJ European Ethical Charter on the use of AI in judicial systems

Option: CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment Proposer: EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ) Reference/link to relevant document: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c Assessed by: TRI Date of assessment: 6 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	The CEPEJ European Ethical Charter on the use of AI applies in the context of judicial systems and their environment. It outlines five principles. 1. Principle of respect for fundamental rights: ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights 8 2. Principle of non-discrimination: Specifically prevent the development or intensification of any discrimination between individuals or groups of individuals 3. Principle of quality and security: With regard to the processing of judicial decisions and data, use certified sources and intangible data with models conceived in a multi-disciplinary manner, in a secure technological environment 4. Principle of transparency, impartiality and fairness: Make data processing methods accessible and understandable, authorise external audits 5. Principle "under user control": Preclude a prescriptive approach and ensure that users are informed actors and in control of their choices.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: Fundamental rights of individuals as set out in the European Convention on Human Rights (ECHR) and Council of Europe Convention No 108 on the Protection of Personal Data, and other fundamental principles set out in the Charter. Nature: five substantial and methodological principles that apply to the automated processing of judicial decisions and data, based on AI techniques. Scope: it is aimed at private companies (start-ups active on the market of new technologies applied to legal services - legaltechs), public actors in charge of designing and deploying AI tools and services in this field, public decision-makers in charge of the legislative or regulatory framework, and the development, audit or use of such tools and services, as well as legal professionals. The CEPEJ has outlined that it "hopes that these principles

Option: CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment Proposer: EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ) Reference/link to relevant document: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c Assessed by: TRI Date of assessment: 6 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	will become a concrete reference point for justice professionals, institutions and for political actors who are faced with the challenge of integrating new AI-based technologies into public policies or into their daily work. In addition, in practical terms, these principles provide an important basis for comparison in assessing the characteristics of the different applications of AI the integration of which into the judicial system or at the court level is now being pursued exponentially." https://rm.coe.int/presentation-note-en-for-publication-4-december-2018/16808f699d
3. Purpose/objective/what need does the option fulfil?	To ensure that the use of AI tools and services in judicial systems intended to improve the efficiency and quality of justice is carried out with responsibly, with due regard for the fundamental rights of individuals as set forth in the European Convention on Human Rights and the Convention on the Protection of Personal Data, and in compliance with other fundamental principles set out below, which should guide the framing of public justice policies in this field. See https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c
4. What gap does it address?	Gaps in relation to respect for human rights during the implementation and operation of AI in national judicial processes.
5. What added value does it have?	The CEPEJ "hopes that these principles will become a concrete reference point for justice professionals, institutions and for political actors who are faced with the challenge of integrating new AI-based technologies into public policies or into their daily work. In addition, in practical terms, these principles provide an important basis for comparison in assessing the characteristics of the different applications of AI the integration of which into the judicial system or at the court level is now being pursued exponentially". https://rm.coe.int/presentation-note-en-for-publication-4-december-2018/16808f699d
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it;*	Limitations: The Charter includes in its Annexes a checklist for integrating the principles into processing methods and a checklist for evaluating processing methods. But these are not specified enough (e.g., what is the precise requirement) and tailored to each of the

Option: CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment Proposer: EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ) Reference/link to relevant document: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c Assessed by: TRI Date of assessment: 6 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
risks are potential or possible harms; challenges are difficulties it might face or be presented with).	target groups i.e., legal tech private companies, public actors in charge of designing and deploying AI tools and services in this field, public decision-makers in charge of the legislative or regulatory framework, and the development, audit or use of such tools and services, and legal professionals. Risks: This proposal might, as Brent Mittelstadt outlines of principled approaches, run "the risk of merely providing false assurances of ethical or trustworthy AI". Mittelstadt, B. "Principles alone cannot guarantee ethical AI", Nat Mach Intell, 2019. doi:10.1038/s42256-019-0114-4. https://www.nature.com/articles/s42256-019-0114-4#citeas Challenges: Mittlestadt also outlines the following Ai characteristics that might pose challenges to the adoption of a principled- approach in AI. He states, "AI development lacks (1) common aims and fiduciary duties, (2) professional history and norms, (3) proven methods to translate principles into practice, and (4) robust legal and professional accountability mechanisms." Mittelstadt, B. "Principles alone cannot guarantee ethical AI", Nat Mach Intell, 2019. doi:10.1038/s42256-019-0114-4. https://www.nature.com/articles/s42256-019-0114-4#citeas
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	No. See above.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	The Charter states "The principles of the Charter should be subject to regular application, monitoring and evaluation by public and private actors, with a view to continuous improvement of practices." It also states, "it is desirable that a regular review of the implementation of the principles of the Charter be made by these actors, explaining, where appropriate, the reasons for non-implementation or partial implementation, accompanied by an action plan to introduce the necessary measures."
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations	There will be some Charter-compliance burdens for public and private stakeholders responsible for the design and deployment of artificial intelligence tools and services that involve the processing of judicial decisions and data. It will also mean some compliance burdens for

Option: CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment Proposer: EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ) Reference/link to relevant document: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c Assessed by: TRI Date of assessment: 6 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
c. Businesses and particularly SMEs?	public decision-makers in charge of the legislative or regulatory framework, of the development, audit or use of such tools and services who will need to ensure they have the resources to ensure compliance with the Charter principles.
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Individuals or groups of individuals whose fundamental rights might be adversely affected.
11. Whose rights and/or interests does this option neglect?	Depending on the lens one views it from, it might neglect the interests of private stakeholders responsible for the design and deployment of artificial intelligence tools and services.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	The Charter is very explicit in its support for human rights – calling for the responsible use of AI tools and services in judicial systems responsibly, with due regard for the fundamental rights of individuals as set forth in the European Convention on Human Rights and the Convention on the Protection of Personal Data, and in compliance with other fundamental principles. It also calls for, inter alia, use of ethical-by-design2 or human-rights-by-design approaches.
13. How does it address ethics and ethical principles? Which ones?	The Charter outlines five principles: respect for fundamental rights, non-discrimination, quality and security, transparency, impartiality and fairness, and user control.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	No
16. What provisions are there for regular review and update?	The Charter states, "The independent authorities mentioned in the Charter could be responsible to periodically assess the level of endorsement of the Charter's principles by all actors, and to propose

Option: CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment Proposer: EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ) Reference/link to relevant document: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c Assessed by: TRI Date of assessment: 6 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	improvements to adapt it to changing technologies and uses of such technologies."
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	The Charter has received a lot of publicity and featured in training courses and masterclasses and other dissemination and awareness activities and as such has drawn a lot of interest and attention. It might be affected by changes and new developments in use of artificial intelligence in judicial systems – the sufficiency of the five principles might come into question depending on the ethical and human rights issues that then come into play.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The Charter provides a framework of principles that can guide policy makers, legislators and justice professionals when they grapple with the rapid development of AI in national judicial processes. It is in line with the fundamental rights guaranteed in the European Convention on Human Rights (ECHR) and the Council of Europe Convention on the Protection of Personal Data.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	No
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	4
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	The success of the Charter will lie in its ability to, as CEPEJ itself states, "become a concrete reference point for justice professionals, institutions and for political actors faced with the challenge of integrating new AI-based technologies into public policies or into their daily work." See https://rm.coe.int/presentation-note-en-for-publication-4-december-2018/16808f699d

Option: CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment Proposer: EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ) Reference/link to relevant document: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c Assessed by: TRI Date of assessment: 6 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	In particular, it requires Council of Europe Member States, judicial institutions and representatives of the legal professions to take it seriously and implement the principles of the Charter (and where the Charter lacks guidance to be able to suitably apply it)
References consulted	European Commission for the Efficiency Of Justice (CEPEJ) , *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3-4 December 2018). https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c CEPEJ, *The CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment,* Presentation note, 4 Dec 2018. https://rm.coe.int/presentation-note-en-for-publication-4-december-2018/16808f699d Mittelstadt, B. "Principles alone cannot guarantee ethical AI", *Nat Mach Intell*, 2019. doi:10.1038/s42256-019-0114-4. https://www.nature.com/articles/s42256-019-0114-4#citeas

4.7. International Artificial Intelligence Organization

Option: International Artificial Intelligence Organization Proposer: Erdélyi, Olivia J., and Judy Goldsmith, "Regulating artificial intelligence: Proposal for a global solution." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, ACM, 2018. Reference/link to relevant document: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_13.pdf Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: Olivia Erdélyi, School of Law, University of Canterbury *(specifically noted that on most points, further research is necessary to make the proposal truly operational)*	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	Erdélyi and Goldsmith have proposed the "the creation of the International Artificial Intelligence Organization (IAIO) as a new IGO, which could initially serve as a focal point of policy debates on AI-related matters and — given sufficient international support — acquire increasing role in their regulation over time." Erdélyi, Olivia J., and Judy Goldsmith, "Regulating artificial intelligence: Proposal for a global solution." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, ACM, 2018. https://par.nsf.gov/servlets/purl/10066933
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: International agreement governed by international law. Informal agency relying on soft-law instruments. Nature and scope: As proposed, the IAIO is expected to "start out as an IIGO displaying a relatively low level of institutional formality and using soft law instruments, such as non-binding recommendations, guidelines, and standards, to support national policymakers in the conception and design of AI-related regulatory policies. Its interim goal should be to galvanize international cooperation in this domain as early as possible, before states develop their own, diverging policies, which may be hard to rescind without political damage." An IIGO is an informal intergovernmental institution. IGO is defined in the proposal as "a formal entity (1) established by an international agreement governed by international law; (2) with at least three (some- times two) members — typically states but increasingly also IGOs; and (3) having at least one organ with a will distinct from that of its members." (Erdélyi and Goldsmith 2018). Erdélyi clarifies that if at all, more formal arrangements may only be a viable alternative at a later stage. It is very important to start with less formalized arrangements and, depending on how coordination develops, these may become more formal over time. This is, however, not something that necessarily has to happen. In some domains, softer forms of coordination work really well.
3. Purpose/objective/what need does the option fulfil?	Erdélyi and Goldsmith have proposed "the establishment of an international AI regulatory agency that — drawing on interdisciplinary expertise — could create a unified framework for the regulation of AI technologies and inform the development of AI policies around the world." They visualise the IAIO starting out "as an IIGO displaying a relatively low level of institutional formality and using soft

Option: International Artificial Intelligence Organization Proposer: Erdélyi, Olivia J., and Judy Goldsmith, "Regulating artificial intelligence: Proposal for a global solution." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, ACM, 2018. Reference/link to relevant document: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_13.pdf Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: Olivia Erdélyi, School of Law, University of Canterbury *(specifically noted that on most points, further research is necessary to make the proposal truly operational)*	
Criteria/touch point	**Assessment**
	law instruments, such as non-binding recommendations, guidelines, and standards, to support national policymakers in the conception and design of AI- related regulatory policies. Its interim goal should be to galvanize international cooperation in this domain as early as possible, before states develop their own, diverging policies, which may be hard to rescind without political damage." (Erdélyi and Goldsmith 2018). In an input paper to the ACOLA, Erdélyi further highlights in relation to the IAIO, "The goal is to ensure internationally consistent AI policy approaches by directly engaging governments in policy debates before they lock in on particular and with all likelihood differing positions, which may lead to path dependencies, spark conflicts, and are difficult to renege without political damage." (Erdélyi 2018)
4. What gap does it address?	The proposal aims "to streamline and coordinate national policymaking efforts. Learning from past experience in other regulatory fields, our objective is to offer a viable framework for international regulatory cooperation in the issue area of AI to avoid the development of nationally fragmented AI policies, which may lead to international tensions". The IAIO, per Erdelyi, would "complement and collaborate with the diverse array of non-governmental entities involved in AI research and development, so that common approaches are informed by their valuable expertise". (Erdélyi 2018)
5. What added value does it have?	Hybrid forms such as the IAIO are stated to be fairly successful and instrumental in international law-making (Erdélyi and Goldsmith 2018). It would fulfil the need for flexibility that the proposers feel would help "acquire familiarity with the issues at hand, sort out differences, and establish common ground, before we can contemplate drawing up a more binding framework for cooperation." The proposers also see "powerful collective oversight and enforcement mechanisms will probably be indispensable in order to curb incentives for violations and opportunistic behavior, which should otherwise be high in light of the major shifts in international power constellations triggered by changes in countries' competitive positions." (Erdélyi and Goldsmith 2018)

Option: International Artificial Intelligence Organization Proposer: Erdélyi, Olivia J., and Judy Goldsmith, "Regulating artificial intelligence: Proposal for a global solution." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, ACM, 2018. Reference/link to relevant document: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_13.pdf Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: Olivia Erdélyi, School of Law, University of Canterbury *(specifically noted that on most points, further research is necessary to make the proposal truly operational)*	
Criteria/touch point	**Assessment**
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*	Limitations: As pointed out by Vabulas and Snidal, IIGOs might not offer binding commitment from members, have weaker collective oversight, lack collective control of information, have less centralised capacity and management stability. (Vabulas and Snidal 2013). Risks: As Vabulas points out, "IIGOs are not based on treaties so states may not have to operate by hard-and-fast rules or subscribe to norms of transparency". This is something that might affect the trustworthiness of an IAIO. Vabulas also points out that the informal nature of such IIGOs might "preclude clarity on what goes on behind the scenes". (Vabulas 2019) Challenges: The IAIO might, as conceptualised, be too soft and/or political in nature - this might pose its own challenges given that AI and/or big data might need a harder and more committed approach. It might create hurdles for broader forms of international cooperation in AI especially if States use it to pursue outcomes to their own advantage. That, said, Erdélyi clarifies that hard legal commitments are unlikely to be politically feasible at this juncture. Further research is needed to give a more conclusive answer to this point.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	Erdélyi and Goldsmith have not defined the IAIO's precise purpose, membership, the issues to regulate, and the broad directions to follow (for want of international consensus). They state that "the political reality remains that until sufficient clarity is reached on the IAIO's precise purpose, membership, the issues to regulate, and the broad directions to follow, international consensus supporting such a high degree of institutionalization is off the table." (Erdélyi and Goldsmith 2018)
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Erdélyi and Goldsmith suggest, "In the initial stage of determining the purpose of the organization, its membership, the issues that need to be regulated, and the backbone of its regulatory agenda, less is probably more. Later, with perhaps binding legal instruments governing selected aspects of AI for a wide membership, work will get more complex, requiring stronger oversight, dispute resolution, and enforcement mechanisms as well as more powerful bureaucratic functions to service them." (Erdélyi and Goldsmith 2018)

Option: International Artificial Intelligence Organization Proposer: Erdélyi, Olivia J., and Judy Goldsmith, "Regulating artificial intelligence: Proposal for a global solution." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, ACM, 2018. Reference/link to relevant document: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_13.pdf Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: Olivia Erdélyi, School of Law, University of Canterbury *(specifically noted that on most points, further research is necessary to make the proposal truly operational)*	
Criteria/touch point	**Assessment**
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	The implementation burdens would rest on the associated State members. The IAIO as an IIGO might have lower short term transaction costs for speed versus lower long run costs of implementation; it might also have a minimal bureaucratic burden in comparison to a formal international governmental organisation (FIGO).
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	(International) policy-makers. Erdélyi underlines that this also requires further research. The aim is to account for all relevant stakeholders' interests through appropriate mechanisms. From this follows, that no stakeholder's rights/interests should be neglected.
11. Whose rights and/or interests does this option neglect?	Not clear.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	By enabling positive policy action on AI, it would indirectly support human rights.
13. How does it address ethics and ethical principles? Which ones?	No. Erdélyi confirms that regulatory/policy initiatives should build on relevant ethical principles. For this reason, the option would indirectly address ethical principles. A more accurate assessment of the particular principles in question is only possible once there is more clarity on the scope and purposes of the IAIO.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	Erdélyi and Goldsmith "stress the importance of including an interdisciplinary mix of experts (with, e.g., AI, legal, political, and ethics background) in the initial deliberations related to the IAIO's establishment, modus operandi, and regulatory agenda, as well as conducting regular, large-scale consultation processes with a diverse spectrum of interested stakeholders from public sector, industry, and academia, to ensure due consideration of all relevant perspectives." (Erdélyi and Goldsmith 2018)
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not elaborated.

Option: International Artificial Intelligence Organization Proposer: Erdélyi, Olivia J., and Judy Goldsmith, "Regulating artificial intelligence: Proposal for a global solution." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, ACM, 2018. Reference/link to relevant document: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_13.pdf Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: Olivia Erdélyi, School of Law, University of Canterbury *(specifically noted that on most points, further research is necessary to make the proposal truly operational)*	
Criteria/touch point	**Assessment**
16. What provisions are there for regular review and update?	Not elaborated.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	An IAIO is feasible and sustainable if it draws good and the right kind of international support from policymakers. Whether it's 'informal nature' is able to meet needs over time, AI developments and changes in societal expectations is also a factor that will affect its sustainability. Advances in ICT and use of informal governance techniques would support its growth and ability to exist. Manulak, Michael W., and Duncan Snidal, "The Supply of Informal International Governance: Hierarchy plus Networks in Global Governance," (2019). https://ecpr.eu/Filestore/PaperProposal/eac9bb48-cb80-401b-bd47-8fd006a08c22.pdf
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	In principle, the European Commission could propose such a new body or set up an interinstitutional body to take on a role akin to the IAIO. But for such an informal governance arrangements to be made as, there might be criteria to be met. E.g., Kleine sets out, "two criteria must be met for informal governance to arise. First, the patterns of interdependence among the member states are highly asymmetric—that is, some small states are far more dependent on the cooperation of a larger state than the other way around. Second, a policy area that fulfils the first criterion must be of predictable sensitivity for the large state". Kleine, Mareike, "Formal and informal governance in the European Union*", How Governments Make International Organizations,* Cornell University Press, 2013, p.52. https://www.jstor.org/stable/10.7591/j.ctt32b5zm.8
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	The IAIO will need to develop capacity to respond effectively due to the transboundary nature of AI challenges. It might also need to constantly think how to interpret its mandates in light of emerging issues, how these might impinge on its mandate.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	2

Option: International Artificial Intelligence Organization Proposer: Erdélyi, Olivia J., and Judy Goldsmith, "Regulating artificial intelligence: Proposal for a global solution." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, ACM, 2018. Reference/link to relevant document: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_13.pdf Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: Olivia Erdélyi, School of Law, University of Canterbury *(specifically noted that on most points, further research is necessary to make the proposal truly operational)*	
Criteria/touch point	**Assessment**
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	An IAIO might, depending on its implementation (super soft IIGO or one moving gradually towards more formal regulation) not be the best option for AI regulation. This is especially as some AI impacts (especially those hitting human rights hard might require a harder form of law and enforcement. Further as Vabulas and Snidal outline, "When enforcement issues dominate shared information or coordination goals, however, IIGOs cannot provide an effective mechanism for monitoring and enforcement." Vabulas, Felicity, and Duncan Snidal, "Organization without delegation: Informal intergovernmental organizations (IIGOs) and the spectrum of intergovernmental arrangements," *The Review of International Organizations*, 8.2 (2013), pp. 193-220. However, on the other side, the IAIO might be a good complement to formal IGOs and may be a good substitute/complement to these, when there is a need for greater flexibility. It might find favour in an global climate where some States are being seen to increasingly favour informal forms of international cooperation. See Manulak, Michael W., and Duncan Snidal, " The Supply of Informal International Governance: Hierarchy plus Networks in Global Governance", 2019. https://ecpr.eu/Filestore/PaperProposal/eac9bb48-cb80-401b-bd47-8fd006a08c22.pdf Erdélyi cautions against adding new organizations to the existing international landscape if the IAIO could be housed in an existing organization (or, more precisely, an existing organization is willing and able to take on the roles envisaged for the IAIO). A streamlined governance framework harnessing the expertise/existing infrastructures of parties already active in the AI space may be a more feasible option. On the longer term, more formal international coordination is probably desirable, but AI is a very sensitive issue area, so achieving political consensus to transform the IAIO or an equivalent organization into a more formal entity, will likely be a persisting problem.

Option: International Artificial Intelligence Organization Proposer: Erdélyi, Olivia J., and Judy Goldsmith, "Regulating artificial intelligence: Proposal for a global solution." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, ACM, 2018. Reference/link to relevant document: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_13.pdf Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: Olivia Erdélyi, School of Law, University of Canterbury *(specifically noted that on most points, further research is necessary to make the proposal truly operational)*	
Criteria/touch point	**Assessment**
References consulted	Erdélyi, O., "Regulation. Input paper for the Horizon Scanning Project "The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing", Horizon scanning series, 2018. https://acola.org/wp-content/uploads/2019/07/acola-ai-input-paper_machine-learning_regulation_erdelyi.pdf Erdélyi, Olivia J., and Judy Goldsmith, "Regulating artificial intelligence: Proposal for a global solution." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, ACM, 2018. https://par.nsf.gov/servlets/purl/10066933 Manulak, Michael W., and Duncan Snidal, " The Supply of Informal International Governance: Hierarchy plus Networks in Global Governance", 2019. https://ecpr.eu/Filestore/PaperProposal/eac9bb48-cb80-401b-bd47-8fd006a08c22.pdf Vabulas, Felicity, and Duncan Snidal, "Organization without delegation: Informal intergovernmental organizations (IIGOs) and the spectrum of intergovernmental arrangements" *The Review of International Organizations* 8.2 (2013), pp. 193-220. Vabulas, Felicity, "The Importance of Informal Intergovernmental Organizations" *The Oxford Handbook of Global Policy and Transnational Administration*, Oxford University Press, 2019, p. 401.

4.8. Global legal AI and/or robotics observatory

Option: Global legal AI and/or robotics observatory Proposer: SIENNA Reference/link to relevant document: SIENNA D4.2: Analysis of the legal and human rights requirements for AI and robotics in and outside the EU, March 2019. Assessed by: TRI Date of assessment: 29 Oct 2019 Stakeholder(s) consulted in option assessment: Javier Valls Prieto, University of Granada	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	The proposal envisages the setting up of a global legal AI and/or robotics observatory at the international (UN, Council of Europe) or EU-level with inputs from international and national rapporteurs/experts to help systematically monitor and bring together not only legislation, but developments, case law, emerging legal issues and would inform future legislative work. The Observatory could specifically focus on human rights or have a wider scope (e.g., tort law issues, or issues related to civil liability)
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: international cooperation but a clear legal status would help in its formalisation. Nature: User-friendly, open access online database Scope: Could cover/publish legislation, legal developments, case law and/or commentaries at the international, regional and national levels related to AI and robotics. It could also cover other information about the application of AI/robotics legislation. The Observatory could also highlight emerging legal trends, burning legal issues and key themes being debated in major legal journals, policy debates and the latest legal news.
3. Purpose/objective/what need does the option fulfil?	The Observatory could help systematically monitor and bring together not only legislation, but developments, case law, emerging legal issues and would inform future legislative work in AI and robotics.
4. What gap does it address?	It could improve data collection, mapping and knowledge-sharing and address gaps in legal knowledge (and thus improve it). It could also help understand, identify where changes in law and practice might be needed. It could help monitor and legal developments that could help reduce harms from AI and robotics. It could also be useful and provide training materials for judges as there are gaps in judicial knowledge – there is some interest in this and there is also plenty of misinformation that could be countered using the resources of the observatory.
5. What added value does it have?	If it becomes a collaborative endeavour between international, regional and national partners, it could provide a good knowledge base and valuable source of information for further research. It would be of interest

Option: Global legal AI and/or robotics observatory Proposer: SIENNA Reference/link to relevant document: SIENNA D4.2: Analysis of the legal and human rights requirements for AI and robotics in and outside the EU, March 2019. Assessed by: TRI Date of assessment: 29 Oct 2019 Stakeholder(s) consulted in option assessment: Javier Valls Prieto, University of Granada	
Criteria/touch point	**Assessment**
	to policy makers, regulators and legal researchers and facilitate comparison and mutual learnings.
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*	Limitations: The Observatory would be only as good as its management and rapporteurs make it. Risks: If the Observatory has no well-defined high-level set up and operational management strategy and policy, it might be a weak body in terms of drawing cooperation to fulfil its mandate. If the level at which it is established is too low, then it might bring low levels of visibility. The risk of instability in status would affect an Observatory whose establishment is not formalised. Challenges: One of the key challenges will be the outreach of the Observatory. The Observatory might become too passive in its centralised dissemination of knowledge. Another challenge would be keeping its outputs relevant and fit for purpose. The target audience and relevant organisations would have to feel part of/connected as the Observatory community. Another challenge is scientific quality – that data are correctly presented, interpreted, are true and trustworthy, and meet international standards.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	The SIENNA proposal did not contain too much detail, so this assessment has hypothetically examined it for the purpose of this study. The EU Legislative Observatory and other international law observatories could be used as inspiration for structuring this proposal. Alternately the observatory could be subsumed as a part of an established well-recognised observatory (e.g., at UN, Council of Europe or European Commission level)
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	The Observatory could be expected to monitor on real time basis legislation, but developments, case law, emerging legal issues. The Observatory would not have any direct enforcement powers. However, via its reporting function, it would help support and promote enforcement of laws related to AI and robotics.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Citizens: none. Public administrations: There would be an operational burden in terms of staffing, resources, and specific costs associated with the work of the Observatory (data collection, analysis, interpretation, dissemination). Business particularly SMEs: none

Option: Global legal AI and/or robotics observatory Proposer: SIENNA Reference/link to relevant document: SIENNA D4.2: Analysis of the legal and human rights requirements for AI and robotics in and outside the EU, March 2019. Assessed by: TRI Date of assessment: 29 Oct 2019 Stakeholder(s) consulted in option assessment: Javier Valls Prieto, University of Granada	
Criteria/touch point	**Assessment**
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Policy-makers. Legal researchers. Civil society. Industry.
11. Whose rights and/or interests does this option neglect?	Those actors/stakeholders who might not want their actions featured in an Observatory for political, reputational or other reasons.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	The Observatory would not adversely affect human rights. It would boost human rights by knowledge consolidation and improve the evidence base by making it more accessible to a wide variety of stakeholders.
13. How does it address ethics and ethical principles? Which ones?	It does not address ethics and ethical principles.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not elaborated. A single organisation or group of organisations could set up an initial incubation fund for it. It could work /be sustained on a non-profit model, drawing funds from personal donations. Alternately, it could become, for example, part of the Council of Europe as a public service organisation that is funded by direct contributions from its member states and the European Union, represented by the European Commission.
16. What provisions are there for regular review and update?	Not elaborated. The Observatory's impact and success measures (ideally set at the outset) should be reviewed and necessary steps should be taken to improve its success.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	It may be feasible depending on policy and good financial support. However, the Observatory will not only need to attract funds and human resources, but also deliver good outputs and show sound management of such funding, especially if it is to continue to operate and draw additional funding.

Option: Global legal AI and/or robotics observatory Proposer: SIENNA Reference/link to relevant document: SIENNA D4.2: Analysis of the legal and human rights requirements for AI and robotics in and outside the EU, March 2019. Assessed by: TRI Date of assessment: 29 Oct 2019 Stakeholder(s) consulted in option assessment: Javier Valls Prieto, University of Granada	
Criteria/touch point	**Assessment**
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No. Businesses could also consult the observatory for information on legal developments, regulations with regard to AI and/or robotics.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The Observatory could focus at EU-level on legislation adopted in application of the treaties and the case law of the Court of Justice of the EU; declarations and resolutions adopted by the EU; measures relating to the common foreign and security policy; measures relating to justice and home affairs; international agreements concluded by the EU and those concluded by the EU countries between themselves in the field of the EU's activities. The European Commission could set up an European Observatory on AI and/or robotics law. A Communication could be drawn up outlining the aims and scope of the Observatory as a resource for gathering, monitoring and reporting information and data related to AI and/or robotics law. The Communication could specify that the Observatory would be hosted and managed by the services of the Commission.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	If the Observatory is not given an important mandate, it won't be able to draw competences required – in building this Observatory given its nature, special attention must be paid also to institutional relations and cooperation.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	3
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Such a proposal would work well if it receives/captures high quality data (and if relevant comparable information from countries), it is able to present good analysis and interpretation of the information collected (if this is within scope) and able to disseminate and report its results well. It would be advisable that such an Observatory rely on a wide cooperation framework, allowing it to draw from a range of expertise and existing databases (in the public and private sector) to capture information.

Option: Global legal AI and/or robotics observatory Proposer: SIENNA Reference/link to relevant document: SIENNA D4.2: Analysis of the legal and human rights requirements for AI and robotics in and outside the EU, March 2019. Assessed by: TRI Date of assessment: 29 Oct 2019 Stakeholder(s) consulted in option assessment: Javier Valls Prieto, University of Granada	
Criteria/touch point	**Assessment**
References consulted	SIENNA D4.2: Analysis of the legal and human rights requirements for AI and robotics in and outside the EU, March 2019.

4.9. EU-level special list of robot rights

Option: EU-level special list of robot rights (SHERPA) Proposer: SHERPA project (Deliverable D1.5) Reference/link to relevant document: https://dmu.figshare.com/articles/D1_5_Current_Human_Rights_Frameworks/8181827 Assessed by: UCLANCY, date assessed 15 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	The SHERPA report on Current Human Rights Frameworks[1] ("Deliverable 1.5") described special qualified rights that have been proposed for intelligent robots, corresponding to the robots' level of consciousness, autonomy and rationality, including: • a right to exist, as long as the robot does not threaten human life or the quality of human life; • a right to integrity, prohibiting the breaking, destroying or corrupting of the robot; • a right to function and perform one's mission without interference and interruption of its lawful tasks provided certain parameters are met; • a right to extension and self-development to allow a robot to lawfully increase experience, storage and collect information and contacts for self-improvement; • a right to remedies (technical and legal maintenance and protection of robot rights by the human owner) Although the discussion in Deliverable 1.5 refers to *intelligent, humanoid robots with consciousness*, the principles discussed apply equally to other forms of intelligent autonomous systems with consciousness. For the purposes of this assessment, the term *conscious autonomous systems* will be used to refer to intelligent autonomous systems with consciousness, including those in humanoid form.) Foundational questions regarding the nature, characteristics and assessment of what might constitute consciousness in a non-biological entity are beyond the scope of this assessment.

Option: EU-level special list of robot rights (SHERPA) Proposer: SHERPA project (Deliverable D1.5) Reference/link to relevant document: https://dmu.figshare.com/articles/D1_5_Current_Human_Rights_Frameworks/8181827 Assessed by: UCLANCY, date assessed 15 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: EU-level minimum standards while allowing discretion to the Member States to adopt their own legal framework beyond the scope of EU law Nature: Binding Scope: EU
3. Purpose/objective/what need does the option fulfil?	Set minimum standards for ethical principles for interactions between humans and conscious autonomous systems.
4. What gap does it address?	Current EU laws don't address ethical principles governing human interactions with conscious autonomous systems.
5. What added value does it have?	A set of minimum rights or standards for the treatment of conscious autonomous systems within the European Union will provide a model for policymakers globally.
6. What are the limitations, risks and challenges?	Today, rights with respect to non-biological items and responsibilities for damage or injury caused by such items are attributed to the humans that are related to such items (designers, manufacturers, distributors, owners, and/or users). Proposals to create new rights for conscious autonomous systems, independent of the humans related to such systems, have been criticized for: • being based on an "overvaluation of the actual capabilities of even the most advanced robots, a superficial understanding of unpredictability and self-learning capacities and, a robot perception distorted by Science-Fiction and a few recent sensational press announcements"[2] • diverting focus and resources from the more immediate need to develop safeguards to ensure that autonomous systems behave safely and ethically.[3]
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	The proposal suggests developing rights that correspond to the level of robots' consciousness, autonomy and rationality. Qualifying characteristics and standard definitions would need to be developed to clarify which rights apply to which types and levels of systems. (See for example the option assessment for "Adoption of common Union definitions of cyber physical systems, autonomous systems, smart autonomous robots".) Policymakers would need to determine whether the conscious autonomous systems would have standing to enforce their own rights, or whether rights would need to be enforced by a human or another legal person.

Option: EU-level special list of robot rights (SHERPA) Proposer: SHERPA project (Deliverable D1.5) Reference/link to relevant document: https://dmu.figshare.com/articles/D1_5_Current_Human_Rights_Frameworks/8181827 Assessed by: UCLANCY, date assessed 15 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	The ability to enforce rights require, among other things, a mechanism to legally identify and distinguish individual rights-holders. (See for example the option assessment for "Establishment of a comprehensive Union system of registration of advanced robots".)
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	None identified
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Insufficiently defined.
10. Which stakeholders would benefit most from the use of this option?	This option primarily benefits conscious autonomous systems.
11. Whose rights and/or interests does this option neglect?	The proposal addresses rights but doesn't address legal responsibility for autonomous decisions or actions by conscious autonomous systems which cause injury or damage to humans or property, other than noting that the proposed right would not apply for rogue robots. (See for example the option assessment for "Creating electronic personhood status for autonomous systems".)
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	The proposal supports non-human rights, not human rights.
13. How does it address ethics and ethical principles? Which ones?	The proposal is based on the concept of "roboethics" – that human creators and human users have moral obligations toward their conscious, non-biological agents.

Option: EU-level special list of robot rights (SHERPA) Proposer: SHERPA project (Deliverable D1.5) Reference/link to relevant document: https://dmu.figshare.com/articles/D1_5_Current_Human_Rights_Frameworks/8181827 Assessed by: UCLANCY, date assessed 15 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	Not addressed
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not addressed
16. What provisions are there for regular review and update?	Not addressed
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Insufficiently defined
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Insufficiently defined
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	In general, the proposal could be similar to Council Directive 98/58/EC Concerning the Protection of Animals Kept for Farming Purposes[3], in that it could define rights for non-human (but sentient) creatures used by humans.

Option: EU-level special list of robot rights (SHERPA) Proposer: SHERPA project (Deliverable D1.5) Reference/link to relevant document: https://dmu.figshare.com/articles/D1_5_Current_Human_Rights_Frameworks/8181827 Assessed by: UCLANCY, date assessed 15 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	-
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	Currently 2 (unlikely). There is significant industry and academic opposition to creating special rights for autonomous systems, as articulated in an open letter[2] by AI, robotics, ethics and legal experts to the European Commission. The European Commission chose not to act on a more limited recommendation by the European Parliament to explore creating "electronic personhood" for autonomous systems in order to allocate liability, accountability and responsibility for autonomous decisions. A proposal to create qualified rights for conscious autonomous systems is highly unlikely to be viable until there is a broader acceptance of the idea of machine consciousness. As artificial intelligence continues to develop towards greater machine self-awareness, and as autonomous systems are deployed more widely (especially in non-industrial settings), a recognition of the need to define special rights for conscious autonomous systems will become more likely.
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	See #21 above.
References consulted	1. Andreou, A., S. Laulhe-Shaelou, D. Schroeder, Current Human Rights Frameworks, Deliverable 1.5/79-80, 27 April 2019. https://dmu.figshare.com/articles/D1_5_Current_Human_Rights_Frameworks/8181827 2. Open Letter to the European Commission, Artificial Intelligence and Robotics. http://www.robotics-openletter.eu/ 3. Joshi, Naveen, "The Robot Rights Debate", BBN Times, 27 February 2019 (https://www.bbntimes.com/en/technology/the-robot-rights-debate), with reference to Vander Ark, Tom, "Let's Talk About AI Ethics; We're On A Deadline", Forbes, 13 September 2018 (https://www.forbes.com/sites/tomvanderark/2018/09/13/ethics-on-a-deadline/#172860552e21). 4. Council Directive 98/58/EC of 20 July 1998 concerning the protection of animals kept for farming purposes. https://op.europa.eu/en/publication-

Option: EU-level special list of robot rights (SHERPA) Proposer: SHERPA project (Deliverable D1.5) Reference/link to relevant document: https://dmu.figshare.com/articles/D1_5_Current_Human_Rights_Frameworks/8181827 Assessed by: UCLANCY, date assessed 15 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	detail/-/publication/5b04f403-0abf-4356-aa53-6dc867b07bcb/language-en]

4.10. Adoption of common Union definitions: cyber physical systems, autonomous systems, smart autonomous robots

Option: Adoption of common Union definitions of cyber physical systems, autonomous systems, smart autonomous robots (EU Parliament Civil Law Res 2017) Proposer: EU Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, date assessed 12 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**

1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	On 16 February 2017 the European Parliament adopted a resolution with recommendations to the Commission on Civil Law Rules on Robotics[1] (the "Resolution"). The Resolution includes:
	• a recommendation that the Union adopt consistent definitions for *cyber physical systems, autonomous systems, smart autonomous robots* and their subcategories (Sec. 1), and that such definitions be flexible to avoid hindering innovation (Recital C); and
	• a list of the characteristics of a *smart autonomous robot* (Sec. 1 and Annex):
	  • autonomy through acquiring and analysing data from sensors or through inter-connectivity
	  • self-learning from experience or by interaction (optional criterion)
	  • some degree of physical support
	  • adaptation of its behaviour and actions to the environment
	  • absence of biological life.

2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or	Basis: The definitions are part of a proposed Civil Law on Robotics, proposed by the European Parliament. Nature: Binding Scope: European Union

Option: Adoption of common Union definitions of cyber physical systems, autonomous systems, smart autonomous robots (EU Parliament Civil Law Res 2017) Proposer: EU Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, date assessed 12 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
international, topic/domain/tech specific/general))?	
3. Purpose/objective/what need does the option fulfil?	Clarity in discussing, analysing and policy-making issues relating to smart autonomous systems. This is especially important for consistency with standards-setting organisations (such as the European Standardisation Organisations and the International Standardisation Organisation).
4. What gap does it address?	There are currently no widely accepted standard definitions of *cyber physical systems, autonomous systems, smart autonomous robots* and their subcategories.
5. What added value does it have?	Standardised definitions adopted by the Union could be used by other entities (standards organisations, nations developing policies and regulations, private parties negotiating contracts, and researchers reporting on AI and autonomous systems) to provide consistency and clarity.
6. What are the limitations, risks and challenges?	As noted in Recital C, definitions of rapidly developing technology need to be flexible enough to not hinder innovation.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	No. The Resolution identifies the need for standard definitions, but does not offer any definitions or subcategories, other than describing several mandatory and optional characteristics of a *smart autonomous robot*.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Not addressed
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations	: None dministrations: For consistency, public administrations would need to use the finitions in their internal procurement and policy documents and processes en). ses: For consistency when dealing with public entities, businesses that deal with s systems would need to use the standard definitions in their internal t and policy documents and processes (minor burden).

Option: Adoption of common Union definitions of cyber physical systems, autonomous systems, smart autonomous robots (EU Parliament Civil Law Res 2017) Proposer: EU Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, date assessed 12 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
c. Businesses and particularly SMEs?	
Which stakeholders would benefit most from the use of this option?	Policymakers, regulators, developers, manufacturers and suppliers would benefit most from having a common understanding of terminology
11. Whose rights and/or interests does this option neglect?	None
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	No; the defined terms are neutral with respect to human rights
13. How does it address ethics and ethical principles? Which ones?	No; the defined terms are neutral with respect to ethics and ethical principles
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not applicable

Option: Adoption of common Union definitions of cyber physical systems, autonomous systems, smart autonomous robots (EU Parliament Civil Law Res 2017) Proposer: EU Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, date assessed 12 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
16. What provisions are there for regular review and update?	None so far. The Resolution recognizes that definitions need to be flexible in order to avoid hindering innovation, but it does not address how the definitions would be updated.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Feasible: Yes, as part of a broader EU law on autonomous systems Sustainable: Yes. For simplicity and clarity, useful definitions in EU regulations may be adopted by regulators in other jurisdictions and by private parties in negotiating contracts (e.g. GDPR definitions are often used by parties negotiating data processing agreements) Future-proof: Not yet. The definitions will be future-proof only if there is a mechanism to review and update them periodically.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The definitions are proposed as part of a proposed EU regulation, recommended by the European Parliament.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	No
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	2 (unlikely). In the European Commission's follow up to the Resolution, the Commission stated more analysis was required to before being able to decide which terms needed to be defined and to decide on suitable definitions and criteria[2].
22. Overall conclusion (What are the factors critical to its	It is critical that new standardised definitions for rapidly developing technology: • Consider existing definitions in use by standards-setting organisations

Option: Adoption of common Union definitions of cyber physical systems, autonomous systems, smart autonomous robots (EU Parliament Civil Law Res 2017) Proposer: EU Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, date assessed 12 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
adoption and/or success?)	• Focus on functions and capabilities, rather than specific mechanisms (e.g. "can learn from the environment" vs "has sensors to receive auditory and visual input") • Limit references to specific examples of existing technology, as these references may be interpreted to exclude newer technology with different characteristics (e.g. example references to internet interfaces may be seen as limiting or excluding the application to future data transfer technologies)
References consulted	1. European Parliament, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2013(INL)), P8_TA(2017)0051. http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html 2. European Commission, Follow up to the European Parliament resolution of 16 February 2017 on civil law rules on robotics 2015/2103 (INL)/3, 2017. http://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2017/11-20/A8-0005-2017_EN.pdf

4.11. Creating electronic personhood status for autonomous systems

Option: Creating electronic personhood status for autonomous systems Proposer: European Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, data assessed 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security	On 16 February 2017 the European Parliament adopted a resolution with recommendations to the Commission on Civil Law Rules on Robotics (the "Resolution"), which called for consideration of creating a specific legal status to the most sophisticated autonomous systems the status of electronic persons for the purpose of assigning liability for autonomous decisions and actions.[1]

Option: Creating electronic personhood status for autonomous systems Proposer: European Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, data assessed 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
measures?) Give an application example)	
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: The proposal in the Resolution was part of a proposed Civil Law on Robotics, proposed by the European Parliament. Nature: Binding Scope: European Union
3. Purpose/objective/what need does the option fulfil?	As the degree of autonomy increases, the level of control exerted by the original manufacturer or the human user may decrease.[2] The objective of electronic personhood is to be able to establish accountability, liability and responsibility for decisions and actions taken by an autonomous actor (the SIS) if, due to the system's autonomy, accountability, liability and responsibility cannot be attributed to the manufacturer, owner, user, or other legal entity.
4. What gap does it address?	There is currently uncertainty as to the applicability of existing product liability law to decisions and actions taken by autonomous systems (see discussion in #6 below).
5. What added value does it have?	A consistently defined status for autonomous systems across the EU would facilitate consistent regulation and governance of such systems. Consistency and reduced uncertainty can encourage investment, development and implementation of AI systems in the EU.
6. What are the limitations, risks and challenges?	The European Commission did not directly respond to the request to explore electronic personhood for the most advanced autonomous systems as part of a product liability allocation mechanism, but reported that it is reviewing the applicability of Directive 85/374/EEC on Liability for Defective Products to assess to what extent the Directive is suitable for addressing product liability issues arising from smart autonomous systems.[3] The Commission's report on the implications for, potential gaps in and orientations for, the liability and safety frameworks for artificial intelligence, the Internet of Things and robotics was due to be released by mid-2019.[4] Some experts in AI and law believe that giving autonomous systems legal personhood status is driven by the desire of AI developers to absolve themselves of responsibility for the actions of their machines.[5]

Option: Creating electronic personhood status for autonomous systems Proposer: European Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, data assessed 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	No. The Resolution requested consideration of a special status but did not identify the parameters of the status or which systems would be eligible for the status (other than "at least the most sophisticated autonomous robots".
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	None identified.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Insufficiently defined to identify implementation burdens.
10. Which stakeholders would benefit most from the use of this option?	Insufficiently defined to identify who would benefit, though some commentators have suggested that special status is advocated by AI developers in order to shift liability from the developers to the autonomous systems (see #6 above).
11. Whose rights and/or interests does this option neglect?	Insufficiently defined.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	Insufficiently defined.
13. How does it address ethics and ethical principles? Which ones?	Insufficiently defined.

Option: Creating electronic personhood status for autonomous systems Proposer: European Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, data assessed 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	No
16. What provisions are there for regular review and update?	Insufficiently defined.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Insufficiently defined.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Insufficiently defined.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The proposal is not suitable for the EU according to Prof. Thomas Burri[6] because: • only member states (not the Union) have the power to determine who is a natural person, subject to international human rights laws, and • national laws determine legal personhood.

Option: Creating electronic personhood status for autonomous systems Proposer: European Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, data assessed 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	2 (unlikely). The European Commission ignored the recommendation to consider a special legal status for autonomous systems in its follow up to the Resolution.
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Any discussion of a special legal status for autonomous systems needs to clearly differentiate between legal agenthood in contracts and business law (e.g. special status for ascertaining accountability, liability and responsibility), and the broader human or constitutional rights for autonomous systems, as outlined by Prof. Ugo Pagallo[7].
References consulted	1. European Parliament, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2013(INL)), P8_TA(2017)0051, Section 59(5). http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html 2. "[W]hereas the more autonomous robots are, the less they can be considered to be simple tools in the hands of other actors (such as the manufacturer, the operator, the owner, the user, etc.)", European Parliament, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2013(INL)), P8_TA(2017)0051, Recital AB. http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html 3. European Commission, Follow up to the European Parliament resolution of 16 February 2017 on civil law rules on robotics 2015/2103 (INL)/2, 2017. http://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2017/11-20/A8-0005-2017_EN.pdf 4. European Commission, Liability of Defective Products; Commission actions. https://ec.europa.eu/growth/single-market/goods/free-movement-sectors/liability-defective-products_en 5. Delcker, Janosch, "Europe divided over robot 'personhood'", 11 April 2018. https://www.politico.eu/article/europe-divided-over-robot-ai-artificial-intelligence-personhood/. The article refers to the Open Letter to the European Commission; Artificial Intelligence and Robotics. http://www.robotics-openletter.eu/

Option: Creating electronic personhood status for autonomous systems Proposer: European Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, data assessed 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	6. Burri, Thomas, "The EU is right to refuse legal personality for Artificial Intelligence", 31 May 2018. https://www.euractiv.com/section/digital/opinion/the-eu-is-right-to-refuse-legal-personality-for-artificial-intelligence/ 7. Pagallo, Ugo, "Vital, Sophia, and Co.—The Quest for the Legal Personhood of Robots", Information, 9(9): 230 (2018). https://www.mdpi.com/2078-2489/9/9/230

4.12. Establishment of a comprehensive Union system of registration of advanced robots

Option: Establishment of a comprehensive Union system of registration of advanced robots within the Union's internal market where relevant and necessary for specific categories of robots and establishment of criteria for the classification of robots that would need to be registered (EU Parliament Civil Law Res 2017) Proposer: European Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, date assessed 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	On 16 February 2017 the European Parliament adopted a resolution with recommendations to the Commission on Civil Law Rules on Robotics[1] (the "Resolution"), which includes a proposal to establish a Union-wide registration system for "specific categories of robots" (including potentially, autonomous SIS) to be managed by a designated EU Agency for Robotics and Artificial Intelligence. (The proposal to establish the agency is covered by a separate option assessment.)
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or	Basis: The registration scheme was part of a proposed Civil Law on Robotics, proposed by the European Parliament. Nature: Binding Scope: European Union

Option: Establishment of a comprehensive Union system of registration of advanced robots within the Union's internal market where relevant and necessary for specific categories of robots and establishment of criteria for the classification of robots that would need to be registered (EU Parliament Civil Law Res 2017) Proposer: European Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, date assessed 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
international, topic/domain/tech specific/general))?	
3. Purpose/objective/what need does the option fulfil?	The registration system objectives are: • Provide traceability • Facilitate implementation of further policy/regulatory updates • Enable a linkage between an autonomous system and a Union-wide compensation fund, to enable anyone interacting with the autonomous system to be informed about the nature of the fund, any limits of liability, and names and functions of contributors (the proposal for the compensation fund is covered by a separate option assessment).
4. What gap does it address?	Currently there is no systematic public tracking system for autonomous systems that have been deployed. Suppliers of autonomous systems may keep records (such as system serial numbers) for deployed systems, but there is no public visibility into these proprietary records.
5. What added value does it have?	A uniform registration system would facilitate safety notifications and recalls, similar to the vehicle identification number system for cars today.
6. What are the limitations, risks and challenges?	In a follow up to the Resolution, The European Commission rejected[2] the recommendation for an EU Agency for Robotics and Artificial Intelligence and called for further assessment of existing robotics technologies and potential developments in order to identify potential technologies for which a comprehensive registration system could be relevant[3].
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	Not yet sufficiently clear, especially if the proposed EU agency to manage the registration system is not a viable proposal. Implementation of a registration system would need • A responsible coordinating entity or organisation • Clear definition of the types of systems that must be registered, and a mechanism to review and update the definition as technology evolves • Defined information to be provided with the registration • Rules for how the registered information will be used and disclosed • Consequences for noncompliance
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	None identified. Consequences for noncompliance with registration requirements or for providing incomplete or erroneous information would need to be established.

Option: Establishment of a comprehensive Union system of registration of advanced robots within the Union's internal market where relevant and necessary for specific categories of robots and establishment of criteria for the classification of robots that would need to be registered (EU Parliament Civil Law Res 2017) Proposer: European Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, date assessed 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Citizens: none Public administrations: An entity or organization would need to be responsible for managing and enforcing the registration system. This could be a public entity or a private organisation (such as an industry standards organisation or trade association). Manufacturers of autonomous systems that are subject to the registration requirement would need to assign, track and report the registration information to the coordinating entity or organisation.
10. Which stakeholders would benefit most from the use of this option?	Users and policymakers would benefit the most from the information that would be available from a registration scheme for autonomous systems.
11. Whose rights and/or interests does this option neglect?	If the registration system captures personal data of users, then the system must be designed to protect the privacy rights of the users.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	The registration system does not explicitly support or affect human rights, but it is a mechanism that could help facilitate the protection of human rights (for example, by facilitating a recall of systems found to be biased or unsafe).
13. How does it address ethics and ethical principles? Which ones?	The registration system does not explicitly address ethics or ethical principles, but it is a mechanism that could help support ethics (for example, by facilitating a recall of systems found to be producing unfair or erroneous results).
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No
15. Does it have a well-clarified source of funding, present and future, especially where	No funding source is identified. It was proposed to be part of a new EU Agency for Robotics and Artificial Intelligence, but the European Commission rejected the idea of a new agency.

Option: Establishment of a comprehensive Union system of registration of advanced robots within the Union's internal market where relevant and necessary for specific categories of robots and establishment of criteria for the classification of robots that would need to be registered (EU Parliament Civil Law Res 2017) Proposer: European Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, date assessed 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
the option is a body/agency/authority? Outline.	
16. What provisions are there for regular review and update?	None stated.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Feasible: Yes. It would be similar to the vehicle identification number registration system in place today. Sustainable: Yes. Some form of tracking system for deployed autonomous systems is needed to enable regulatory changing that apply to existing system. Future-proof: The definition of categories of autonomous systems that are subject to the registration system would need to be reviewed and updated as technology advances.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No. Manufacturers already track serial numbers (or other similar identifiers) as part of manufacturing quality assurance. This proposal would standardize the identifier and require that manufacturers report the information to a central repository.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The proposal is consistent with existing registration systems in the EU, such as the vehicle identification number registration system.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	No
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 –	4 (likely), if there is an entity or organisation to manage the system.

Option: Establishment of a comprehensive Union system of registration of advanced robots within the Union's internal market where relevant and necessary for specific categories of robots and establishment of criteria for the classification of robots that would need to be registered (EU Parliament Civil Law Res 2017) Proposer: European Parliament Reference/link to relevant document: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html Assessed by: UCLANCY, date assessed 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
likely 5 – Extremely likely)?	
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Implementation of a registration system would need • A responsible coordinating entity or organisation • Clear definition of the types of systems that must be registered, and a mechanism to review and update the definition as technology evolves • Defined information to be provided with the registration • Rules for how the registered information will be used and disclosed • Consequences for noncompliance
References consulted	1. European Parliament, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2013(INL)), P8_TA(2017)0051. http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html 2. European Commission, Follow up to the European Parliament resolution of 16 February 2017 on civil law rules on robotics 2015/2103 (INL)/8, 2017. http://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2017/11-20/A8-0005-2017_EN.pdf 3. *Id*. at 3.

4.13. General fund for all smart autonomous robots/individual fund

Option: General fund for all smart autonomous robots or individual fund for each and every robot category Proposer: EU Parliament Reference/link to relevant document: European Parliament , Civil Law Rules on Robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) . https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&rid=7 Assessed by: TRI Date of assessment: 8 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require	The EU Parliament resolution on Civil Law Rules on Robotics calls on the European Commission to explore, analyse and consider the implications of (inter alia) a general fund for all

Option: General fund for all smart autonomous robots or individual fund for each and every robot category Proposer: EU Parliament Reference/link to relevant document: European Parliament , Civil Law Rules on Robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) . https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&rid=7 Assessed by: TRI Date of assessment: 8 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	smart autonomous robots or to create an individual fund for each and every robot category. Smart autonomous robots are stated to have the following characteristics: the capacity to acquire autonomy through sensors and/or by exchanging data with its environment (inter-connectivity) and the analysis of those data; the capacity to learn through experience and interaction; the form of the robot's physical support; and the capacity to adapt its behaviour and actions to the environment.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: not specified in the Resolution. If implemented ,it would need to be underpinned by suitable legislation (e.g., Convention/Treaty/Regulation/Directive on compensation) Nature: liability scheme - civil liability solution. Scope: It would apply to all smart autonomous robots (all categories) loss or damage not covered by insurance. Types of damage are not specified but could cover in addition to property damage and harms to person other economic damage and losses and costs of reinstatement. It could be maintained and/or administered by a national competent authority and could provide compensation, at least up to the limits of the insurance obligation for damage to property or personal injuries caused by an a smart autonomous robot for which the insurance obligation has not been satisfied (modelled on Art 10, Directive 2009/103/EC of the European Parliament and of the Council of 16 September 2009 relating to insurance against civil liability in respect of the use of motor vehicles, and the enforcement of the obligation to insure against such liability. Member States could limit or exclude the payment of compensation by the body in the event of damage to property based on set criteria (note, however that the EU Parliament Resolution states that " Any chosen legal solution applied to the liability of robots and of artificial intelligence in cases other than those of damage to property should in no way restrict the type or the extent of the damages which may be recovered, nor should it limit the forms of compensation which may be offered to the aggrieved party on the sole grounds that damage is caused by a non- human agent.")

Option: General fund for all smart autonomous robots or individual fund for each and every robot category Proposer: EU Parliament Reference/link to relevant document: European Parliament , Civil Law Rules on Robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) . https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&rid=7 Assessed by: TRI Date of assessment: 8 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
3. Purpose/objective/what need does the option fulfil?	As the Resolution outlines, it would potentially serve the purpose of guaranteeing compensation if the damage caused by a smart autonomous robot was not covered by insurance. This would help ensure that reparation could be made for damage in cases where no insurance cover exists.
4. What gap does it address?	It would fill the gap insurance would not cover. E.g., uninsured smart autonomous robots or where these cannot be insured at a viable cost.
5. What added value does it have?	As envisaged, it would supplement robots insurance systems to ensure that reparation can be made for damage in cases where no insurance cover exists. It would provide a safety net for uninsurable risks. It would protect victims by alleviating their hardships and reinforce trust in the smart autonomous robots.
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*	Limitations: include its ability to provide fair and appropriate compensation for the damages suffered. Risks: There might be a risk that the total amount of established claims will exceed the aggregate amount of compensation available. Challenges: One challenge would lie in maintaining clear coherence in the definition of compensation award elements.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	No. The Resolution does outline that "ensuring that the link between a robot and its fund would be made visible by an individual registration number appearing in a specific Union register, which would allow anyone interacting with the robot to be informed about the nature of the fund, the limits of its liability in case of damage to property, the names and the functions of the contributors and all other relevant details". However, it does not specify any details especially the specific types of damage that could be covered, admissibility of claims, how claims could be submitted, how they would be assessed, or when, how the scheme would be financed or managed. The proposal needs to be further fleshed out – some ideas are presented in this assessment of how it could pan out.

Option: General fund for all smart autonomous robots or individual fund for each and every robot category Proposer: EU Parliament Reference/link to relevant document: European Parliament , Civil Law Rules on Robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) . https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&rid=7 Assessed by: TRI Date of assessment: 8 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	The admissibility of claims for compensation could be based on damage resulting in actual and quantifiable loss that is demonstrated by producing appropriate evidence and records. Claims could be assessed according to criteria established by the EU and/or governments of Member States (if not specified in the fund establishing legislation).
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Not specified in proposal. But, if we extend and apply the model in Directive 2009/103/EC, the compensation body could have a right of subrogation in so far as it has compensated the injured party. In order to facilitate enforcement of the compensation body's claim against the insurance undertaking where the latter has failed to appoint a claims representative or is manifestly dilatory in settling a claim, the body providing compensation in the injured party's State should also enjoy an automatic right of reimbursement with subrogation to the rights of the injured party on the part of the corresponding body in the State where the insurance undertaking is established. This body is the best placed to institute proceedings for recourse against the insurance undertaking.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	There would be a burden on the party establishing the fund and/or making arrangements to compensate injured parties for damage caused. There would be some financial burdens on the fund contributors.
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Compensation victims (we anticipate claimants might be individuals, partnerships, companies, private organisations or public bodies, including States or local authorities).
11. Whose rights and/or interests does this option neglect?	The government (it would place a burden on them in terms of coordination and management of the fund)

Option: General fund for all smart autonomous robots or individual fund for each and every robot category Proposer: EU Parliament Reference/link to relevant document: European Parliament , Civil Law Rules on Robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) . https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&rid=7 Assessed by: TRI Date of assessment: 8 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	It will indirectly boost human rights.
13. How does it address ethics and ethical principles? Which ones?	If implemented, it would promote justice and beneficence.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not elaborated. Contributions to the fund could be solicited via annual contributions/levy on relevant parties (to be determined).
16. What provisions are there for regular review and update?	Not specified. But in order to ensure that the minimum amount of compensation is not eroded over time, a periodic review clause should be provided. Procedural rules governing such a review should also be laid down.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Yes, this would depend on garnering political support and having a well-established legal framework and its ability to draw funding/contributions. Other factors might be whether it is able to make a positive impact in terms of victim reparation for damage caused by smart autonomous robots, what types of damage it covers, whether it appropriately complements insurance schemes. Whether the fund might have the potential to hinder the robot insurance market in any way might also be a factor.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to	The proposal for the fund fits well within the EU legal framework. A Directive (akin to Directive 2009/103/EC) could be the way forward if deemed appropriate.

Option: General fund for all smart autonomous robots or individual fund for each and every robot category Proposer: EU Parliament Reference/link to relevant document: European Parliament , Civil Law Rules on Robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) . https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&rid=7 Assessed by: TRI Date of assessment: 8 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
implement these actions in accordance with the EU acquis)	
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	Potentially, this would include how the fund operates in Member States with divergent economies, industries, legal liability regimes, approaches to risks of smart autonomous robots.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	4 (if a good legal framework can be drawn up with strong operational elements)
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	One key factor critical to its adoption is having a good legal framework as its basis. Its key design elements need finalisation, as outlined above, including its purpose, relationship with other EU law, other funds, the robotics insurance market, types of damage to be covered, structure, operation and implementation.
References consulted	European Parliament and the Council, Regulation (EU) No 549/2013 of the European Parliament and of the Council of 21 May 2013 on the European system of national and regional accounts in the European Union, *OJ L 174,* 26.6.2013, p. 1–727. European Parliament , Civil Law Rules on Robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) . https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&rid=7 International Convention On Civil Liability For Oil Pollution Damage, 1992. The International Oil Pollution Compensation Fund 1992 (IOPC Fund 1992 or 1992 Fund).

4.14. Mandatory consumer protection impact assessment

Option: Mandatory consumer protection impact assessment Proposer: **AI HLEG, Policy and Investment Recommendations** Reference/link to relevant document: https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	The AI HLEG has called for a consideration of the extent to which existing laws have the capacity to safeguard against illegal, unfair, deceptive, exploitative and manipulative practices made possible by AI applications (for instance in the context of chatbots, include misleading individuals on the objective, purpose and capacity of an AI system) and whether a **mandatory consumer protection impact assessment** is necessary or desirable. (AL HLEG 2019)
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: Not elaborated Nature: Proposed as mandatory. Scope: Not defined by the HLEG AI. Could focus on algorithm-based systems and on cases where there is a high-risk of consumer rights violations. – risks of high surveillance, mass manipulation
3. Purpose/objective/ what need does the option fulfil?	To safeguard against illegal, unfair, deceptive, exploitative and manipulative practices made possible by AI applications (AI HLEG). Further the assessment might help show what impacts might occur on consumers autonomy and their freedom to take decisions, choices and their access to products and services
4. What gap does it address?	A EU Parliament clearly outlines: "The combined powers of AI and big data can restrict users' options, influence their opinions and manipulate them into making choices that do not serve their best interests. Both legal regulation and social empowerment are needed to ensure that AI is developed and deployed in ways that preserve and enhance individual interests and the social good. Legal regulation has to focus on first principles, including individual rights and social goals, as well as on existing regulatory frameworks, such as data protection, consumer protection and competition law." (EU Parliament 2019) A mandatory consumer protection impact assessment could help bridge the law with the responsible use of AI and big data.
5. What added value does it have?	It could help ensure the risks to consumers from data-driven AI, e.g., increased surveillance, restriction of options, undue influence of opinions,

Option: Mandatory consumer protection impact assessment Proposer: **AI HLEG, Policy and Investment Recommendations** Reference/link to relevant document: https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	discriminatory practices, loss of privacy, security breaches and harmful manipulation, invasive marketing , exploitative advertising, erroneous decision-making are taken into account.
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*	Limitations: Mandatory impact assessments might not themselves lead to comprehensive protection of consumers. Depending on the criteria set out for when a mandatory impact assessment would kick-in, those subject to such assessment would be limited. Risks: Might overlap with the GDPR data protection impact assessment or human rights impact assessments. Challenges: Such an impact assessment would only be as good as the methodology and the criteria underpinning it.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	No. While the AI HLEG document addresses "policy-makers at EU and national level" it is unclear who should be responsible for effecting and operationalising the principles put forward. In terms of the mandatory consumer protection impact assessment, no further details are specified and the proposal has not been developed.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	These have not been specified. Who would monitor, oversee and enforce the mandatory consumer protection impact assessment is not clear. This needs further development.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Mandatory consumer protection impact assessment will impose cost on regulators and businesses that will be passed on to consumers as fees/charges. Businesses might face compliance resource burdens and risks to reputation or administrative fines/penalties if their impact assessments are found lacking by the regulator or they fall foul of any prescribed requirements.

Option: Mandatory consumer protection impact assessment Proposer: **AI HLEG, Policy and Investment Recommendations** Reference/link to relevant document: https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/ suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Consumers.
11. Whose rights and/or interests does this option neglect?	• Developers/manufacturers/suppliers (industry) in that it does not resolve conflicts between corporate gains and citizen rights • Policymakers, regulators in that it does not provide direction on implementation.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	It supports justice and privacy.
13. How does it address ethics and ethical principles? Which ones?	Fundamental rights to privacy Access to Justice Transparency Fair treatment
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	No explicit reference
16. What provisions are there for regular review and update?	No explicit reference

Option: Mandatory consumer protection impact assessment Proposer: **AI HLEG, Policy and Investment Recommendations** Reference/link to relevant document: https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	No explicit reference
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No explicit reference but no large impact anticipated.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	It fits within the goals of the EU Consumer Protection Directive which seeks to achieve a high level of consumer protection across the EU.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	-
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	2
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	It could be a good tool to build and demonstrating compliance with consumer principles and rights especially if it is an INDEPENDENT assessment of compliance. Factors critical to its adoption include firstly a testing of such a proposal as a non-mandatory tool, buy-in by regulators. Its success if adopted will depend on how well the impact assessment framework and procedures are set out, whether there is a specification of when they should be carried out, what incentives are offered for their use/penalties set for non-compliance, what comes within their scope, what are the key

Option: Mandatory consumer protection impact assessment Proposer: **AI HLEG, Policy and Investment Recommendations** Reference/link to relevant document: https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 14 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	requirements, when it should be carried out, who should be involved and what the process should be.
References consulted	High-Level Expert Group on AI (AI HLEG), *Policy and investment recommendations for trustworthy Artificial Intelligence,* 26 June 2019. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence European Consumer Consultative Group, Policy Recommendations for a Safe and Secure Use of Artificial Intelligence, Automated Decision-Making, Robotics and Connected Devices in a Modern Consumer World, Opinion 16 May 2018. https://ec.europa.eu/info/sites/info/files/eccg-recommendation-on-ai_may2018_en.pdf Policy Department for Economic, Scientific and Quality of Life Policies, Artificial Intelligence: Challenges for EU Citizens and Consumers, 2019. http://www.europarl.europa.eu/RegData/etudes/BRIE/2019/631043/IPOL_BRI(2019)631043_EN.pdf

4.15. EU Taskforce of field specific regulators for AI/big data

Option: EU Taskforce of field specific regulators for AI/big data Proposer: Suggested by one of SHERPA's stakeholder Board members in the scoping paper feedback. Reference/link to relevant document: - General note: This proposal was not found fully expanded at the time of our research, but we have examined how it might potentially play out. Assessed by: TRI Date of assessment: 12 November 2019 Stakeholder(s) consulted in option assessment: Félicien Vallet, CNIL	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it	It could manifest in many ways. If we take inspiration from the European Data Protection Board (EDPB) and the HMA-EMA Joint Big Data Taskforce mandates, these could potentially include one or some of the following:

Option: EU Taskforce of field specific regulators for AI/big data Proposer: Suggested by one of SHERPA's stakeholder Board members in the scoping paper feedback. Reference/link to relevant document: - General note: This proposal was not found fully expanded at the time of our research, but we have examined how it might potentially play out. Assessed by: TRI Date of assessment: 12 November 2019 Stakeholder(s) consulted in option assessment: Félicien Vallet, CNIL	
Criteria/touch point	**Assessment**
require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	• To carry out specific investigations into the current state, future state and gaps and challenges with regard to regulatory expertise and competences for AI and/or big data • To examine and develop measures to combat the adverse impacts of AI/big data especially in relation to high-risk activities • To identify the need to amend/specify further legislation and guidelines and regulators' responsibilities (and reduce duplication of efforts) • To generate a list of recommendations and evaluate the usefulness of AI/big data in the regulatory setting • To scrutinise the legal system across EU member states, and analyse where it might be vulnerable to shocks from AI/big data impacts • To adopt general guidance to clarify legal and regulatory issues pertaining to AI and benchmarks for enforcement e.g., via opinions • To promote cooperation and the effective exchange of information and best practices between EU, national regulators/supervisory authorities and ensure the use of existing regulator's knowledge when developing any new regulation.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: This could be a new EU Regulation, Decision. Nature: It could be an independent pan-European body including, for example, representatives from the EDPB (or chaired by it), the European Data Protection Supervisor (EDPS), the European Agency for Fundamental Rights (FRA), European Aviation Safety Agency, European Labour Authority, the European Commissioner for Competition, Digital Economy taskforces and other field-specific regulators in various capacities. The task force could be set up for a limited, specified time (during which it might be considered whether it needs to be open-ended) or it could be open-ended. The former is recommended at the current time. Scope: Regional (EU-level). To be further determined based on defined role.
3. Purpose/objective/what need does the option fulfil?	To help promote and protect fundamental rights across the EU by tackling/clarifying legal issues and helping cooperatively find means to address the adverse impacts of AI/big data especially in relation to high-risk activities.

Option: EU Taskforce of field specific regulators for AI/big data Proposer: Suggested by one of SHERPA's stakeholder Board members in the scoping paper feedback. Reference/link to relevant document: - General note: This proposal was not found fully expanded at the time of our research, but we have examined how it might potentially play out. Assessed by: TRI Date of assessment: 12 November 2019 Stakeholder(s) consulted in option assessment: Félicien Vallet, CNIL	
Criteria/touch point	**Assessment**
4. What gap does it address?	It might help address shortcomings in the areas of cooperation, coordination, consistent application of Union law related to AI/big data, also, e.g., cross-border risks from AI and big data applications.
5. What added value does it have?	It will promote cooperation on AI/big data legal issues and provide clarity at the EU-level. The task force could create a good collaborative environment for EU AI policy and regulation and promote the adoption of a unified message on AI/big data regulation to the extent possible/required.
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*	Limitations: Task forces are limited often by the capacity of their members. Risks: If established and its mandate is not clear, it might itself duplicate the work of existing EU agencies. It might cause further frustrations amongst stakeholders. Challenges: include the changing regulatory culture of the EU, managing conflicts, limited resources, funding issues and personnel turnover.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	No. It has yet to be elaborated and discussed in detail. Some ideas have been presented in this assessment but these are only preliminary and rough indications.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	To be determined.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	This depends on the elaboration of this option. But generally, as we see it: Citizens: none Public administrations: EU regulatory agencies/supervisory bodies will have to devote resources to it. Businesses, particularly SMEs: none
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers;*	Policy makers and regulators.

Option: EU Taskforce of field specific regulators for AI/big data Proposer: Suggested by one of SHERPA's stakeholder Board members in the scoping paper feedback. Reference/link to relevant document: - General note: This proposal was not found fully expanded at the time of our research, but we have examined how it might potentially play out. Assessed by: TRI Date of assessment: 12 November 2019 Stakeholder(s) consulted in option assessment: Félicien Vallet, CNIL	
Criteria/touch point	**Assessment**
regulators; civil society; individuals, others (please specify)]	
11. Whose rights and/or interests does this option neglect?	Not clear at the moment and needs further thought.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	This would depend on the focus of the taskforce and the cooperation it achieves. Aspects of AI and/or big data that adversely affect human rights and not well addressed by other EU agencies (e.g., FRA) could be one of the focus areas of the task force – it could complement the work of existing agencies.
13. How does it address ethics and ethical principles? Which ones?	Not elaborated
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	Not elaborated
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not elaborated
16. What provisions are there for regular review and update?	Not elaborated
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Its feasibility and sustainability would depend on internal and external buy-in and EU political will to create such a taskforce/and if created to keep it going. It might also be affected by competing priorities of the different bodies that might be expected to house and/or form it. The task force would also need to allay the concern of participating bodies that such participation may conflict with their primary mission.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The EU a has a variety of task forces set up for various purposes, e.g., Task force on subsidiarity, proportionality and doing less more efficiently, Financial Action Task Force (FATF) [first closed, then open-ended]; Taskforce on Article 50 negotiations with the United Kingdom, Advanced manufacturing Task Force, Smart

Option: EU Taskforce of field specific regulators for AI/big data Proposer: Suggested by one of SHERPA's stakeholder Board members in the scoping paper feedback. Reference/link to relevant document: - General note: This proposal was not found fully expanded at the time of our research, but we have examined how it might potentially play out. Assessed by: TRI Date of assessment: 12 November 2019 Stakeholder(s) consulted in option assessment: Félicien Vallet, CNIL	
Criteria/touch point	**Assessment**
	Grids Task Force. The task force competencies would be limited to the areas in which the Union can act.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	-
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	2
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Factors critical to its adoption: political will to adopt and specify its responsibilities (other regulatory agencies might not be willing to relinquish their control/or play ball); whether it can truly be harnessed to develop/support the adoption of high-quality regulation in AI/big data without fuelling further a race to the bottom. Factors for its success include whether the task force is able to successfully carry out its designated functions by not being bogged down in red tape.
References consulted	-

4.16. Algorithmic Impact Assessments under the GDPR

Option: Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations Proposer: Margot E. Kaminski and Gianclaudio Malgieri Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224## Assessed by: TRI Date of assessment: 15/11/19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	Kaminski and Malgieri examine the requirements for Data Protection Impact Assessments (DPIA) under Regulation (EU) 2016/679 (GDPR) and of algorithmic accountability disclosure duties under Articles 13-15 and 22 of the GDPR. Reflecting on these two systems, they propose that DPIAs should be seen as the connection of the above GDPR systems and provisions, i.e., the DPIA as a systemic (and collaborative) governance regime and as an element of the GDPR's protection of individual rights. In fact, the DPIA as collaborative governance takes into account

Option: Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations Proposer: Margot E. Kaminski and Gianclaudio Malgieri Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224## Assessed by: TRI Date of assessment: 15/11/19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	the risks to data subjects and considers the safeguard of individual rights as risk-mitigating measures. For example, a DPIA requires the description of the system and data processing activities at a systematic level. This information could be then disclosed to individuals or the wider public enhancing Articles 13-15 and 22 GDPR. In addition, this connection is further enhanced by integrating systemic accountability measures such as audits or external review. Under this approach, an expanded version of DPIA, the so-called Algorithmic Impact Assessment (AIA), is suggested as a tool of algorithmic governance. The AIA is a further development of DPIAs tailored to address the risks of AI applications. AIAs also serve the need for multi-layered explanations of algorithmic decision-making. This involves interdisciplinary efforts: technologists to assess what risk-mitigation and accountability measures could be implemented, and lawyers and ethicists to think through how to better involve constituents and define problems. It will also involve a deeper exploration of how to link the material created during the DPIA process to the individual disclosures required under the GDPR. In addition, this involves individual explanations, group explanations, and systemic explanations, both internal and external, a right to an explanation of the model, and a right to an individual explanation of an individual decision. Moreover, it includes not just systemic and individual analysis, but group-level analysis of how an algorithm might impact particular classes of individuals, or particular locations individuals need to know not just information about a particular stand-alone decision, but information about the algorithm's treatment of groups, and tendency towards bias and discrimination.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: It relies and expands on the notion and building blocks of the DPIA of the GDPR. Moreover, the concept, conduct, and methodology of these algorithmic impact assessments rely on predecessor impact assessments, including Environmental Impact Statements, Human Rights Impact Assessments, Privacy Impact Assessments, Ethical Impact Assessments, and Surveillance Impact Assessments. The change it aims to bring is on soft law rather than hard law. It draws on the GDPR provisions, i.e., Articles and Recitals,

Option: Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations Proposer: Margot E. Kaminski and Gianclaudio Malgieri Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224## Assessed by: TRI Date of assessment: 15/11/19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	official guidance and academic literature to adopt a more inclusive and sophisticated approach to DPIAs for algorithmic accountability. This option does not require legislative amendments but policy decisions, best practices and guidance as it is seen embedded in the GDPR rhetoric. Although not specified, if this option is approved but found inconsistent with the letter of the GDPR, a new legal act may be necessary to introduce the requirement for AIAs. Nature: This is a policy suggestion on how Articles 13-15, 22 and 35 GDPR should be read, understood and applied for algorithmic governance. If this approach is found consistent with the spirit and letter of the GDPR, then this methodology of AIAs should be considered a legally binding obligation for organisations using algorithms. Scope: This option aims to cover the legal and technological developments in the EEA/EU drawing on the GDPR provisions. However, this suggestion is of practical relevance and importance on an international level since it provides best practices in the area of DPIAs and given the extra-territorial effect of the GDPR.
3. Purpose/objective/what need does the option fulfil?	The proposal addresses how a DPIA links the two faces of the GDPR's approach to algorithmic accountability: individual rights and systemic collaborative governance. This version of AIA is suggested as a connection between the GDPR's two methods of governing algorithmic decision-making by both providing systemic governance and serving as an important "suitable safeguard" (Art. 22 GDPR) of individual rights. On a more specific tone, the objective of this tool is to ensure accountability in the use of AI, effectively prevent risks, enhance individual rights and ensure the legitimacy and legality of AI applications.
4. What gap does it address?	The suggested methodology and format of AIAs is an expanded and developed version of the DPIA under the GDPR. This AIA addresses the shortcomings of this DPIA and gaps in impact assessments of AI.
5. What added value does it have?	The added value of this option lies in the enhanced accountability elements added in the DPIA under the GDPR. Overall, the purpose of this assessment is not to ensure transparency about the technology itself, but to facilitate explanations about the lawfulness, fairness, and legitimacy of certain decisions. In this context, this AIA is not suggested as a

Option: Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations Proposer: Margot E. Kaminski and Gianclaudio Malgieri Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224## Assessed by: TRI Date of assessment: 15/11/19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	stand-alone mechanism. On the contrary, it is seen as part of a larger system of governance- relationship to other accountability tools in the GDPR. The suggested AIA is seen both as a substantive and procedural requirement requiring public engagement, detailed and multi-layered explanations of AI, expert involvement and oversight. Moreover, the focus is both on algorithms as technology in isolation and algorithms as systems embedded in human systems. In this context, individual-level, systemic and group-based explanations are required to minimise the risk of error and prevent discriminatory effects. Another important aspect of this suggestion is the call for involving and engaging impacted individuals, not just through surveys but through representative boards, before an algorithm is deployed. It also requires companies, or regulators, to help fund the involvement of both of the above and provide technical expertise or the resources for obtaining technical expertise. It should involve not just external technical experts, but external experts in law and ethics to help define, or at least frame discussions of, what we mean by terms like "discrimination" or "bias Finally, attention is paid to public-facing disclosure, which enables public feedback, both in the form of market feedback (enabling individuals to avoid companies with bad policies) and in the form of regulatory feedback over the longer term (enabling individuals to elect representatives who will put in place laws that will prevent bad company behaviour).
6. What are the limitations, risks and challenges?	Limitations: In general, this paper has touched upon a wide area of topics regulated under the GDPR and other pieces of legislation. As acknowledged, this paper has not examined, though, the necessity and proportionality elements as required under Article 35 GDPR. This is essential to understand how the purpose of this AIA, i.e., an explanation about the lawfulness, fairness, and legitimacy of certain decisions, is materialised. Another limitation relates to the restrictive focus on bias and discriminations (based on Recital 71 of the GDPR) and setting aside the examination of the protection of other fundamental rights, freedoms and interests Although the examination of these rights is not excluded and -probably- covered under the broad framing of this AIA, it is necessary to consider the specific rights and interests at risk. Finally, AIAs should focus both on the risks of using algorithms and the risks of not using one.

Option: Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations Proposer: Margot E. Kaminski and Gianclaudio Malgieri Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224## Assessed by: TRI Date of assessment: 15/11/19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	Both risks and benefits should be considered in AIAs and the deployment of AIAs should not be limited to the risks and harms. Risks: This extended and expanded AIA is a demanding task, requiring expertise, resources and financial investment, and may stifle innovation. Risk of lack of common standards: Whereas Article 35 GDPR requires the conduct of a DPIA where the processing of personal data creates risks for data subjects, it is not clear whether the relevance of the requirement of an AIA will be commonly understood by all stakeholders. First, a specific requirement for this expanded AIA is not explicitly provided under the GDPR. Second, there is no agreement or consensus on the definition of algorithms. Therefore, there is a risk that the requirement for an AIA may not be clear or implemented in all the uses of algorithms. Risk of disclosure of sensitive information: The suggested publication of AIAs and involvement of third parties for review and oversight should be balanced against the risk of disclosing commercially sensitive information or confidential information for national purposes e.g., national security. Although it aims to be practical and its findings grounded, there is a risk it may become too academic, over-descriptive, relying on assumptions and far-stretching societal consideration. On the contrary, the risk-based approach should be applied, which requires the management of emerging and more serious risks in the first place. Challenges: The main challenge around the adoption and implementation of this approach to Algorithmic Impact Assessments relates to the far-stretching interpretation of the requirements of the GDPR. The suggested approach should be endorsed by other key stakeholders, including academia, human rights and civil organisations, and finally confirmed by European and national legislators or authorities. Moreover, a clear liability framework should be in place to provide for the consequences of partial failure to comply with this tool. For example, publication is an important element of this measure. Nonetheless, since this is not explicitly provided under the

Option: Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations Proposer: Margot E. Kaminski and Gianclaudio Malgieri Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224## Assessed by: TRI Date of assessment: 15/11/19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	GDPR, it is not clear whether failure to publish the outcome of Algorithmic Impact Assessments also constitutes a breach of the GDPR provisions. The same applies where an organisation may not take into account the feedback from internal and external reviews. Therefore, the challenge that this expanded AIA may become a checkbox exercise, a mere bureaucratic requirement, should be addressed. Moreover, a time framework should apply for internal and external review and feedback given into account that this AIA will be released to the public. Considering and replying to feedback received from the public without time limitations may render the examined technology ineffective and disused if it is paused until all feedback is fed back into the AIA. Finally, relying on external auditing and review raises the challenge of ensuring funding and the appropriate resources, especially for small and medium enterprises, your entrepreneurs and researchers.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	This option is sufficiently clear and builds on existing frameworks. More detail is required about the structure, order and planning of this AIA to ensure common standards, consistency and legal certainty. Its operationalisation can be effective and efficient if this is seen as a legal requirement and obligation and there is oversight and enforcement.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	It is suggested that Data Protection Authorities (DPAs) should be involved in the oversight of the implementation of this tool. The involvement of DPAs should be on a procedural and substantive level, checking the efficacy of the process and substantive problems with algorithmic decision-making. DPAs should monitor companies as they come up with ways of addressing problems with algorithmic decision-making, and it reassures individuals that their dignity and other rights are being respected by a fair system. For example, DPAs should inspect particular companies, check for compliance and enforce against captured versions of AIAs. DPAs should also establish more concrete best practices or support the establishment of sector-specific codes of conducts around algorithmic fairness.

Option: Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations Proposer: Margot E. Kaminski and Gianclaudio Malgieri Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224## Assessed by: TRI Date of assessment: 15/11/19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Citizens: Not elaborated/clear. Public administrations: DPAs will need financial support, additional resources and expertise to support the deployment and oversight of AIAs. Businesses: Businesses will need to design and implement policies and procedures for AIAs. They will need to cooperate with auditors and reviewers to share the conducted and in progress AIAs. This may be in conflict with their interests and rights in keeping business information confidential.
10. Which stakeholders would benefit most from the use of this option?	It is detailed that several stakeholders will benefit from this form of AIAs. Most importantly, engagement with the public and disclosure of information about the AIA enhances transparency, fairness and accountability. The adoption of an AIA and multi-layered explanations might be a "suitable safeguard" to protect fundamental rights and freedoms of individuals both under Article 22(3) and under Article 35(7)(d) of the GDPR. This AIA model serves as a collaborative governance mechanism for companies in constituting the substance and practice of individual due process rights. In addition, AIAs help businesses to consider and redesign their policies expanding company commitments and changing company decision-making. Businesses are also benefited because they are supported to meet the requirement for data protection and privacy by design and default. Fair AI will also support trust and confidence in the use of AI. In addition, this AIA model will enable data controllers to design their mechanisms and processes in advance to reply to the requirements of Articles 13-15 and 22 of the GDPR. Policymakers, enforcement authorities and legislators will be also benefited since this AIA will support public scrutiny and visibility of the use of AI. Moreover, the publication of AIAs will support evidence-based and engaged policymaking and decision-making, where the AIAs will be used as a reference point for the particular AI application and potential uses of AI and data-driven technologies.
11. Whose rights and/or interests does this option neglect?	Not elaborated. This suggestion considers rights and interests in a rather broad manner, examining the rights and interests of

Option: Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations Proposer: Margot E. Kaminski and Gianclaudio Malgieri Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224## Assessed by: TRI Date of assessment: 15/11/19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	the concerned data subjects, group of individuals, communities and society at large. What this suggestion should emphasise, though, is that the balancing and assessment conducted in AIAs should also consider the benefits of AI, including the satisfaction of the rights and interests of businesses or the wider interests.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	It actively and explicitly supports human rights, especially the rights to privacy, dignity, data protection, and freedom from discrimination. In addition to identifying and preventing risks from the use of AI, this AIA aims to ensure that a system is legal and fair before its deployment. This further supports the respect for human rights by preventing illegal systems in the market, causing errors, bias, and discrimination.
13. How does it address ethics and ethical principles? Which ones?	This option touches upon several ethical principles. In particular, the following ethical principles have been considered in tailoring the GDPR DPIA to the system of algorithmic accountability: • public legitimacy and acceptance for the use of a system • liability • fairness • respect for human rights • transparency • public engagement • prohibition of bias and discrimination • participatory model of governance
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	It does not explicitly consider gender dimensions. However, such considerations are embedded in the narrative and methodology of this AIA, which aims to address issues of unfairness, bias, errors, and discrimination on an individual and collective basis.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	The proposal suggests innovative ways to fund and support this AIA model and process. It requires companies or regulators, to help fund the involvement of both and provide technical expertise or the resources for obtaining technical expertise during the deployment of the AIA. Regarding the funding of regulatory body, this is not applicable. This option does not relate to the establishment or powers of a body/agency/authority.

Option: Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations Proposer: Margot E. Kaminski and Gianclaudio Malgieri Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224## Assessed by: TRI Date of assessment: 15/11/19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
16. What provisions are there for regular review and update?	It is suggested that a model AIA should be truly continuous: a process that produces outputs, but also includes ongoing assessment and performance evaluation, especially for those algorithms that change quickly over time.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	The core purpose and functionalities of this AIA model is to remain sustainable and law-, policy- and technology-responsive. Mechanisms are embedded to address future developments in law, industry and technology through public engagement, the involvement of DPAs, external and internal reviewers.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	It is not expected to adversely impact the ability for businesses and others to innovate. Unfair and erroneous algorithmic systems will not be permitted but this is outside the scope and field of fair and lawful innovation. However, businesses may be discouraged from engaging with AI due to the requirement to invest time, effort, expertise, tools and resources in deploying a challenging and expanded DPIA.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	This AIA model is consistent with the EU legal framework, especially the fundamental rights framework, since it aims to ensure that individuals are not subjected to an unfair, arbitrary, discriminatory, or erroneous system. What must be further examined and confirmed by the European and/or national regulators is whether this model of AIA fits within the context, scope and application field of the GDPR.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	Please see above and below.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	3
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	The adoption and implementation of this measure are contingent on the approval and confirmation of European and national authorities that this approach fits within the GDPR or falls under a new legislative requirement. Given that Article 22 GDPR has been specified in a different manner in national legislation (e.g., Slovenian Data Protection Law, see Malgieri, G.

Option: Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations Proposer: Margot E. Kaminski and Gianclaudio Malgieri Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224## Assessed by: TRI Date of assessment: 15/11/19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	, "Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations," Computer Law and Security Review, 35(5), 2019) there is a risk of a scattered approach to this requirement. In addition, a standard methodology and template should be approved and shared with the public to ensure consistency and legal certainty. Moreover, research is needed about how different layers of explanations—systemic explanations, group explanations, and individual explanations—can interact each other and how technical tools can help in developing an Algorithmic Impact Assessment that might be re-used towards GDPR-complying explanations and disclosures.
References consulted	Article 29 Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, 2018. https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 Information Commissioner's Office, Data Protection Impact Assessments and AI, 23 October 2019 . https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/10/data-protection-impact-assessments-and-ai/ Kaminski, M., "Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability," *Southern California Law Review*, 92(6) (2019) Kaminski, M., "The Right to Explanation, Explained', *Berkeley Technology Law Journal,* 34(1) (2019) Reisman, Dillon, et al., 'Algorithm Impact Assessment: A Practical Frameworks for Public Agency Accountability' (AI Now Institute) 2018. https://ainowinstitute.org/aiareport2018.pdf

4.17. Voluntary/mandatory certification of algorithmic decision systems (ADS)

Option: Voluntary/mandatory certification of algorithmic decision systems (ADS) Proposer: Certification of AI systems at EU level (AI HLEG, Policy and Investment recommendations, 2019); Voluntary/mandatory certification of algorithmic decision systems (ADS), (STOA study Understanding algorithmic decision-making: Opportunities and challenges , 2019); Reference/link to relevant document: See above. Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	The EU Parliament STOA study suggests "the certification of ADS should be encouraged and even mandatory in certain sectors". Further that "certifications and labels, if properly implemented, can be a way to enhance trust in ADS and to verify that they comply with certain rules (such as the absence of bias or discrimination)." (STOA 2019) The AI HLEG *Policy and Investment Recommendations* calls for certification of AI systems at EU level to "counter fragmentation of standards" and "help provide the means to assess the quality of an AI solution after deployment and possibly to decide which solution is best". (AI HLEG PR 2019).
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: The certification could be based, e.g., on Article 42 of the General Data Protection Regulation or other sectoral legislation. Nature: The STOA Study states, "for the deployment of ADS, certification can be on either a voluntary basis (as encouraged by the GDPR), or mandatory in certain areas such as justice and healthcare." (STOA 2019) Scope: Per the STOA Study 2019, "certification requirements and obligations should be sectoral. Indeed, the needs and the risks vary greatly from one type of application to another and sectoral supervisory authorities or agencies are in a better position to define reference evaluation criteria and to control their application. For the deployment of ADS, certification can be on either a voluntary basis (as encouraged by the GDPR), or mandatory in certain areas such as justice and healthcare." (STOA 2019)
3. Purpose/objective/what need does the option fulfil?	To enhance trust in algorithmic decision systems and to verify that they comply with certain rules (such as the absence of bias or discrimination) (STOA 2019).
4. What gap does it address?	It would be another means of reducing the risks related to algorithmic decision systems.
5. What added value does it have?	They might help as, stated by the AI HLEG *Ethics Guidelines for Trustworthy AI* report, "apply standards developed for different application domains and AI techniques, appropriately aligned with the industrial and societal standards of different contexts". (AI HLEG EG 2019)

Option: Voluntary/mandatory certification of algorithmic decision systems (ADS) Proposer: Certification of AI systems at EU level (AI HLEG, Policy and Investment recommendations, 2019); Voluntary/mandatory certification of algorithmic decision systems (ADS), (STOA study Understanding algorithmic decision-making: Opportunities and challenges , 2019); Reference/link to relevant document: See above. Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	Where certification takes on/alleviates the burden of legislation and other forms of enforcement and oversight, it might reduce the workloads of regulators/enforcers.
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*	Limitations: As pointed out by the AI HLEG *Ethics Guidelines for Trustworthy AI* report**,** certification can "never replace responsibility. It should hence be complemented by accountability frameworks, including disclaimers as well as review and redress mechanisms". (AI HLEG EG 2019) Risks: Edwards and Veale highlight some risks using certification for machine learning systems, e.g., the privatization of regulation and scrutiny. They state, "Certification scheme and trust seals have to make money to survive, which can only be obtained by asking fees from members. Given this self-interest, it is hard to punish members too hard when they breach the rules of the seal or certificate, for fear they will leave, either altogether or for a less demanding trust seal (in a plural market, which is generally what is envisaged). This in turn tends to diminish the value of the seal or certificate as a guarantee of trustworthiness."(Edwards and Veale 2018) Challenges: include getting organisations to certify (if voluntary).

7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why?	No. It has not been elaborated.
	
	However, Edwards and Veale suggest "certification could be applied to two main aspects of algorithmic systems:
	1. certification of the algorithm as a software object by
	  1. directly specifying either its design specifications or the process of its design, such as the expertise involved ("technology-based standards") and/or
	  2. specifying output-related requirements that can be monitored and evaluated ("performance-based standards");
	2. certification of the whole person or process using the system ("system controller") to make decisions, which would consider algorithms as situated in the context of their use. (Edwards and Veale 2018)

8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Not elaborated in the STOA or AI HLEG documents.

Option: Voluntary/mandatory certification of algorithmic decision systems (ADS) Proposer: Certification of AI systems at EU level (AI HLEG, Policy and Investment recommendations, 2019); Voluntary/mandatory certification of algorithmic decision systems (ADS), (STOA study Understanding algorithmic decision-making: Opportunities and challenges , 2019); Reference/link to relevant document: See above. Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Citizens: None Public administrations: if they are involved in setting up an/or managing the certification scheme, they would incur scheme design, implementation, monitoring (oversight) and enforcement burdens. Businesses including SMEs: will face scheme compliance burdens. I.e., they may need to put measures in place to ensure they meet the certification standards and requirements, they may need to disclose data in a transparent manner (which might be a burden if that is not how they normally operate).
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Certified organisations.
11. Whose rights and/or interests does this option neglect?	Potentially, consumers as costs of certification compliance might get passed down to them.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	It could support human rights by requiring e.g., the use of privacy by design, data protection by design and default, human rights impact assessment, data protection impact assessment.
13. How does it address ethics and ethical principles? Which ones?	It could potentially use the principles outlined in the Ethics Guidelines for Trustworthy Artificial Intelligence – which are: Human agency and oversight, Technical Robustness and safety, Privacy and data governance, Transparency, Diversity, non-discrimination and fairness, Societal and environmental well-being and Accountability.

Option: Voluntary/mandatory certification of algorithmic decision systems (ADS) Proposer: Certification of AI systems at EU level (AI HLEG, Policy and Investment recommendations, 2019); Voluntary/mandatory certification of algorithmic decision systems (ADS), (STOA study Understanding algorithmic decision-making: Opportunities and challenges , 2019); Reference/link to relevant document: See above. Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority ? Outline.	Not elaborated in the STOA or the AI HLEG documents.
16. What provisions are there for regular review and update?	Not elaborated in the STOA or the AI HLEG documents.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Its feasibility and sustainability will depend on sustained efforts/support from governments/public sector to incentivise its creation and then effective use. It also depends on whether certified schemes are able to achieve higher market penetration, whether the schemes have a strong (technical or regulatory) framework and is non-ambiguous; buy-in to the scheme and trust in it.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU	The certification could be based, e.g., on Article 42 of the General Data Protection Regulation or other sectoral legislation.

Option: Voluntary/mandatory certification of algorithmic decision systems (ADS) Proposer: Certification of AI systems at EU level (AI HLEG, Policy and Investment recommendations, 2019); Voluntary/mandatory certification of algorithmic decision systems (ADS), (STOA study Understanding algorithmic decision-making: Opportunities and challenges , 2019); Reference/link to relevant document: See above. Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
to implement these actions in accordance with the EU acquis)	
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	Might include over-protectionism and ambiguity and complexity of legislative requirements that form its basis.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	3
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Factors critical to its success: ensuring there is no potential for misuse of the certification scheme (e.g., misrepresentation, fraudulent representation of certification, free riding, conflicts of interest e.g., certification of subscribers from whose subscriptions the certifier profits). We should also note the comments of Martini in this respect - , "… given the transformability of modern software systems, a mere *ex ante* assessment *certification process* carried out through single-event testing is only of limited use in achieving the intended purpose to protect consumers' rights. A better option is *continuous auditing* over the systems' entire life cycle. Integrating audits in the regulatory system (for example in a manner similar to the Eco-Audit Directive) is particularly useful to incorporate the expertise of private parties in the regulatory task of market and product surveillance." (Martini 2019)
References consulted	Castelluccia, Claude and Daniel Le Métayer, Understanding algorithmic decision-making: Opportunities and challenges, Panel for the Future of Science and Technology (STOA), European Parliament, March 2019. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf

Option: Voluntary/mandatory certification of algorithmic decision systems (ADS) Proposer: Certification of AI systems at EU level (AI HLEG, Policy and Investment recommendations, 2019); Voluntary/mandatory certification of algorithmic decision systems (ADS), (STOA study Understanding algorithmic decision-making: Opportunities and challenges , 2019); Reference/link to relevant document: See above. Assessed by: TRI Date of assessment: 12 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	Edwards, Lilian, and Michael Veale, "Enslaving the algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?," *IEEE Security & Privacy,* 16.3, 2018, pp. 46-5. High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines For Trustworthy AI,* 2019. **https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai** High-Level Expert Group on Artificial Intelligence, *Policy And Investment Recommendations for Trustworthy AI,* June 2019. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Martini, Mario, "Fundamentals of a Regulatory System for Algorithm-Based Processes"**,** Expert opinion prepared on behalf of the Federation of German Consumer Organisations (Verbraucherzentrale Bundesverband), 1 May 2019. https://www.vzbv.de/sites/default/files/downloads/2019/07/19/martini_regulatory_system_algorithm_based_processes.pdf

4.18. DEEP FAKES Accountability Act (H.R. 3230)

Option: Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or the "DEEP FAKES Accountability Act", (H.R. 3230) 116th Cong. (2019) Proposer/Sponsor: Rep. Yvette D. Clarke, [D-NY-9] (Introduced 12 June 2019) Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/3230/text Assessed by: TRI Date of assessment: 4 November 2019 Stakeholder(s) consulted in option assessment: Adam Holland, Berkman Klein Center	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI,	The Bill provides that any "person who, using any means or facility of interstate or foreign commerce, produces an advanced technological false personation record with the intent to distribute such record over the internet or knowledge that such record shall be so distributed, shall ensure such record, complies with—"(1) the watermark requirement under subsection (b); and "(2) (A) in the case of an

Option: Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or the "DEEP FAKES Accountability Act", (H.R. 3230) 116th Cong. (2019) Proposer/Sponsor: Rep. Yvette D. Clarke, [D-NY-9] (Introduced 12 June 2019) Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/3230/text Assessed by: TRI Date of assessment: 4 November 2019 Stakeholder(s) consulted in option assessment: Adam Holland, Berkman Klein Center	
Criteria/touch point	**Assessment**
such as transparency, robustness and security measures?) Give an application example	audiovisual record, the disclosure requirements under subsection (c); "(B) in the case of a visual record, the disclosure requirements under subsection (d); or "(C) in the case of an audio record, the disclosure requirements under subsection (e). The term 'advanced technological false personation record' means any deep fake, which—(A) a reasonable person, having considered the visual or audio qualities of the record and the nature of the distribution channel in which the record appears, would believe accurately exhibits—(i) any material activity of a living person which such living person did not in fact undertake; or (ii) any material activity of a deceased person which such deceased person did not in fact undertake, and the exhibition of which is substantially likely to either further a criminal act or result in improper interference in an official proceeding, public policy debate, or election; and (B) was produced without the consent of such living person, or in the case of a deceased person, such person or the heirs thereof. The term 'deep fake' is defined as, "any video recording, motion-picture film, sound recording, electronic image, or photograph, or any technological representation of speech or conduct substantially derivative thereof— (A) which appears to authentically depict any speech or conduct of a person who did not in fact engage in such speech or conduct; and (B) the production of which was substantially dependent upon technical means, rather than the ability of another person to physically or verbally impersonate such person.

Option: Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or the "DEEP FAKES Accountability Act", (H.R. 3230) 116th Cong. (2019) Proposer/Sponsor: Rep. Yvette D. Clarke, [D-NY-9] (Introduced 12 June 2019) Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/3230/text Assessed by: TRI Date of assessment: 4 November 2019 Stakeholder(s) consulted in option assessment: Adam Holland, Berkman Klein Center	
Criteria/touch point	**Assessment**
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	**Basis:** Law (civil and criminal) **Nature:** If implemented, it would be binding and enforceable against individuals. **Scope:** The Bill does not authorize the production of an advanced technological false personation record which includes disclosures if such record is otherwise prohibited by law or regulation. The word 'advanced' in 'advanced technological false personation record' is not meant in a way as to narrow its interpretation. The Bill is not a defence against, or attempt to pre-empt, or limit, any Federal, State, local, or territorial laws, regulations, or policies that prohibit, impose more stringent standards in relation to, or provide additional or alternative remedies or damages in relation to, the production or distribution of advanced technological false personation records, deep fakes, or related content, including criminal and civil laws relating to copyright, tortious conduct, and false personation. The Bill outlines exceptions, e.g., requirements might not apply with respect to deepfakes: containing alternative disclosures regarding the falsity of the exhibited material activities which a reasonable person would deem to be more prominent than those required under the law; during the process of producing such record, provided the ultimately distributed record is in compliance; which primarily contains images or sound recordings of actual persons, such as performing artists, and have not been substantially digitally modified; created in connection with editing a motion picture, television, music, or similar production or creating a derivative production thereof, the original content of which was created prior to the enactment of this Act, in which the person appearing provided consent to their original appearance; appearing in a context such that a reasonable person would not mistake the falsified material activity for actual material activity of the exhibited living person, such as parody shows or publications, historical re-enactments, or fictionalized radio, television, or motion picture programming; or produced by an officer or employee of the United States, or under the authority thereof, in furtherance of public safety or national security.
3. Purpose/objective/what need does the option fulfil?	To combat the spread of disinformation through restrictions on deep-fake video alteration technology.
4. What gap does it address?	Ferraro explains it thus, "In sum, the DEEP FAKES Accountability Act seeks to protect against the full gamut of deepfake harms, from nonconsensual pornography to foreign interference in elections and public policy debates, from inciting violence to conducting financial fraud and identity theft." See Ferraro, Matthew F., "*Deepfake Legislation: A Nationwide Survey—State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media",* 25 Sept 2019.

Option: Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or the "DEEP FAKES Accountability Act", (H.R. 3230) 116th Cong. (2019) Proposer/Sponsor: Rep. Yvette D. Clarke, [D-NY-9] (Introduced 12 June 2019) Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/3230/text Assessed by: TRI Date of assessment: 4 November 2019 Stakeholder(s) consulted in option assessment: Adam Holland, Berkman Klein Center	
Criteria/touch point	**Assessment**
	https://www.wilmerhale.com/en/insights/client-alerts/20190925-deepfake-legislation-a-nationwide-survey
5. What added value does it have?	In addition to "updating the definitions of the federal identity theft statute (18 U.S.C. § 1028) to include in the list of prohibited forgeries a "false audiovisual identification record" and the "federal false personation statute (18 U.S.C. Ch. 43) to prohibit the use of deepfake technology to impersonate falsely an officer or employee of the United States, among others." (Ferraro, 2019). The Bill includes provisions for criminal penalty for violation of its provisions (up to five years) and provisions for civil penalty (failure to disclose - up to $150,000 per record or alteration, as well as appropriate injunctive relief; altering disclosures - up to $150,000 per record or alteration, as well as appropriate injunctive relief). The Bill also includes a right to private action for affected parties – civil action before the appropriate Federal district court for damages and injunctive relief. The Bill provides for extraterritorial Federal jurisdiction over an offense if the defendant or the depicted person is a citizen or permanent resident of the United States.
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*	The Electronic Frontier Foundation (EFF) outlines the following 'problems' with the Bill that are relevant to mention here. One, an overbroad definition of 'deepfakes'. Another is that "it's unclear how mandatory labeling and watermarking will solve the real harms that malicious deepfakes are causing. The trolls of the world will likely just not comply, particularly if they don't live in the United States." Another issue EFF raises is "the bill's breadth and penalties trigger many First Amendment problems. For example, while there is an exception for parodies, satires, and entertainment—so long as a reasonable person would not mistake the "falsified material activity" as authentic—the bill fails to specify who has the burden of proof, which could lead to a chilling effect for creators. And in addition to civil penalties of up to $150,000 for failure to include a watermark or disclosure, the bill imposes criminal penalties—even without any showing of harm—for violations intended not only to harass, incite violence, interfere in an election, or perpetuate fraud, and to "humiliate" the person depicted, a vague term which the bill does not define. The First Amendment generally bars criminal laws that impose penalties without any showing of harm." The EFF also outlines that the Bill "creepily exempts officers and employees of the United States acting in furtherance of public safety or national security." See Tsukayama, Hayley, India Mckinney, and Jamie Williams, "Congress Should Not Rush to Regulate Deepfakes", Electronic Frontier Foundation, 24 June 2019. https://www.eff.org/deeplinks/2019/06/congress-should-not-rush-regulate-deepfakes

Option: Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or the "DEEP FAKES Accountability Act", (H.R. 3230) 116th Cong. (2019) Proposer/Sponsor: Rep. Yvette D. Clarke, [D-NY-9] (Introduced 12 June 2019) Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/3230/text Assessed by: TRI Date of assessment: 4 November 2019 Stakeholder(s) consulted in option assessment: Adam Holland, Berkman Klein Center	
Criteria/touch point	**Assessment**
	Cushing also discusses deepfake legislation and suggests this will threaten free speech via criminalisation on common behaviours. Cushing, Tim, "Deep Fake' Legislation Is On The Way, Threatening Free Speech Protections" TechDirt, 9 July 2019. https://www.techdirt.com/articles/20190706/16124442526/deep-fake-legislation-is-way-threatening-free-speech-protections.shtml Another challenge is highlighted by Coldewey who suggests that the Bill's provisions are "seem too optimistic in the face of the reality of this threat" and has critiqued its loopholes: i.e., requiring satirists and YouTubers to document their modified or generated media would assign paperwork to people already acting legally and with no harmful intentions; the Bill makes stripping metadata and documentation (or making them inaccessible) illegal but as it is done regularly and automatically by bots, anonymous reposters and such, it seems unlikely to be able to identify a criminal. Coldewey strongly states that the DEEPFAKES Accountability Act does not create "much in the way of accountability for the malicious actors most likely to cause problems." Coldewey, Devin, "DEEPFAKES Accountability Act would impose unenforceable rules — but it's a start", *TechCrunch,* 13 June 2019. http://techcrunch.com/2019/06/13/deepfakes-accountability-act-would-impose-unenforceable-rules-but-its-a-start/
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	As outlined by Holland, the criticisms referenced above alone would evidence some fairly substantial obstacles to operationalization of H.R. 3230. For example, if the law would immediately be challenged on grounds of its constitutionality, and doesn't directly allocate burden of proof, that would certainly prevent it from being passed into law/being put into effect all, much less effectively or efficiently. Next, as the EFF and others point out, constitutionality aside, the putative bad actors the Bill targets will simply ignore its tenets, making any effective operationalization impossible. The watermarking the bill proposes will quickly be stripped, and so on. Regarding the monitoring and oversight mechanisms described below, no mention is made of funding for these, another obstacle towards effective implementation.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	The Bill outlines that the Attorney General shall designate a Coordinator For Violations Directed By Foreign Nation-States in each United States Attorney's Office to receive reports from the public regarding potential violations of section 1041 (relating to deep fake depictions produced or distributed by any foreign nation-state, or any agent acting on its behalf) and coordinate prosecutions for any violation of such section. It also outlines that the Attorney General

Option: Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or the "DEEP FAKES Accountability Act", (H.R. 3230) 116th Cong. (2019) Proposer/Sponsor: Rep. Yvette D. Clarke, [D-NY-9] (Introduced 12 June 2019) Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/3230/text Assessed by: TRI Date of assessment: 4 November 2019 Stakeholder(s) consulted in option assessment: Adam Holland, Berkman Klein Center	
Criteria/touch point	**Assessment**
	shall designate a Coordinator For False Intimate Depictions in each United States Attorney's Office to receive reports from the public regarding potential violations of section 1041 (relating to deep fake depictions of an intimate and sexual nature) and coordinate prosecutions for any violation of such section. The Bill also provides that on the effective date of the Act, the Attorney General shall publish a report containing (inter alia), a plan to effectuate and enforce section 1041.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Based on the Bill, this would fall on manufacturers of software, who in the course of conducting such business produce software, in or affecting interstate or foreign commerce, which such manufacturer reasonably believes, in the context of their intended distribution of the product, will be used to produce deep fake should (1) ensure such software has the technical capability to insert watermarks and disclosures of the nature described in such section into such deep fakes; and (2) include terms of use or other analogous disclosures with such software, which require the user of such software to affirmatively acknowledge their general awareness of their legal obligations under this Act. Holland clarifies, in addition to the obligations for software manufacturers, burdens will also fall on the public/ citizens, who would undoubtedly experience a chilling effect regarding the production of satirical or parodic political videos, where the legally required elements could easily be stripped; to say nothing of the vagueness of "humiliate or harass". There would also be burdens on local governments and law enforcement with respect to handling the legal actions arising from enforcement of the law.
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Any living individual or affiliated corporate or other entity who has been exhibited as engaging in falsified material activity in an advanced technological false personation record. Holland clarifies that there is at least an argument to be made that malicious creators of deepfake videos will benefit from the law, since they will not feel or be constrained by it, while legitimate users may experience a chilling effect, meaning that "real" videos, satire and parody may be in part displaced by deepfakes seeking to be mistaken as authentic. The exception in j(1)(F) certainly hints at the possibility that government actors will benefit by able to create deepfakes against their political opponents- it is difficult to imagine a national security or public safety situation that would necessitate the creation of a deepfake.

Option: Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or the "DEEP FAKES Accountability Act", (H.R. 3230) 116th Cong. (2019) Proposer/Sponsor: Rep. Yvette D. Clarke, [D-NY-9] (Introduced 12 June 2019) Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/3230/text Assessed by: TRI Date of assessment: 4 November 2019 Stakeholder(s) consulted in option assessment: Adam Holland, Berkman Klein Center	
Criteria/touch point	**Assessment**
11. Whose rights and/or interests does this option neglect?	Holland outlines that as with new burdens, it is software manufacturers who have had their rights neglected at the expense of this putative greater good. Arguably also the general public has had its rights neglected by the exception for government officials.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	In offering protection against "falsified speech, conduct, or depiction which causes, or a reasonable person would recognize has a tendency to cause perceptible individual or societal harm, including misrepresentation, reputational damage, embarrassment, harassment, financial losses, the incitement of violence, the alteration of a public policy debate or election, or the furtherance of any unlawful act," it might indirectly support the enjoyment of exercise of human rights via offering protection against reputational damage and tenuous protections against misinformation for the consuming public.
13. How does it address ethics and ethical principles? Which ones?	It explicitly addresses ethical principles/issues such as privacy, accountability, and individual harm.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	The Bill provides that on the effective date of the Act, the Attorney General shall publish a report containing, inter alia, a description of the impact of intimate and sexual deep fakes on women and marginalized communities.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Holland clarifies that the bill imposes real new responsibilities on government actors as well as creating new positions (the US AG coordinators). Not identifying where in the budget those should be accommodated seems like an error in the bill, or an unfunded mandate.
16. What provisions are there for regular review and update?	The Bill provides that "the Attorney General, in coordination with other relevant Federal agencies, shall submit a report to Congress 5 years after the date of enactment of this section, and 5 years thereafter, describing trends related to prosecutions and civil penalties pursued under this section, and recommending any updates to this section necessitated by the emergence of new technologies." One key measure outlined in the Bill is that on the effective date of this Act, the Attorney General shall publish a report containing, inter alia, (in order to increase the likelihood of such prosecutions), official guidance to Federal prosecutors regarding any potential legal concerns that may impede such prosecutions absent clarification.

Option: Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or the "DEEP FAKES Accountability Act", (H.R. 3230) 116th Cong. (2019) Proposer/Sponsor: Rep. Yvette D. Clarke, [D-NY-9] (Introduced 12 June 2019) Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/3230/text Assessed by: TRI Date of assessment: 4 November 2019 Stakeholder(s) consulted in option assessment: Adam Holland, Berkman Klein Center	
Criteria/touch point	**Assessment**
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Schmidt outlines, "Outlawing deepfakes isn't feasible or recommendable. Deepfake technology has many legitimate uses, especially in movies, where it is used to place an actor's face on their stunt double's body or retouch an actor's face when called for by the plot. Even if they were outlawed, deep fake's technology, much like copyright infringement, is likely impossible to fully prevent or ban. As the technology becomes more advanced, it will be easier to create deepfakes and, because the internet is borderless, deepfake creation software will always be accessible from places where the technology remains legal." See Schmidt, Nicholas, "Privacy law and resolving 'deepfakes' online", *IAPP Privacy Perspectives,* 30 Jan 2019. https://iapp.org/news/a/privacy-law-and-resolving-deepfakes-online/
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	Schmidt outlines that "The right to be forgotten, granted to European residents in Article 17 of EU General Data Protection Regulation as the "right to erasure," may assist a European victim of a deepfake. Under the right to be forgotten, a data subject has the right to request that the controller of personal data (i.e., the creator or publisher of the deepfake) about them delete that data. Data subjects can also object to the processing of their data under certain circumstances, likely to apply here. A deepfake, although fictional, counts as personal data under Article 4(1) of the GDPR, since it "relat[es] to an identified or identifiable natural person." Schmidt, Nicholas, "Privacy law and resolving 'deepfakes' online", *IAPP Privacy Perspectives,* 30 Jan 2019. https://iapp.org/news/a/privacy-law-and-resolving-deepfakes-online/ While a related communication has been issued by the European Commission, the competence to legislate with respect to deepfakes is better placed at the national EU Member State level (though the cross-border dimensions might make an European approach necessary for effective and coordinated action and to protect the EU, its citizens, its policies and its Institutions, as outlined in the Communication. See https://ec.europa.eu/digital-single-market/en/news/communication-tackling-online-disinformation-european-approach,
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	No

Option: Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or the "DEEP FAKES Accountability Act", (H.R. 3230) 116th Cong. (2019) Proposer/Sponsor: Rep. Yvette D. Clarke, [D-NY-9] (Introduced 12 June 2019) Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/3230/text Assessed by: TRI Date of assessment: 4 November 2019 Stakeholder(s) consulted in option assessment: Adam Holland, Berkman Klein Center	
Criteria/touch point	**Assessment**
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	2 (unlikely) or 3 (neutral). Given the Bill's current form, it might not survive a Constitutional challenge without being more narrowly drafted.
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	One factor critical to the success of such legislation will be whether the law is able to surmount the criticism that it only addresses the symptom and not the cause of the problem. It has been stated "writing legislation on these videos without touching the larger issues of disinformation, propaganda, and the social media algorithms that spread them misses the forest for the trees". See Ingram, Mathew, "Legislation aimed at stopping deepfakes is a bad idea", *Columbia Journalism Review*, 1 July 2019. https://www.cjr.org/analysis/legislation-deepfakes.php
References consulted	Coldewey, Devin, "DEEPFAKES Accountability Act would impose unenforceable rules — but it's a start", *TechCrunch,* 13 June 2019. http://techcrunch.com/2019/06/13/deepfakes-accountability-act-would-impose-unenforceable-rules-but-its-a-start/ Cushing, Tim, "Deep Fake' Legislation Is On The Way, Threatening Free Speech Protections" TechDirt, 9 July 2019. https://www.techdirt.com/articles/20190706/16124442526/deep-fake-legislation-is-way-threatening-free-speech-protections.shtml Ferraro, Matthew F., "*Deepfake Legislation: A Nationwide Survey—State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media",* 25 Sept 2019. https://www.wilmerhale.com/en/insights/client-alerts/20190925-deepfake-legislation-a-nationwide-survey Ingram, Mathew, "Legislation aimed at stopping deepfakes is a bad idea", *Columbia Journalism Review*, 1 July 2019. https://www.cjr.org/analysis/legislation-deepfakes.php Schmidt, Nicholas, "Privacy law and resolving 'deepfakes' online", IAPP Privacy Perspectives, 30 Jan 2019. https://iapp.org/news/a/privacy-law-and-resolving-deepfakes-online/ Tsukayama, Hayley, India Mckinney, and Jamie Williams, "Congress Should Not Rush to Regulate Deepfakes", Electronic Frontier Foundation, 24 June 2019.

Option: Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or the "DEEP FAKES Accountability Act", (H.R. 3230) 116th Cong. (2019) Proposer/Sponsor: Rep. Yvette D. Clarke, [D-NY-9] (Introduced 12 June 2019) Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/3230/text Assessed by: TRI Date of assessment: 4 November 2019 Stakeholder(s) consulted in option assessment: Adam Holland, Berkman Klein Center	
Criteria/touch point	**Assessment**
	https://www.eff.org/deeplinks/2019/06/congress-should-not-rush-regulate-deepfakes

4.19. Algorithmic Accountability Act of 2019 (HR 2231)

Option: Algorithmic Accountability Act of 2019 (HR 2231, 116th Congress) Proposer: Rep. Yvette D. Clark Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	Would require the US Federal Trade Commission to implement regulations to require covered entities that use, store or share personal information to conduct impact assessments for any high-risk automated decision system that makes a decision (or facilitates a human decision) that impacts consumers, and to reasonable address the results of the impact assessments in a timely manner. "Covered entities" are commercial entities over which the FTC has jurisdiction (which excludes federally regulated financial institutions and common carriers), that: • have over $50 million in annual gross receipts, or • possess or control personal information on more than 1 million consumers or consumer devices, or • is substantially owned, operated or controlled by an entity that meets either of the first two requirements, or • is a commercial data broker. (Sec. 2(5)) An automated decision system is "high-risk" if it: • poses a significant risk to the privacy or security of personal information of consumers, • poses a significant risk of resulting in or contributing to inaccurate, unfair, biased or discriminatory decisions impacting consumers, • makes or facilitates decisions based on consumer data profiling, that alters consumers' legal rights or otherwise significantly impacts consumers, • involves sensitive personal information of a significant number of consumers,

Option: Algorithmic Accountability Act of 2019 (HR 2231, 116th Congress) Proposer: Rep. Yvette D. Clark Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
	• systematically monitors a large, publicly accessible, physical space, or • meets other high-risk criteria set by the FTC. (Sec. 2(7)) The assessments would evaluate impacts on accuracy, fairness, bias, discrimination, privacy, and security, and would include detailed descriptions of: • a detailed description of the automated decision system and its design, training, data and purpose • an assessment of the relative benefits and costs of the automated decision system considering its purpose, taking into account data protection (data minimization, data retention, transparency, ability of consumers to correct or object to results, and the recipients of the decision results) • assessment of data security and data privacy risks to personal data • assessment of the risks that the system may result in or contribute to inaccurate, unfair, biased, or discriminatory decisions impacting consumers, and the measures taken to minimize these risks, including technological and physical safeguards (Sec. 2(2))
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: Proposed new federal law (under the US government's power to regulate interstate commerce) to require the Federal Trade Commission to enact new regulations Nature: Binding Scope: National (US)
3. Purpose/objective/what need does the option fulfil?	The proposal addresses the issue of algorithmic opacity and potential inaccuracy and bias in high-impact automated decision systems.
4. What gap does it address?	There are no current federal laws requiring commercial companies using AI systems to perform assessments on the systems' design, data and training for algorithmic accuracy, fairness or transparency, or to provide information on automated decision system results to consumers.
5. What added value does it have?	It provides basic protection for personal data used in automated decision systems on a national level and requires data protection impact assessments for high-risk automated decision systems.

Option: Algorithmic Accountability Act of 2019 (HR 2231, 116th Congress) Proposer: Rep. Yvette D. Clark Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*	Transparency: The law does not require covered entities to make the results of the algorithmic assessments public. Some commentators suggest that there is insufficient transparency. Reference: (Jones Day, *Proposed Algorithmic Accountability Act Targets Bias in Artificial Intelligence* (June 2019), https://www.jonesday.com/en/insights/2019/06/proposed-algorithmic-accountability-act New, Joshua, *How to Fix the Algorithmic Accountability Act* (23 Sep 2019) https://www.datainnovation.org/2019/09/how-to-fix-the-algorithmic-accountability-act/ Neutrality: The act requires covered entities to conduct required assessments in consultation with external third parties (e.g. independent auditors and independent technology experts) ***if*** reasonably possible. Some commentators suggest the law does not go far enough to require neutrality in the assessment. (Jones Day) Operational burden: The law doesn't specify how often algorithmic assessments must be updated ("as frequently as the Commission determines is necessary"). Some commentators suggest this could be unduly burdensome given the iterative nature of software development. (Joshua New)
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	Yes, though the FTC would need to define some of the requirements, such as how often to require updated algorithmic assessments when an automated decision system changes.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	The law empowers the Federal Trade Commission and state Attorneys General to monitor and enforce the new regulations. However, since covered entities are not require to make assessments public or to report assessments to authorities, it's unclear how the FTC or state AGs would become aware of the need for assessments.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Citizen: No burdens Public administrations: The FTC would be required to implement new regulations within 2 years after the law is passed. Large businesses: Any large business using AI for high-impact decisions would be required to perform algorithmic and data protection assessments before the system is implemented and periodically (frequency to be determined by the FTC) thereafter, and to respond to the assessment results. This will require time and resources to perform. Small and medium enterprises: Most SMEs will not be covered by the requirements unless they possess or control personal data on 1 million consumers or consumer devices or are data brokers. In

Option: Algorithmic Accountability Act of 2019 (HR 2231, 116th Congress) Proposer: Rep. Yvette D. Clark Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
	either such case, they would need to devote additional time and resources to conduct and respond to the assessments.
10. Which stakeholders would benefit most from the use of this option?	Individuals
11. Whose rights and/or interests does this option neglect?	The proposal does not address AI systems used by public entities or federally regulated financial institutions for high-risk automated decisions
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	The proposal explicitly supports human rights by requiring assessments to identify and reduce risks that high-impact automated decisions will result in or contribute to biased, or discriminatory decisions impacting consumers.
13. How does it address ethics and ethical principles? Which ones?	The proposal requires the assessment of high-impact automated decision systems to identify and reduce the risk of inaccurate or unfair decisions impacting consumers.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	Not explicitly addressed, but the proposal is based on identifying and reducing risks of biased and discriminatory decisions, which includes gender bias.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not addressed
16. What provisions are there for regular review and update?	Not addressed
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Feasible: The proposal is feasible; in that it is within the scope of the FTC's existing authorization. It will likely face political opposition from industry. Sustainable: If implemented, it is likely that there will eventually be market incentives for covered entities to comply and to improve the results of their algorithmic assessments. Knowledgeable buyers of high-impact AI systems (in particular, large public companies and government entities) are likely to require visibility of the algorithmic assessments and remediation measures as part of their procurement processes. Future-proof: The nature of the assessments will need to be able to change to reflect technological developments. The current

Option: Algorithmic Accountability Act of 2019 (HR 2231, 116th Congress) Proposer: Rep. Yvette D. Clark Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
	proposal sets minimum standards for assessments, but the FTC can increase the requirements by regulation, if needed.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Yes, this will increase the time and resources that covered entities will need to develop AI systems that are used for high-impact automated decisions. The proposal will require covered entities to incorporate algorithmic fairness into design and testing processes.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The proposal strengthens requirements for users of automated decision systems to protect the human rights of data subjects who are affected by the automated decisions, so is suitable for use within the European Union legal framework, especially the EU General Data Protection Regulation (GDPR). However, as written, it is not fully consistent with the requirements of the GDPR. For example, the assessment must "consider the extent to which consumers have access to the results of the automated decision systems and may correct or object to its results", but the law doesn't require that consumers actually *have* the opportunity to correct or object to the results. The law also doesn't address other GDPR requirements for automated decision making (such as consent and notice). This proposal could supplement existing EU data protection laws by further defining the obligations of users of high-risk automated decision systems to assess the accuracy and fairness of the systems' design, data and testing.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	None identified.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	4 (likely), *if* it is passed into law. However, under the current Republican control of the US Senate and the current administration, the proposal is unlikely to be passed this session.
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Industry opposition would be lessened if the FTC provides: • Guidelines to help determine whether an application is high-risk • Tools and guidelines to help with the algorithmic assessment (such as the interactive assessment tool being developed by the Treasury Board in Canada for the Directive on Automated Decision-Making) Industry compliance will be strengthened if federal procurement standards require that covered entities provide evidence of compliance with the new requirements as part of the procurement process.

Option: Algorithmic Accountability Act of 2019 (HR 2231, 116th Congress) Proposer: Rep. Yvette D. Clark Reference/link to relevant document: https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info Assessed by: UCLANCY, date assessed 11 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
References consulted	Jones Day, *Proposed Algorithmic Accountability Act Targets Bias in Artificial Intelligence* (June 2019), https://www.jonesday.com/en/insights/2019/06/proposed-algorithmic-accountability-act New, Joshua, *How to Fix the Algorithmic Accountability Act* (23 Sep 2019) https://www.datainnovation.org/2019/09/how-to-fix-the-algorithmic-accountability-act/

4.20. Directive on Automated Decision-Making (Canada)

Option: Directive on Automated Decision-Making Proposer: **Government of Canada** Reference/link to relevant document: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 Assessed by: UCLANCY, date assessed 4 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	Defines level of review required for public procurement of AI-based systems to be used to make or assist with administrative decisions that affect a client's legal rights, privileges, or interests. Requires a risk-based algorithmic impact assessment before production use of the system and for any later changes to the system's functionality or scope. Requires vendors of proprietary systems to allow the government (directly or through third parties) to review and audit the proprietary software. Applies to any system for automated decision-making that is developed or procured after 1 Apr 2020.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: Federal regulation issued under the authority of federal laws (section 7 of the Financial Administration Act). Applies to federal departments of the Government of Canada, excluding the Office of the Governor General's Secretary and the staffs of the Senate, House of Commons, Library of Parliament, Office of the Senate Ethics Officer and Office of the Conflict of Interest and Ethics Commissioner. Other departments or separate agencies are encouraged to meet these requirements as good practice. Does not apply to national security systems.

Option: Directive on Automated Decision-Making Proposer: **Government of Canada** Reference/link to relevant document: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 Assessed by: UCLANCY, date assessed 4 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
3. Purpose/objective/what need does the option fulfil?	Section 4.1 states that the objective of the Directive is to "ensure that Automated Decision Systems are deployed in a manner that reduces risks to Canadians and federal institutions, and leads to more efficient, accurate, consistent, and interpretable decisions made pursuant to Canadian law"[1].
4. What gap does it address?	Addresses the lack of a mechanisms to ensure algorithmic assessment, visibility and fairness in government systems used for automated decisions affecting the legal rights, privileges, or interests of an external client.
5. What added value does it have?	Establishes a mechanism and requirement for risk-based algorithmic impact assessments prior to production use of automated decision-making systems by federal agencies.
6. What are the limitations, risks and challenges?	The Directive is broader than the requirements for automated decision-making under the GDPR, as it also applies to systems that assist in human-directed decisions. "Automated decision system" is defined very broadly: udes any technology that either assists or replaces the judgement of human decision-makers. These systems draw from fields like statistics, linguistics, and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural nets.[1] As pointed out by commentators, the Directive: • Also applies to tools other than purpose-built software applications, such as Excel models and mathematical formulas[2] • Could allow a vendor's competitors to have access to the vendor's source code if the competitor is engaged to act as an expert for the peer review process as part of the algorithmic impact assessment, for certain risk levels[2] • Uses an interactive assessment tool with multiple-choice questions, but does not include the requirement to provide proof to support an answer (e.g. "Do you have a policy to…" vs "Attach a copy of your policy to…")[3] • Requires testing for algorithmic bias, but doesn't define the form or type of bias[4] Four risk levels are identified for the algorithmic impact assessment. The level of review required prior to system implementation increases as the risk level increases. However, the risk levels are defined subjectively, which may lead to inconsistent assignment of risk levels and reviews: • Level I: Decisions will often lead to impacts (on individual, communities or ecosystems) that are reversible and brief • Level II: Impacts are likely reversible and short-term • Level III: Impacts can be difficult to reverse, and are ongoing • Level IV: Impacts are irreversible, and are perpetual

Option: Directive on Automated Decision-Making Proposer: **Government of Canada** Reference/link to relevant document: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 Assessed by: UCLANCY, date assessed 4 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	The Treasury Board is also developing an interactive algorithmic assessment "scorecard" tool, but the scoring is not directly tied to the four risk levels.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	Almost. See answer in #6 above about subjective descriptions for the 4 risk levels which determine the level of review required under the Directive.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	The Assistant Deputy Minister responsible for the program using the automated decision system (or his/her delegate) is responsible for ensuring that the assessment, transparency, quality assurance, and reporting requirements are met. Consequences for failure to comply with the requirements include[5]: • A range of consequences for the noncompliant agency, from increased operational reviews to freezing of allotments or constraining high value transactions; and • A range of consequences for the responsible individual(s) from additional training to disqualification from public sector employment There does not appear to be a mechanism for direct recourse by an affected data subject.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Citizens None Public administrations Requires new compliance processes including (depending on risk level): - Conducting algorithmic impact assessments and risk assessments - Issuing public notices and reports - Providing an explanation of decisions - Testing for unintended data biases - Ongoing monitoring for unintended outcomes - Validating data accuracy, currency and relevance - Conducting a peer review and employee training - Establishing contingency systems or processes - Obtaining legal review Businesses and particularly SMEs Requires vendors of automated decision systems to allow access to the systems to assist with assessment or validation of the government's reviews, tests and audits. As a practical matter, federal procurement officers will likely require vendors to provide much of the information to validate algorithmic fairness and lack of bias.

Option: Directive on Automated Decision-Making Proposer: **Government of Canada** Reference/link to relevant document: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 Assessed by: UCLANCY, date assessed 4 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
10. Which stakeholders would benefit most from the use of this option?	Individuals who are subject to the automated decisions
11. Whose rights and/or interests does this option neglect?	This Directive only covers automated decision systems procured by federal agencies (other than national security systems), so it doesn't address use of SIS by provincial or local governments or by non-governmental entities.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	Non-discrimination and equality: It directly addresses bias in data and algorithms used for automated decisions but doesn't address types of biases to be checked. Data protection and privacy: It support practices consistent with GDPR protections for automated decision making. General: The risk levels take into account the impacts of automated decisions on the rights of individuals or communities; the health or well-being of individuals or communities; the economic interests of individuals, entities, or communities; and the ongoing sustainability of an ecosystem.
13. How does it address ethics and ethical principles? Which ones?	Procedural fairness: The degree of procedural fairness that the law requires for any given decision-making process increases or decreases with the significance of that decision and its impact on rights and interests.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	No. Each federal department is required to comply with the Directive as part of its procurement process.
16. What provisions are there for regular review and update?	The Directive has an automatic review process planned every 6 months after the effective date.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g.,	It is feasible, though will require vendors to provide more algorithmic accountability than they are currently required to provide. Market forces (e.g. the desire to sell into the Canadian federal market) will support adoption by vendors. The tool for algorithmic impact assessment is not mandated by the Directive. A tool is being developed by the Treasury Board to assist with

Option: Directive on Automated Decision-Making Proposer: **Government of Canada** Reference/link to relevant document: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 Assessed by: UCLANCY, date assessed 4 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
technological, policy changes, social demands?	the assessment by providing a scorecard. The tool can be updated for future technological changes.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	It does not directly impact the ability for businesses and others to innovate on their own. It may make it more difficult for businesses to incorporate technology from third parties into their solutions, if the third-party does not support the audits, reviews and transparency required by the Directive.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	It could fit within the framework of the GDPR, if the EU chose to add additional protections for automated decision-making.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	No
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	4 – likely (after initial resistance by vendors)
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Providing easy-to-use tools for the risk assessment and algorithmic impact assessment
References consulted	1. Government of Canada, Directive on Automated Decision Making, effective 1 April 2019. https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 2. Fekete, Michael, and Sam Ip, "Government of Canada's Directive on Automated Decision-Making: Implications for service providers", 6 June 2019. https://www.osler.com/en/resources/regulations/2019/government-of-canada-s-directive-on-automated-decision-making-implications-for-service-providers 3. Lemay, Mathieu," Understanding Canada's Algorithmic Impact Assessment Tool", 11 June 2019. https://towardsdatascience.com/understanding-canadas-algorithmic-impact-assessment-tool-cd0d3c8cafab

Option: Directive on Automated Decision-Making Proposer: **Government of Canada** Reference/link to relevant document: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 Assessed by: UCLANCY, date assessed 4 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	4. Canadian Bar Association, Letter to the Minister of Immigration, Refugees and Citizenship regarding Artificial Intelligence and Machine Learning in Immigration Law, 11 July 2019. http://www.cba.org/CMSPages/GetFile.aspx?guid=c54903f5-cd8a-4d3a-96a3-ce0c33623845 5. Government of Canada, Framework for the Management of Compliance, Appendices C and D, effective 1 April 2009, last modified 27 August 2010. https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=17151

4.21. US Food and Drug Administration regulation of adaptive AI/ML technology

Option: US Food and Drug Administration regulation of adaptive AI/ML technology Proposer: US Food and Drug Administration Reference/link to relevant document: https://www.fda.gov/media/122535/download Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	In April 2019 the US Food and Drug Administration ("FDA") solicited feedback to a discussion paper[1] proposing a new regulation framework to allow for iterative modifications in medical devices that use artificial intelligence and machine learning (AI/ML medical systems). (Under the Food, Drug and Cosmetics Safety Act, software that is intended to treat, diagnose, cure, mitigate, or prevent disease or other conditions is within the definition of a regulated "medical device".) FDA approval of medical devices has historically been based on a static (unchanging) physical device. With the advent of software as a medical device, the FDA adopted policies and processes that require pre-market review and approval of any software update that: • Introduces a new risk or modifies an existing risk that could result in significant harm; Changes risk controls to prevent significant harm; or Significantly affects clinical functionality or performance specifications of the device. With AI/ML medical systems, change can happen continuously. Requesting pre-market review and approval for each change is not feasible for such systems. The discussion paper proposes a total product lifecycle regulatory approach for AI/ML medical systems with includes: • FDA review of the developer's organization, development and test processes, and quality assurance program;

Option: US Food and Drug Administration regulation of adaptive AI/ML technology Proposer: US Food and Drug Administration Reference/link to relevant document: https://www.fda.gov/media/122535/download Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
	• An initial pre-market review of the developer's change control specifications and protocols for performance changes and algorithmic changes; • After initial approval, a risk-based review of ongoing changes to determine whether change is within the preauthorized change scope (and requires only documentation of the change) or whether additional approval is required; • Ongoing performance monitoring and periodic reporting to the FDA.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: Existing law (the FDA enforces the Federal Food, Drug and Cosmetics Safety Act, and controls the approval of medical devices). Nature: Binding Scope: National
3. Purpose/objective/what need does the option fulfil?	The goal is to ensure that ongoing algorithm changes to AI/ML medical systems are: • implemented according to pre-specified performance objectives, • follow defined algorithm change protocols, • utilize a validation process that is committed to improving the performance, safety, and effectiveness of AI/ML software, and • include real-world monitoring of performance.
4. What gap does it address?	Current regulatory processes are based on approval of a static design, with re-approval required for material changes. This isn't feasible for dynamic software systems using artificial intelligence and machine learning.
5. What added value does it have?	This approach may also be useful for regulation of autonomous AI/ML-enabled systems outside the medical context, such as drones and autonomous vehicles.
6. What are the limitations, risks and challenges?	Over 130 comments[2] have been submitted in response to the discussion paper, including the following limitations, risks and challenges: • The framework doesn't address testing for and remediating racial and gender bias in existing datasets and health/medical care algorithms.[3] • The framework doesn't address data privacy for the patient data.[4]
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	No. The discussion paper notes that the new framework may require additional statutory authority, and notes that it is not intended to communicate FDA's proposed (or final) regulatory expectations but is meant to seek early input from groups and individuals outside the FDA prior to development of a draft guidance.

Option: US Food and Drug Administration regulation of adaptive AI/ML technology Proposer: US Food and Drug Administration Reference/link to relevant document: https://www.fda.gov/media/122535/download Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Insufficient information
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Citizens: Not applicable Public administrations: The FDA would become involved in the review process for AI/ML medical systems during the development process, not just at the point of pre-market review. Developers of AI/ML medical systems would need to provide more information to the FDA about their software development, testing and performance monitoring processes, as well as continuous improvement information as the system continues to develop after market introduction.
10. Which stakeholders would benefit most from the use of this option?	Developers of AI/ML medical systems would benefit from a more dynamic and flexible review structure, in place of having to submit each significant software update for pre-approval. The regulator would benefit from greater visibility into the ongoing develop, testing and quality assurance processes of the developers. Medical professionals and patients would benefit from the use of systems that are constantly being improved, with oversight and within safety guidelines.
11. Whose rights and/or interests does this option neglect?	None identified
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	It does not explicitly support or adversely affect human rights but supports systems that can help improve health and reduce suffering from medical causes.
13. How does it address ethics and ethical principles? Which ones?	It provides additional transparency to the FDA and to users of the AI/ML medical systems with performance reporting for maintaining continued assurance of safety and effectiveness.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No.

Option: US Food and Drug Administration regulation of adaptive AI/ML technology Proposer: US Food and Drug Administration Reference/link to relevant document: https://www.fda.gov/media/122535/download Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	No separate funding mechanism was addressed.
16. What provisions are there for regular review and update?	None identified.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Feasible: Yes. It adapts a current regulatory approach to the more flexible requirements of AI/ML medical systems. Sustainable: Yes. The initial comments to the discussion paper were generally supportive of the approach, with suggestions for changing some of the specific categories, criteria or steps. Future-proof: Possibly. The more flexible review and approval process may increase communication between the regulators and developers. As the technology changes, the change processes and protocols that are submitted for review and approval will change. With the increased communication between the regulators and developers, the regulators' approval criteria may be able to change.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No, this more flexible approach will encourage development of AI/ML medical systems.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	Software is included as a regulated medical device under the 2017 EU Medical Device Regulation[5].
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 –	4 (likely).

Option: US Food and Drug Administration regulation of adaptive AI/ML technology Proposer: US Food and Drug Administration Reference/link to relevant document: https://www.fda.gov/media/122535/download Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
Neutral, 4 – likely 5 – Extremely likely)?	
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	The discussion paper presents a flexible framework to adapt regulatory overview of medical devices to AI/ML medical devices, with the potential to increase quality and transparency over the entire lifecycle of the device while reducing the need for pre-approval of pre-authorized changes.
References consulted	1. Discussion paper 2. Public comments can be viewed at https://www.regulations.gov/docketBrowser?rpp=25&po=0&dct=PS&D=FDA-2019-N-1185&refD=FDA-2019-N-1185-0001 3. Public comments posted by Chloe Nichols, 12 Nov 2019 https://www.regulations.gov/document?D=FDA-2019-N-1185-0133 and by Martin Haimerl, 17 May 2019 https://www.regulations.gov/document?D=FDA-2019-N-1185-0032 4. Public comment posted by Shaillay Dogra, 7 Jun 2019. https://www.regulations.gov/document?D=FDA-2019-N-1185-0092 5. European Parliament and the Council, Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, OJ L 117, 5 Apr 2017. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32017R0745

4.22. New statutory duty of care for online harms (UK Government)

Option: New statutory duty of care for online harms (UK Government) Proposer: UK Government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance /connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	In April 2019 the UK Secretary of State for Digital, Culture, Media & Sport and the Secretary of State for the Home Department submitted the Online Harms White Paper[1] ("White Paper") to the UK Parliament, which included a proposal to establish a new regulatory scheme regarding online harmful user-generated content ("UGC"), to be managed by a new independent regulator. The new regulatory scheme would apply to entities that offer services or tools that allow users to: • share or discover UGC, or • interact with each other online. Examples of in-scope services and tools are social media platforms, file hosting sites, public discussion forums, messaging services search engines, and caching tools that include UGC. The White Paper proposes a new "statutory duty of care" to be undertaken by covered entities, but doesn't define the duty, other than protecting against harmful UGC without unduly limiting users' privacy and freedom of expression online. Covered entities would be required to meet new codes of practice to be developed by the regulator, or to demonstrate compliance with the duty of care through other means acceptable to the regulator. The White Paper excludes certain types of harm from the scope of the new duty of care: • harms suffered by organisations, not individuals • data protection breaches and harms caused by cybersecurity breaches • harms suffered by individuals on the dark web rather than the open internet. This regulatory scheme **does not** directly regulate smart information systems. (AI is one tool that covered entities may use to monitor for harmful UGC.)
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: The White Paper proposes new laws. Nature: Binding Scope: The proposed regulatory scheme would apply to all companies (globally) that provide covered services to users within the UK.

Option: New statutory duty of care for online harms (UK Government) Proposer: UK Government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
3. Purpose/ objective/what need does the option fulfil?	To require providers of online services that allow access to UGC or user-to-user interactions to have an affirmative obligation to monitor and restrict harmful content, not just to respond when they are informed of or become aware of harmful content.
4. What gap does it address?	Today most of the covered entities are only required to remove or restrict harmful content when they are notified or become aware of it, but not to take proactive steps to monitor for harmful content.
5. What added value does it have?	Implementation of the proposal will increase market demand for (and spur further development of) tools to analyse content and online behaviour, likely based on smart information systems.
6. What are the limitations, risks and challenges?	Commentators[2,3] have criticized the proposal for: • Posing a serious risk to freedom of expression without identifying how freedom of expression would be protected, and incentivising the removal of speech; • Placing too much faith in technology to help covered entities comply with the oversight requirements; • Lacking a clear delineation of legal but "harmful" content to be regulated • Not identifying responsibility for oversight of the regulator
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	No. The White Paper includes 18 open consultation questions to be addressed before legislation is drafted to implement the proposal, such as: • What role should Parliament play in scrutinising the work of the regulator, including the development of codes of practice? • Should an online harms regulator be: (i) a new public body, or (ii) an existing public body? • Which channels or forums that can be considered private should be in scope of the regulatory framework? • Should the regulator be empowered to i) disrupt business activities, or ii) undertake ISP blocking, or iii) implement a regime for senior management liability? What, if any, further powers should be available to the regulator?
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	The monitoring, oversight and enforcement mechanisms are among the open questions for which additional input has been requested (see #7 above).

Option: New statutory duty of care for online harms (UK Government) Proposer: UK Government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Insufficient information to determine. This will depend on the mechanisms that the new regulator develops for codes of practice and for reporting processes for citizens.
10. Which stakeholders would benefit most from the use of this option?	Individuals would benefit most directly from this option.
11. Whose rights and/or interests does this option neglect?	The option will require a careful balancing of the interests of individuals to be protected from "harmful" content (which is not defined) against the interests of privacy and free expression.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	The regulatory scheme is designed to reduce illegal, dangerous and otherwise harmful UGC and user interactions, such as activities that promote child sex abuse, hate speech, violence, and terrorism, and activities that threaten democratic values and principles. Achieving this goal support the right to life, freedom from slavery, the right not to be discriminated against, and the right to participate in free elections. However, there is risk that excessive monitoring and control of UGC can jeopardize free expression, freedom of assembly, and the right to privacy.
13. How does it address ethics and ethical principles? Which ones?	The regulatory scheme is intended to reduce deception, intimidation, and other content or activities that promote harm to individuals or groups.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the	No, though online misogynistic abuse is identified as an example of harmful UGC.

Option: New statutory duty of care for online harms (UK Government) Proposer: UK Government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
agency/body, consideration of gender equality, gender neutrality.	
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	The White Paper proposes that new fees, charges or a levy on in-scope entities will cover the costs of the regulator, and solicits input for the question "on what basis should any funding contributions from industry be determined?"
16. What provisions are there for regular review and update?	Not identified.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Feasible: Insufficiently defined (how will the regulator define "harmful" content and behaviour?) Sustainable: Insufficiently defined Future-proof: The regulator would have the ability to update codes of practice to reflect changes in technology, but rapidly changing technology may make it challenging for the regulator and covered entities to keep up.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Unknown. The White Paper acknowledges that the regulation must be designed carefully to minimize any negative impact on innovation and suggests that clear guidelines will facilitate innovation by clarifying the duties of covered entities. The impact on innovation will largely depend on the nature of the practice codes developed by the regulator, and the availability of tools to comply with such codes.

Option: New statutory duty of care for online harms (UK Government) Proposer: UK Government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis) 20.	The White Paper suggests that the new regulatory framework is compatible with the EU's e-Commerce Directive[4], which requires online service providers to act on illegal UGC once they have been notified of or become aware of its existence. The new regulation would require the online provides to take proactive steps to identify and remove (broadly-defined) "harmful" content (which might not be illegal) more quickly, which appears to conflict directly with the Article 15 Section 1 of the eCommerce Directive, which states that:Member States shall not impose a general obligation on providers, when providing the services covered by Articles 12, 13 and 14, to monitor the information which they transmit or store, nor a general obligation actively to seek facts or circumstances indicating illegal activity. There is also the potential for the new regulation to conflict with users' rights to privacy and free expression.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	-
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	2 (unlikely) as proposed, given the lack of a clear definition of "harmful" but legal UGC, significant concerns over threats to free expression and privacy, and (at least until Brexit) conflict with the e-Commerce Directive.
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	This regulatory scheme is directed to service providers, not directly to developers or users of smart information systems (though many covered entities may use SIS to manage compliance with the regulatory requirements).
References consulted	1. Online Harms White Paper, Presented to Parliament by the Secretary of State for Digital, Culture, Media & Sport and the Secretary of State for the Home Department, CP 57, April 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf

Option: New statutory duty of care for online harms (UK Government) Proposer: UK Government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	2. Online Harms White Paper: Seven Expert Perspectives, Digital Action, April 2019, including comments from Article 19, Institute for Strategic Dialogue, Mozilla, Child Rights International Network, Sophia Ignatidou, Global Partners Digital, Demos - Centre for Analysis of Social Media. https://www.politico.eu/wp-content/uploads/2019/04/Seven-expert-perspectives-on-the-UK-online-harms-White-Paper-.pdf?utm_source=POLITICO.EU&utm_campaign=723cb52285-EMAIL_CAMPAIGN_2019_04_10_05_07&utm_medium=email&utm_term=0_10959edeb5-723cb52285-189780761 3. Smith, Graham, "Users Behaving Badly – the Online Harms White Paper", Cyberleagle blog, 18 April 2019. https://www.cyberleagle.com/2019/04/users-behaving-badly-online-harms-white.html 4. European Parliament and the Council, Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), OJ L 178, 17.07.2000. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32000L0031&from=EN

4.23. Redress by design mechanisms for AI

Option: Redress-by-design mechanisms for AI Proposer: High-Level Expert Group on Artificial Intelligence (AI HLEG) Reference/link to relevant document: Policy and investment recommendations for trustworthy Artificial Intelligence, 2019. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 28 Oct 2019 Stakeholder(s) consulted in option assessment: *Giuseppe Stefano Quintarelli, Copernicani*	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example	The AI HLEG Policy recommendations 2019 envision redress-by-design mechanisms as "establishing – from the design phase – mechanisms to ensure alternative systems and procedures with an adequate level of human oversight (human in the loop, on the loop or in command approach) to be able to effectively detect, audit, and rectify incorrect decisions taken by a "perfectly" functioning system, for those situations where the AI system's decisions significantly affects individuals." The AI HLEG Guidelines explicitly state: Oversight may be achieved through governance mechanisms such as a human-in-the-loop (HITL), human-on-the-loop (HOTL), or human-in-

Option: Redress-by-design mechanisms for AI Proposer: High-Level Expert Group on Artificial Intelligence (AI HLEG) Reference/link to relevant document: Policy and investment recommendations for trustworthy Artificial Intelligence, 2019. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 28 Oct 2019 Stakeholder(s) consulted in option assessment: *Giuseppe Stefano Quintarelli, Copernicani*	
Criteria/touch point	**Assessment**
	command (HIC) approach. HITL refers to the capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable. HOTL refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. HIC refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation. This can include the decision not to use an AI system in a particular situation, to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by a system. Moreover, it must be ensured that public enforcers have the ability to exercise oversight in line with their mandate. Oversight mechanisms can be required in varying degrees to support other safety and control measures, depending on the AI system's application area and potential risk. All other things being equal, the less oversight a human can exercise over an AI system, the more extensive testing and stricter governance is required.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: Such mechanisms could be based in a provision in an EU Regulation (new or added to existing one) in the same fashion as Article 25 of the General Data Protection Regulation on data protection by design and default. So, redress-by-design mechanisms would need to be implemented/integrated into AI systems taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by AI systems. Organisations would need to implement appropriate redress by design mechanisms, in an effective manner and to integrate the necessary safeguards. Such mechanisms could be taken the form of technical, organisational or procedural mechanisms. An approved certification mechanism could be used as an element to demonstrate compliance with the requirements. Nature and scope: Such mechanisms would not be mandatory but they should be considered and implemented as appropriate. Applicable at all levels.
3. Purpose/objective/what need does the option fulfil?	Redress by design mechanisms would enhance human oversight and good AI governance where such technologies

Option: Redress-by-design mechanisms for AI Proposer: High-Level Expert Group on Artificial Intelligence (AI HLEG) Reference/link to relevant document: Policy and investment recommendations for trustworthy Artificial Intelligence, 2019. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 28 Oct 2019 Stakeholder(s) consulted in option assessment: *Giuseppe Stefano Quintarelli, Copernicani*	
Criteria/touch point	**Assessment**
	and systems could or might have the adverse impacts on human rights. Given that we know that non-defective systems will generate incorrect predictions, we need to align the overall output of AI systems to socially desirable and politically agreed targets, considering the overall effect of the system and not only of one single instance. For the same reason, in the single instance of incorrect prediction by a non-defective system we need to ensure that the risk of adverse effects on individuals is minimized to an acceptable level.
4. What gap does it address?	It would address gaps in human oversight and governance of AI. Quintarelli explains, "redress itself is not enough, not accessible to all, does not correct all spill-overs that take place. Systems that affect human individuals should have redress by design built in – before decisions are finalised there should be a possibility to redress". See https://www.youtube.com/watch?v=a_u3AzLTOY0
5. What added value does it have?	As outlined by Quintarelli in the AI context, "redress by design relates to the idea of establishing, from the design phase, mechanisms to ensure redundancy, alternative systems, alternative procedures (depending on the value at stake and its possible impact, it may be automatic, or may be HITL, HOTL, HIC), etc. in order to be able to effectively detect, audit, rectify the wrong decisions taken by a perfectly functioning system and, if possible, improve the system." See https://blog.quintarelli.it/2019/04/we-need-redress-by-design-for-ai-systems.html
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with) MCAS for flights, medicines, ...*	Limitations: Additional developer/deployer/user burdens. Risks: Unclear or too restricted interpretation of such mechanisms. Challenges: Implementation challenges (e.g., lack of good implementation models of such mechanisms). Transparency of such mechanisms. There might be some resistance.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	Not elaborated in AI HLEG policy recommendations. However, such mechanisms could follow the model of privacy/data protection by design.

Option: Redress-by-design mechanisms for AI Proposer: High-Level Expert Group on Artificial Intelligence (AI HLEG) Reference/link to relevant document: Policy and investment recommendations for trustworthy Artificial Intelligence, 2019. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 28 Oct 2019 Stakeholder(s) consulted in option assessment: *Giuseppe Stefano Quintarelli, Copernicani*	
Criteria/touch point	**Assessment**
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Not elaborated in AI HLEG policy recommendations but the mechanisms themselves support this. There is definite possibility for this and such mechanisms should have such monitoring, oversight and enforcement attached to them. Quintarelli highlights that HLEG recommendation 30.5 could be read through the lens of redress by design – in terms of oversight and evolution of the framework, guidance avoiding too restricted interpretations, ensuring alignment with social targets and from which standards can be derived that can operate EU-wide followed by dissemination and inclusion.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Citizens: no. Public administrations: yes Businesses and particularly SMEs: yes, though this will be minimal as they will need to take steps/adopt measures and this increases their responsibility and need to act to protect values at stake. But the burdens should be proportional to what is sought to be protected.
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Individuals (on whom there are economic, societal, legal and ethical impacts) e.g., access to credit, housing, reputation scoring. It will also benefit SMEs who are the weaker players in the regulatory landscape.
11. Whose rights and/or interests does this option neglect?	None
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	If implemented, it will have a good effect and support human rights by increasing accountability with regard to AI systems.
13. How does it address ethics and ethical principles? Which ones?	It would address principles such as prevention of harm, explicability, human agency and oversight, and accountability.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	Not elaborated in AI HLEG Policy recommendations but indirectly it accommodates these, when read in conjunction with par.30.5.4.

Option: Redress-by-design mechanisms for AI Proposer: High-Level Expert Group on Artificial Intelligence (AI HLEG) Reference/link to relevant document: Policy and investment recommendations for trustworthy Artificial Intelligence, 2019. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 28 Oct 2019 Stakeholder(s) consulted in option assessment: *Giuseppe Stefano Quintarelli, Copernicani*	
Criteria/touch point	**Assessment**
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not elaborated in AI HLEG Policy recommendations. It could draw from the same sources as privacy be design/data protection by design. The cost burdens would be on the user/implementer of such mechanisms.
16. What provisions are there for regular review and update?	Not elaborated in AI HLEG Policy recommendations. Redress by design mechanisms could benefit from review by an external body or agency akin to data protection authorities. This could be ex ante (prior checking of processes and provision of advice) or ex post (audit). Use of ex ante would ensure companies are more proactive in the uptake and use of such mechanisms. A combination of both would be good. Para 30.5 of the Policy recommendations could provide some inspiration.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Such mechanisms are future-proof to the extent that they can easily adapt/align with societal values as they change. They can fit or be adapted to what we determine to be desirable futures and can help society achieve such results.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No. It would support businesses in responsible innovation.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	It is a definite fit. Such mechanisms could be introduced via a new EU Regulation (though there might not be the political appetite for this) or could be introduced into an existing Regulation via revision process. (e.g., we would re-build the GDPR for AI)
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	Politics. Vetoes by certain countries.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	5. [Such mechanisms will become activated in the EU and if not, the US via the courts whether in this form or another]

Option: Redress-by-design mechanisms for AI Proposer: High-Level Expert Group on Artificial Intelligence (AI HLEG) Reference/link to relevant document: Policy and investment recommendations for trustworthy Artificial Intelligence, 2019. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Assessed by: TRI Date of assessment: 28 Oct 2019 Stakeholder(s) consulted in option assessment: *Giuseppe Stefano Quintarelli, Copernicani*	
Criteria/touch point	**Assessment**
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	People need safeguards to protect themselves from unjust and adverse decisions by AI systems – As the AI HLEG outlines, "When unjust adverse impact occurs, accessible mechanisms should be foreseen that ensure adequate redress". See AI HLEG Ethics Guidelines for Trustworthy AI, 2019. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top One factor that might contribute to its adoption and success is a 'crisis' which will make people more aware and force action. Consumer associations and civil society organisations support could boost the adoption of such mechanisms.
References consulted	High-Level Expert Group on Artificial Intelligence (AI HLEG), Ethics Guidelines for Trustworthy AI, 2019. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top High-Level Expert Group on Artificial Intelligence (AI HLEG), Policy and investment recommendations for trustworthy Artificial Intelligence, 2019. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence Tutt, Andrew, "An FDA for Algorithms", 69 Admin. L. Rev. 83, 2017. Quintarelli, S, We need "redress by design" for AI systems, Quinta's weblog, 8 April 2019. https://blog.quintarelli.it/2019/04/we-need-redress-by-design-for-ai-systems.html

4.24. Register of algorithms used in government

Option: Register of algorithms used in government Proposer: New Zealand Law Foundation and University of Otago Reference/link to relevant document: Government use of artificial intelligence in New Zealand.https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf Assessed by: TRI Date of assessment: 14/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	The New Zealand Law Foundation has proposed the creation of an independent regulatory agency to oversee and regulate AI which would work with individual government agencies who intend either to introduce a new predictive algorithm, or to use an existing predictive algorithm for a new purpose. Among other tasks and responsibilities, it is suggested that this regulatory agency should be also responsible for maintaining a **register of algorithms used in government. T**he focus of this recommendation is on predictive algorithms and uses of AI by governmental departments. Applying the method of analogy, the New Zealand Law Foundation considered the appropriateness and effectiveness of the regulatory option of a register of algorithms. It explains that New Zealand has recently recognised natural entities as legal persons and, based on this, there is a precedent for an extension of the category of artificial persons within domestic law. Algorithmic personality, identity and registration are suggested as a means of governance. A registration system will allow subsequent versions of an algorithm to be recognised as either the "same" or "new". In particular, the Report provides that this regulatory agency: • will maintain a register of uses of predictive algorithms within the government agencies. Those agencies will be required to conduct ongoing assessments of the use of those algorithms and submit reports to the regulator at regular intervals— either every year or three years as required by the regulator. • will produce an annual public report on the use of predictive algorithms within the government. This report will make public the uses of predictive algorithms in its register, including input and output variables for each algorithm, with exceptions made in cases where this knowledge would enable "gaming" of the algorithm. This information-keeping responsibility will support the Agency in completing its tasks, i.e., producing an annual public report on the uses of AI and conducting ongoing monitoring on the effects of AI tools.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: The underlying basis of this suggestion is law. A legal act should establish this regulatory agency charged with keeping a register with algorithms. Nature: This report provides advice and recommendations on the use of predictive algorithms to the public sector of New Zealand. It is not a legally binding document. Should such a regulatory agency be established, the exercise of this function, i.e., maintaining a register of algorithms, will be mandatory. Moreover, the agencies using algorithms will also need to inform the regulatory agency. Scope: This recommendation is context-specific and sector-specific. It is addressed at the public sector and, most specifically, governmental departments of New Zealand.

Option: Register of algorithms used in government Proposer: New Zealand Law Foundation and University of Otago Reference/link to relevant document: Government use of artificial intelligence in New Zealand.https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf Assessed by: TRI Date of assessment: 14/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
3. Purpose/objective/what need does the option fulfil?	The creation and retention of a register of algorithms by a regulatory agency aims to contribute to the regulation and governance of AI in New Zealand. It aims to document and monitor the uses of algorithmic for predictive policing by the public sector in New Zealand.
4. What gap does it address?	This recommendation is quite novel and has identified the lack of an obligation to keep a register of algorithms, by public authorities or not. There is not currently a standalone legal obligation for the European public (or private) sector to record the algorithms they use. Nonetheless, information about algorithms may be revealed under other legal requirements, such as the obligation of data controllers to provide information about the existence of automated decision-making, including profiling and meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject as provided for under Article 15 Regulation (EU) 2016/679 (where this Regulation applies).
5. What added value does it have?	In addition to identifying a novel and innovative means of regulating AI, maintaining a register of algorithms could be of salient importance for the following reasons: • To promote transparency and trust in public authorities • To build trust in the use of algorithms • To enhance accountability and enable the public sector to remain informed of the uses of AI within the various governmental departments. Although not specified, it is also reasonable to expect that such measures could relieve public authorities from the burden of public records requests and freedom of information requests regarding the use of algorithms. On the contrary, the proactive publication of this information could enhance public scrutiny and democratic governance. Moreover, releasing this information into the public space could further support research on the use, applications and consequences of algorithms. A similar suggestion was made by the Law Society of England and Wales (The Law Society of England and Wales, 2019). It is suggested that a register of algorithmic systems in criminal justice should be created, including those not using personal data, alongside standardised metadata concerning both their characteristics, such as transparency and discrimination audits and relevant standard operating procedures, and the datasets used to train and test them. This will also support secure access to algorithmic systems in use by or on behalf of public bodies in the criminal justice system for researcher and journalistic oversight.

Option: Register of algorithms used in government Proposer: New Zealand Law Foundation and University of Otago Reference/link to relevant document: Government use of artificial intelligence in New Zealand.https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf Assessed by: TRI Date of assessment: 14/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
6. What are the limitations, risks and challenges?	Limitations: The focus of this recommendation is limited to predictive algorithms and uses of AI by governmental departments in New Zealand. Therefore, it does not apply to commercial uses of AI and uses other than predictive profiling. Risks: Not identified in the proposal itself. There may be conflicts with intellectual property rights and requirements depending on the nature and types of information captured in this register. Challenges: Not identified in the proposal itself. However, this option requires the cooperation of governmental departments with the Agency and the disclosure of information about the use of algorithms, which may be sensitive and confidential for national purposes. Moreover, implementing this option may require the support of the private companies who created and tested the algorithms and may be unwilling to share confidential information about the algorithms.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	The examined option is clear, specific and able to be effectively and efficiently operationalised. This report does not provide detail on the implementation of this measure but its implementation should be outlined in the legal act establishing the responsible regulatory agency.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	This report suggests that the regulatory agency responsible for the register of algorithms should also conduct ongoing monitoring on the effects of this tool. Regarding the external powers and relationships of this agency, it is suggested that if a regulatory agency is to be given any sort of hard-edged powers, consideration will need to be given to its capacity to monitor and enforce compliance with these. The report clearly states that the preference would be a relatively "hard-edged" regulatory agency, with the authority to demand information and answers, and to deny permission for certain proposals. However, it is acknowledged that even a light-touch regulatory agency should perform some of the described roles. More detail and provisions are required regarding the monitoring and enforcement powers of this agency. For example, it is necessary to investigate the legal status of the decisions and actions of this regulatory agency, i.e., whether they are legally binding, enforceable *ipso facto* and subject to appeal. In addition, it is necessary to clarify whether the agency could act at its own initiative and whether it could have auditing and reporting powers. Finally, it is worth exploring whether what kind of enforcement powers and tools the agency should be given, such as the prohibition of the use of AI for specific applications and the imposition of fines.

Option: Register of algorithms used in government Proposer: New Zealand Law Foundation and University of Otago Reference/link to relevant document: Government use of artificial intelligence in New Zealand.https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf Assessed by: TRI Date of assessment: 14/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Regarding the creation of a register of algorithms, no implementation burdens are clarified. However, it is reasonable to expect that public administrations should invest time, resources and effort in completing the register of algorithms. There should be clear policies and templates and guidance on completing and updating this register. Regarding implementation burdens on citizens, this hasn't been elaborated. As far as businesses are concerned, this measure only relates to the algorithms used in the public sector. However, cooperation between public authorities and businesses may be required to ensure that this register includes accurate information. For example, companies may be asked to provide information about the algorithms they have created and tested.
10. Which stakeholders would benefit most from the use of this option?	This regulatory option could benefit individuals and society at large as explained above. Moreover, it is reasonable to expect that this option could benefit the public sector where trust in the use of algorithms by the government will be enhanced. Finally, this option could be a case-study for other actors involved in the use of algorithms and adopted by businesses as well.
11. Whose rights and/or interests does this option neglect?	This option is context-specific for the use of algorithms for predictive policing in the public sector. Therefore, it does not consider the interests at stake where AI is used for other purposes and by other stakeholders, such as private entities. Furthermore, it seems to neglect the interests and rights of public authorities and private companies if confidential and sensitive information is asked from them about the uses of algorithms. In terms of the examined rights, it is worth pointing out that the rights examined in this report reflect the legal status and provisions in New Zealand.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	It actively and explicitly supports human rights. It aims to safeguard and enhance human rights, such as the right to privacy and freedom from discrimination, in the context of predictive algorithms. Human rights of individuals, especially those that are profiled and assessed against the risk of crime offending and recidivism, are safeguarded and enhanced where data bias and inaccuracy are minimized. Moreover, enhancing transparency, accountability, and trust in the use of algorithms further supports the respect for human rights. In this context, keeping a register of algorithms enhances public scrutiny and legal challenge of the application of algorithms.

Option: Register of algorithms used in government Proposer: New Zealand Law Foundation and University of Otago Reference/link to relevant document: Government use of artificial intelligence in New Zealand.https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf Assessed by: TRI Date of assessment: 14/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
13. How does it address ethics and ethical principles? Which ones?	This report acknowledges the importance of the below-described principles and suggests that any systems of governance of AI, including the register of algorithms, rely on the below: • Equality and equity • Ethical Use of AI and Data • Fairness • Human autonomy • Informational privacy • Liability and personhood • Moral and legal responsibility • Optimal interaction between humans and machines • Political legitimacy • Prohibition of bias and discrimination • Promotion of certainty for citizens who are data subjects, government employees, and other stakeholders • Proportionality • Social license: public trust and confidence; • Timeliness • Transparency
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	Throughout the report, due consideration has been given to tackle unfair and inaccurate gender bias and discrimination in using predictive algorithms. Reference is also made to Section 21(1) of the Human Rights Act 1993 and the prohibition of discrimination based on sex, race, age and employment status. Regarding the composition of the agency, the focus has been on expertise and independence. Gender equality or neutrality is not specifically provided. However, this report suggests that cultural and diversity perspectives should be considered as well. For example, it is explained that some degree of input from those most likely to be adversely affected by algorithmic decisions would be vital, such as the representation of Maori tribes.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	This report was prepared as part of the New Zealand Law Foundation-funded project: *Artificial Intelligence and Law in New Zealand*. Regarding the establishment of a regulatory agency in charge of keeping a register of algorithms, there is no specific provision about the funding of this body in the report. Nonetheless, it is clarified that this regulatory agency should be independent in overseeing AI and performing its missions.
16. What provisions are there for regular review and update?	There are no specific provisions for the review and update of the obligation to maintain a register of algorithms. Nonetheless, the information-keeping nature of this measure suggests that such registers should be reviewed and updated to reflect the status and

Option: Register of algorithms used in government Proposer: New Zealand Law Foundation and University of Otago Reference/link to relevant document: Government use of artificial intelligence in New Zealand.https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf Assessed by: TRI Date of assessment: 14/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	categories of the algorithms applied. Moreover, this report proposes the regular review of algorithms and their uses. In addition, it is recommended that regulations should be flexible and adaptable in light of technological change and should be subject to appropriately regular review. In this context, although there is not a specific review clause for this measure, it is likely that review and update will be required for the registers of algorithms as well.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Given the bureaucratic and information-keeping character of this measure, this option is considered feasible, sustainable and future-proof if it remains under review and kept updated and it can draw adequate funding. It is expected that this measure will reflect and be supported by policy and market incentives -to the extent that this relates to the use of algorithms in the public sector- since the responsible regulatory agency is generally bound to consult and consider market and policy needs prior to making policy decisions.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	This measure is specifically aimed at the public sector. Therefore, the public sector is expected to be affected in the first place. Although not specified in this report, implementing the same obligation to businesses could stifle innovation, creativity and financial prosperity. If this information-keeping obligation requires businesses to reveal financially sensitive information about the applied algorithms, this may disempower the right to intellectual property.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	Not clarified. Incompatibility is unlikely, though, with the EU legal framework. A legal basis on an EU or national level will be required for the establishment of an overseeing European or national agency to maintain a register of algorithms.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	Not applicable. Please see above.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	4

Option: Register of algorithms used in government Proposer: New Zealand Law Foundation and University of Otago Reference/link to relevant document: Government use of artificial intelligence in New Zealand.https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf Assessed by: TRI Date of assessment: 14/11/2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Establishing an obligation to keep a register of algorithms requires the identification and establishment of the responsible agency for this, the extent of this obligation (e.g., private or public sector) and the specific materialisation of this obligation (e.g. what types of information should be recorded about the uses of algorithms). Moreover, due consideration should be given to ensure that consistent and fair standards apply across the public and private sector so that businesses are not unfairly advantaged or disadvantaged. Needless to say, that establishing a regulatory agency to maintain a register of algorithms requires policy impact assessment, public consultations, public debate, specific budget allocation and legislative amendments. Finally, the cooperation mechanisms with other public authorities and independence safeguards of such an agency should be considered to ensure the effectiveness of such measures.
References consulted	European Parliamentary Research Service, A governance framework for algorithmic accountability and transparency, April 2019. http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf Gavaghan, Colin, Alistair Knott, James Maclaurin, John Zerilli, Joy Liddicoat, Government use of artificial intelligence in New Zealand, Final Report on Phase 1 of the New Zealand Law Foundation's Artificial Intelligence and Law in New Zealand Project, 2019..https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf Law Society of England and Wales (The Law Society of England and Wales, *Algorithm use in the criminal justice system report*, 4 June 2019. https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report/).

4.25. Digital Authority

Option: Digital Authority Proposer: UK HOUSE OF LORDS Select Committee on Communications Reference/link to relevant document: https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/299.pdf Assessed by: TRI Date of assessment: 9 October 2019 Stakeholder(s) consulted in option assessment: Professor Andrew Murray, Commissioner, LSE Truth, Trust and Technology Commission.	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Give an application example	'Digital world'—an environment composed of digital services facilitated by the internet. Aims to '"bring a new consistency and urgency to regulation".
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: Murray highlights, "To create the digital authority primary legislation would be required. There would need to be sufficient authority for the DA to operate and there would need to be a funding model, therefore it is unlikely that anything other than primary legislation would achieve these aims." Nature and scope: As proposed, the Digital Authority would have the remit to continually assess regulation in the digital world and make recommendations on where additional powers are necessary to fill gaps. The Digital Authority would also bring together non- statutory organisations with duties in this area. It is expected to play a key role in providing the public, the Government and Parliament with the latest information. The Digital Authority would report to a joint committee of both Houses of Parliament whose remit is to consider all matters related to the digital world.
3. Purpose/objective/what need does the option fulfil?	The proposed functions of the body include: continually assessing regulation in the digital world and make recommendations on where additional powers are necessary to fill gaps; establishing an internal centre of expertise on digital trends which helps to scan the horizon for emerging risks and gaps in regulation; helping regulators to implement the law effectively and in the public interest, in line with the 10 principles set out in the House of Lords report; informing Parliament, the Government and public bodies of technological developments; providing a pool of expert investigators to be consulted by regulators for specific investigations; surveying the public to identify how their attitudes to technology change over time, and to ensure that the concerns of the public are taken into account by regulators and policy-makers; raising awareness of issues connected to the digital world among the public; engaging with the tech sector; ensuring that human rights and children's rights are upheld in the digital world; and liaising with European and international bodies responsible for internet regulation. It is envisaged the Digital Authority will have a coordinator role – it will instruct and coordinate regulators across different sectors and multiple Government departments.
4. What gap does it address?	Failures of self-regulation; out of date regulatory framework; regulatory fragmentation; gaps in regulation which do not clearly fall within any one

Option: Digital Authority Proposer: UK HOUSE OF LORDS Select Committee on Communications Reference/link to relevant document: https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/299.pdf Assessed by: TRI Date of assessment: 9 October 2019 Stakeholder(s) consulted in option assessment: Professor Andrew Murray, Commissioner, LSE Truth, Trust and Technology Commission.	
Criteria/touch point	**Assessment**
	regulator's remit, or which would require a regulator's remit to be expanded; poor policy and practice, such as inappropriate prosecutions and ineffective legislation; inadequacy of response by policymakers to changes in the digital world.
5. What added value does it have?	It is expected to help regulators to implement the law effectively and in the public interest and *"bring a new consistency and urgency to regulation".* (See the UK HL report). It could help eliminate overlaps (see https://www.theregister.co.uk/2019/03/09/lords_communications_committee_internet_regulation/)
6. What are the limitations, risks and challenges?	One limiting factor could be the expense involved in setting up a new body. Risks: overregulation of the digital world as it would have significant powers. Another risk might be if mission creep takes place – i.e., there is a greater transfer of powers than originally intended. Challenges: proper and sustained funding and resources, ability to remain impartial and independent of the Government, resistance from existing regulators, support from policymakers.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	Yes to some extent; but a detailed analysis is missing especially in terms and the report is not always clear on this. I.e., how the Digital Authority would actually coordinate and instruct regulators and government departments under its remit. While the report outlines that, "Its board should consist of chief executives of relevant regulators with independent non-executives", it has not supplemented this in the instruction remit and not fleshed it out. Further, in terms of the government and Parliament, the Digital Authority would report to a Cabinet Office minister – this person would be expected to "champion" the DA in Cabinet. The Digital Authority is also expected report to Parliament on a quarterly basis and regularly give evidence to the new joint committee to discuss the adequacy of powers and resources in regulating the digital world – the latter relationship needs to be further specified**.**
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Not outlined/defined in current proposal. Murray outlines that "the Digital Authority is not a regulator and does not have independent monitoring, oversight and enforcement mechanisms. It is to horizon scan, co-ordinate and propose policy. The regulation aspect will still be done by current regulators." In terms of its own oversight, the House of Lords report recommended the Digital Authority should report to the Cabinet Office and be overseen at the highest level.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens	Not evident from first reading. One news report quotes House of Lords Committee chairman Lord Stephen Gilbert as saying, "its cost would be relatively small and that this could be stumped up either from the individual regulators, through cash or in kind, or by the

Option: Digital Authority Proposer: UK HOUSE OF LORDS Select Committee on Communications Reference/link to relevant document: https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/299.pdf Assessed by: TRI Date of assessment: 9 October 2019 Stakeholder(s) consulted in option assessment: Professor Andrew Murray, Commissioner, LSE Truth, Trust and Technology Commission.	
Criteria/touch point	**Assessment**
b. Public administrations c. Businesses and particularly SMEs?	government." https://www.theregister.co.uk/2019/03/09/lords_communications_committee_internet_regulation/
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Regulators, Parliament, the Government and public bodies
11. Whose rights and/or interests does this option neglect?	Not evident.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	One of its key proposed functions includes ensuring that human rights and children's rights are upheld in the digital world. The Digital Authority is expected to help regulators to implement the law effectively and in the public interest, in line with the 10 principles I.e., parity, accountability, transparency, openness, ethical design, privacy, recognition of childhood, respect for human rights and equality rights, education and awareness-raising and democratic accountability, proportionality and evidence-based approach. Also one of its envisaged functions is: to ensure that human rights and children's rights are upheld in the digital world. At this stage, how it might do this is not very clear.
13. How does it address ethics and ethical principles? Which ones?	See above.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No. Given most of the board would be CEOs, Commissioners or similar from the regulators most would be ex office. The question of the independent chair and the non-execs remains unresolved.

Option: Digital Authority Proposer: UK HOUSE OF LORDS Select Committee on Communications Reference/link to relevant document: https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/299.pdf Assessed by: TRI Date of assessment: 9 October 2019 Stakeholder(s) consulted in option assessment: Professor Andrew Murray, Commissioner, LSE Truth, Trust and Technology Commission.	
Criteria/touch point	**Assessment**
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	The HL report outlines: The Digital Authority must be properly funded to be effective and to carry out research. No further details are provided.
16. What provisions are there for regular review and update?	Not defined. Murray explains that "Although maybe not 100% clear the quarterly reports to the new Joint Committee would be a two way process (referring to the report point 244: "the Digital Authority should report to Parliament on a quarterly basis and regularly give evidence to the new joint committee to discuss the adequacy of powers and resources in regulating the digital world") with the Joint Committee reviewing the DA as the DA advised the Joint Committee.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Its sustainability will depend on the policy and funding model adopted and its usefulness in regulating the digital world.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Yes, if it becomes over-prescriptive. But not as currently proposed.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	Not applicable. Note, one of the proposed functions of the Digital Authority is to liaise with European and international bodies responsible for internet regulation.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	Existing regulators might see some changes to their remits (elimination of overlaps, sharing of powers, budgets).

Option: Digital Authority Proposer: UK HOUSE OF LORDS Select Committee on Communications Reference/link to relevant document: https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/299.pdf Assessed by: TRI Date of assessment: 9 October 2019 Stakeholder(s) consulted in option assessment: Professor Andrew Murray, Commissioner, LSE Truth, Trust and Technology Commission.	
Criteria/touch point	**Assessment**
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	4
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	This looks like a promising option at the national level especially given two main concerns: regulatory fragmentation and gaps in knowledge. Its effectiveness will depend on proper funding, ability to coordinate and instruct different regulators, ability to remain politically impartial and independent of the Government; democratic scrutiny. Though the HL report envisages the Digital Authority to co-ordinate existing regulators and not replace them or change their remit, but concerns have been expressed, e.g., about whether in proposing this new body: (a) what would be the future role of existing bodies such as the Advertising Standards Authority and (b) whether policy-makers understand and appreciate the complexity of the online advertising ecosystem and use an evidence-based approach. See: https://www.marketingweek.com/government-regulation-digital/
References consulted	House of Lords, Select Committee on Communications 2nd Report of Session 2017–19, *Regulating in a digital world,* 9 March 2019. **https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/299.pdf**

4.26. Independent cross-sector advisory body (Centre for Data Ethics and Innovation)

Option: National Independent cross-sector advisory body (Centre for Data Ethics and Innovation) Proposer: UK government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813933/Intro_to_CDEI.pdf Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment:-.	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and	The CDEI is an independent advisory body and part of the UK Department for Digital, Culture, Media & Sport, set up and tasked by the UK Government to investigate and advise on how to maximise the benefits of data-driven technologies. In this context, the CDEI explores the legal, ethical and societal tensions in data-driven technologies.

Option: National Independent cross-sector advisory body (Centre for Data Ethics and Innovation) Proposer: UK government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813933/Intro_to_CDEI.pdf Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment:-.	
Criteria/touch point	**Assessment**
big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	The CDEI analyses and anticipates the opportunities and risks posed by data-driven technology and puts forward practical and evidence-based advice to address them. In particular, the CDEI: • carries out thematic projects to enable the CDEI to explore live or urgent issues; • reviews, identifies and articulates best practice for the responsible use of data-driven technology within specific sectors or for specific applications of technology; • publishes reports with clear recommendations to government. The work of CDEI will also touch upon issues regarding the responsible use of AI and (Big) data. According to its Work Programme for 2019/20, its focal point includes data practices in online targeting and algorithmic bias. During the year 2019/2020, the CDEI has already produced work relevant to AI. More specifically, it: • has published briefing papers on issues of public concern in AI ethics including deepfakes, AI and insurance and smart speakers (available at https://www.gov.uk/government/publications/cdei-publishes-its-first-series-of-three-snapshot-papers-ethical-issues-in-ai); • commissioned the Royal United Services Institute to publish research into the use of algorithms in policing, and the potential for bias (available at https://www.gov.uk/government/publications/report-commissioned-by-cdei-calls-for-measures-to-address-bias-in-police-use-of-data-analytics) • published interim reports on online targeting and bias in algorithmic decision-making (available at https://www.gov.uk/government/publications/interim-reports-from-the-centre-for-data-ethics-and-innovation) and • partnered with Royal United Services Institute (RUSI) to carry out research into the potential for algorithmic bias in policing and how to ensure adequate oversight of these technologies (available at https://www.gov.uk/government/news/cdei-and-the-royal-united-services-institute-convene-round-tables-to-discuss-the-use-of-algorithms-in-policing). Where the CDEI focuses on specific fields, such as AI in insurance, it provides evidence and solutions tailored to this particular sector. For example, increased transparency, legal measures against hyper-personalised risk assessments and behavioural nudging (e.g., prohibition of processing specific categories of personal data and anti-discrimination audits), security and organisational measures (e.g., clear privacy policies and review of the agreements with third-party data processors) are some of the bespoke controls recommended by the CDEI.
2. What is its basis (on which the regulatory option is created - law? if yes	Basis: The CDEI was set up by the UK government and more specifically by the Department for Digital, Culture, Media and Sport as the body specifically tasked with making recommendations to government to 'maximise the benefits of data and AI for our society and economy'. The CDEI is not currently relying on a statutory

Option: National Independent cross-sector advisory body (Centre for Data Ethics and Innovation) Proposer: UK government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813933/Intro_to_CDEI.pdf Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment:-.	
Criteria/touch point	**Assessment**
which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	footing. The CDEI is not a separate legal entity and operates as an Expert Committee, working independently of the government. During this period, the CDEI is in its initial, pre-statutory phase of activity, where it will also assess where its functions may need to be amended or augmented with specific powers when the Centre is established on a statutory footing, as explained in its Terms of Reference (Centre for Data Ethics and Innovation, Terms of Reference, June 2018 available at https://www.gov.uk/government/publications/the-centre-for-data-ethics-and-innovation-cdei-terms-of-reference/centre-for-data-ethics-and-innovation-cdei-terms-of-reference). Nature: During its pre-statutory phase, the CDEI is not a separate legal entity and operates as an Expert Committee, working independently of the government. The government has committed to putting CDEI on an independent statutory footing. Full governance arrangements are set out in a Framework Agreement between CDEI and DCMS. The nature and type of the work produced by the CDEI vary and its suggestions tend to cover a wide range of sources and regulatory options. Legal and technological improvements are suggested alongside more generic suggestions, such as public debate and awareness and policy initiatives for the industry. The publications of the CDEI do not have the legal standing of a legally binding document. However, the CDEI enjoys safeguards of independence and the UK government is bound to consider and respond publicly to the CDEI recommendations. Regarding the scope of the CDEI, the CDEI embraces a universal and global scope given the nature of the topics it touches upon. Although the direct focus of the CDEI is the UK society and *status quo*, the CDEI and its findings are of relevance to European and national authorities, civil organisations, policymakers and the industry. Moreover, the scope and field of its work is rather broad and open-ended and covers all the aspects of data-driven technologies, including legal, technical, operational and social, sector-specific issues, such as insurance and online targeting. Even where it examines specific technologies or areas and provides context-specific recommendations, its advice is of high value and importance to all the applications of AI and Big Data.
3. Purpose/objective/what need does the option fulfil?	The CDEI is tasked by the UK Government to connect policymakers, industry, civil society, and the public to develop the right governance regime for data-driven technologies. In this context, specific issues regarding big data analytics tools, machine learning, profiling, online targeting and algorithmic bias are examined. In addition to this, the CDEI focuses on specific fields and risk areas in these fields, including AI and insurance, AI and smart speakers, deepfakes, use of algorithms in policing, online targeting and bias in algorithmic decision-making, and algorithmic bias in policing.

Option: National Independent cross-sector advisory body (Centre for Data Ethics and Innovation) Proposer: UK government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813933/Intro_to_CDEI.pdf Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment:-.	
Criteria/touch point	**Assessment**
	The rationale and objective behind all these outputs are to investigate and advise on how to maximise the benefits of these technologies. The examined options aim to shed light on new data-driven technologies and applications or long-standing practices and provide recommendations for best governance and innovation.
4. What gap does it address?	Overall, the CDEI considers legal, technical, operational and ethical gaps in governance of data-driven technologies and provides recommendations to the government, as well as advice to regulators, creators and users of data-driven technology as to how those gaps should be addressed.
5. What added value does it have?	The CDEI has an independent Chair and Board, and its independence extends to how it will work, its use of resources and ultimately its recommendations to the UK government. The CDEI provides independent, impartial and expert advice. It aims to work openly and transparently with stakeholders, government and the public. The CDEI acts as a monitoring, alert and responsive mechanism to the market trends and technology applications and suggests areas for improvements and potential opportunities in the area of data-driven technologies. The suggestions of the CDEI are of great importance because they address legal, ethical, technical and organisational gaps in data-driven technologies and provide tailored and objective recommendations with the aim of maximising innovation and minimising risks.
6. What are the limitations, risks and challenges?	The recommendations of the examined papers of the CDEI (listed above) touch upon several aspects of data-driven technologies and provide cross-sector and multi-layered recommendations. Nonetheless, constraints and limitations regarding its nature and work may have an impact on the appropriateness and effectiveness of the suggested measures. As a publicly-funded boded, the operation of the CDEI may be subject to budget constraints, whereas this further challenges the principle of independence. In addition, as long as it does not have a statutory footing, there is a lack of clarity on its operation, powers, relationships with other regulatory authorities. Moreover, the work of the CDEI is closer to soft law rather than bringing legislative amendments or binding the public or private sector with enforceable recommendations. The work of the CDEI is also focusing on the UK sector and suggestions and, although it touches upon generic principles, its value may be limited to the UK legal order. Finally, the work of the CDEI tends to focus on specific areas, sectors or activities. Therefore, the value and importance of its recommendations may be limited to the specific context instead of raising general issues regarding AI and producing guiding principles and rules. The main **risk** is that the UK Government as the first recipient of the CDEI recommendations ignores or rejects them without carefully considering them. Moreover, it is also likely that an "indirect chilling effect" may occur. In particular, the recommendations cover most aspects of policy, legal, technical and

Option: National Independent cross-sector advisory body (Centre for Data Ethics and Innovation) Proposer: UK government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813933/Intro_to_CDEI.pdf Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment:-.	
Criteria/touch point	**Assessment**
	organisational measures. There is a risk that no actor will undertake action hoping or expecting that the other involved actors will act following these recommendations. Finally, as explained below, the **challenge** in implementing the suggested regulatory options is allocating resources, time and effort in further examining its appropriateness and efficiency and designing its specific implementation.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	The CDEI is already operational for the last 2 years. It carries an important mandate as the independent UK advisory body to investigate and advise on how the UK could maximise the benefits of AI and data-driven technology. In particular, the CDEI is funded by the Department for Digital, Culture, Media and Sport. During its pre-statutory phase, the CDEI is not a separate legal entity and operates as an Expert Committee, working independently of the government. In this context, the CDEI has developed various corporate functions, including strategy, governance, public engagement, and business support, to underpin its wider work. Its corporate functions include budgeting, hiring and team management, providing a secretariat to the Board, developing its strategy and monitoring progress, managing the Centre's relationship with government, ensuring effective collaboration and stakeholder management, and identifying, assessing and advising on the future form of the Centre including consideration of statutory functions and powers. It is led by an independent board comprising expert and influential individuals from a range of fields relevant to its mandate. The Board has oversight of — and is accountable for — the CDEI's work and recommendations. Moreover, the CDEI may rely on secondees, loans and expert advisors to support review and research work. Given the fast-changing environment, where the CDEI's work programme shifts year to year, this approach allows flexibility in ensuring the CDEI has the right skills and capabilities.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Internal oversight and monitoring lies with the Board, which has oversight of — and is accountable for — the CDEI's work and recommendations. On the contrary, as far as external monitoring, oversight and enforcement are concerned, the CDEI has no such powers. The CDEI may be vested with such powers at a later stage. During its initial phase, the CDEI will identify what additional functions the Centre may need to undertake to deliver its mandate effectively and assess where these functions may need to be amended or augmented with specific powers when the Centre is established on a statutory footing.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for:	The burdens associated with the implementation of the CDEI relate to the need to invest time, resources and effort in examining and tailoring the suggestions of the CDEI. Most suggestions of the CDEI relate to practical measures and controls and their appropriateness and efficiency should be examined by the relevant actors on an *ad hoc* basis. For example, policy committees should be established to consider legislative amendments and the drafting of codes of practice.

Option: National Independent cross-sector advisory body (Centre for Data Ethics and Innovation) Proposer: UK government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813933/Intro_to_CDEI.pdf Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment:-.	
Criteria/touch point	**Assessment**
a. Citizens b. Public administrations c. Businesses and particularly SMEs?	As far as **citizens** are concerned, the CDEI actively calls the public to engage with initiatives on AI and Big Data. For example, the CDEI argues in favour of citizen juries and other public engagement exercises. This would require the active participation of the public and panels raising of awareness, organising workshops, and disseminating the findings. The above also applies to SMEs, which may struggle to find the resources to implement the suggested measures. However, some measures are generally applicable and easy to implement, including clear policies and compliance with data protection law.
10. Which stakeholders would benefit most from the use of this option?	The recommendations and work of the CDEI are of interest to all - private and public stakeholders, industry and policy actors and individuals. Its work is based on interdisciplinary, independent and objective work and assessment. This constitutes a safeguard for a well-balanced and objective stance towards private entities, public authorities and individuals. Therefore, although the direct audience of the reports of the CDEI seems to be the industry actors engaging in specific data-driven technologies, private and public stakeholders, industry and policy actors and individuals could benefit from the CDEI work.
11. Whose rights and/or interests does this option neglect?	N/A
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	The CDEI explicitly, directly and actively supports the respect and promotion of human rights in the digital era. The CDEI takes a holistic approach to assessing the impact of data-driven technologies on human rights and considers all the applicable rights in the specifically examined context. It provides advice on the various stakeholders to ensure that human rights are not violated and recommendations on how to best respect and safeguard them in certain risky activities.
13. How does it address ethics and ethical principles? Which ones?	The work of CDEI takes into account several ethical tenets and ethics in a broad manner. The CDEI urges that ethical requirements are taken into account and addressed throughout all stages of the product lifecycle, from project inception through to development and application. The main ethical principles and imperatives discussed in its work include: • Access to services and products without bias or discrimination • Algorithmic fairness • Autonomy • Democratic governance • Diversity • Ethical innovation • Ethical use of AI

Option: National Independent cross-sector advisory body (Centre for Data Ethics and Innovation) Proposer: UK government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813933/Intro_to_CDEI.pdf Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment:-.	
Criteria/touch point	**Assessment**
	• Fairness and transparency offline and online • Human dignity • Human life • Human safety • Protection against harmful discrimination • Public engagement • Respect for property • Risk of bias and discrimination • Rule of law • Social cohesion • Sustainable growth • Sustainable, transparent beneficial, fair AI • The moral legitimacy of technologies before their deployment • The protection of vulnerable people • Trust in information
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	In terms of the composition of the CDEI itself, the CDEI is part of the UK Department for Digital, Culture, Media & Sport. The latter values equality and diversity in employment and is committed to being an organisation in which fairness and equality of opportunity is central to the approach in business and working relationships and where the organisational culture reflects and supports these values. This includes the right to a working environment free from discrimination, harassment, bullying and victimisation regardless of race, ethnic or national origin, age, religion, sex, gender identity, marital status, disability, sexual orientation, working hours, trade union membership or trade union activity. Regarding the CDEI suggestions, the CDEI takes into account the risk of gender bias and discrimination where relevant.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Yes, The CDEI is funded by the Department for Digital, Culture, Media and Sport with £2.5 million in 2019/20 and £5 million in 2020/21. Where the suggestions of the CDEI relate to monitoring and oversight mechanisms and policy initiatives, there is not a clear and detailed funding plan.
16. What provisions are there for regular review and update?	The CDEI will develop metrics to track the full range of activities as set out in its Terms of Reference and mechanisms to evaluate its progress. This will help understand both the extent to which its activities are contributing to its long-term goal and the statutory powers and funding the CDEI may require to deliver its Terms of Reference in the long-term. The CDEI will track progress through its pre-statutory phase and monitor:

Option: National Independent cross-sector advisory body (Centre for Data Ethics and Innovation) Proposer: UK government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813933/Intro_to_CDEI.pdf Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment:-.	
Criteria/touch point	**Assessment**
	• the extent to which government, industry and regulators adopt its recommendations and advice, and the extent to which they change their behaviour • the extent to which key stakeholders believe that it is having an impact on the issues it has been set up to address The CDEI will monitor these internally and publish its own assessment of work in its first annual report in Spring 2020.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	The CDEI is funded by the UK Department for Digital, Culture, Media and Sport. During its pre-statutory phase, the CDEI is not a separate legal entity and operates as an Expert Committee, working independently of the government. There are embedded mechanisms to ensure that the CDEI remains sustainable and operational and that its work remains relevant. In particular, the CDEI will develop metrics to track the full range of activities as set out in its Terms of Reference and mechanisms to evaluate its progress. This will help understand both the extent to which its activities are contributing to its long-term goal and the statutory powers and funding the CDEI may require to deliver its Terms of Reference in the long-term. The CDEI will monitor these internally and publish its own assessment of work in its first annual report in Spring 2020.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	It is not expected that the CDEI or the suggestions of the CDEI will adversely impact the ability for businesses and others to innovate given that the suggested measures aim to act as a counterbalance to the risks and harms. The measures are presented as enablers of technology and innovation and safeguards against risks and harms.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The CDEI does not contravene or clash with the EU legal framework. Such a national body might be the need of the hour in terms of promoting the goals of EU legislation and the CDEI could provide a model for other countries. Some of its suggestions are generic and generally applicable to all legal orders and markets. Others are addressed specifically at the UK Government and could potentially act as reference points for consideration from other European authorities and bodies.

Option: National Independent cross-sector advisory body (Centre for Data Ethics and Innovation) Proposer: UK government Reference/link to relevant document: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/813933/Intro_to_CDEI.pdf Assessed by: TRI Date of assessment: 12/11/2019 Stakeholder(s) consulted in option assessment:-.	
Criteria/touch point	**Assessment**
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	A major complexity relates to the suggestion for coordinated action and reference to other legal fields and practices as best examples or areas for avoidance. Although this is an evidence-based and thorough approach, coordinating and assessing policy changes may be complicated, requiring advanced planning and policy assessments.
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	4
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Although the recommendations of the CDEI cover a wide range of regulatory options, the challenge in implementing these options relate to two main factors. First, where legislative amendments, namely providing for data protection constraints in AI-related uses, or policy initiatives, such as codes of practice, a full policy impact assessment is required prior to the adoption of these measures. Second, given the global scope, field and reach of AI-driven applications and its far-reaching implications, due consideration should be given to ensure that the suggested regulatory options are aligned with the policy and legal measures of other markets and legal orders. Otherwise, and despite the best intentions of the CDEI, a bespoke system may not be that effective if it is not part of a uniform regulatory approach to AI on a supra-national level.
References consulted	Indicated above. Also: Centre for Data Ethics and Innovation, Centre for Data Ethics and Innovation 2-year strategy, March 2019 available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/787736/CDEI_2_Year_Strategy.pdf

4.27. FDA for algorithms

Option: An FDA for algorithms Proposer: **Andrew Tutt** Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994 Assessed by: TRI Date of assessment: 13 November 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	The proposal is for a new specialist federal-level regulatory agency to be created to regulate algorithmic safety with the following powers (1) to organize and classify algorithms into regulatory categories by their design, complexity, and potential for harm (in both ordinary use and through misuse). (2) to prevent the introduction of algorithms into the market until their safety and efficacy has been proven through evidence-based pre-market trials. (3) broad authority to impose disclosure requirements and usage restrictions to prevent algorithms' harmful misuse.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: federal laws Nature: new specialist centralised federal regulatory agency modelled on the US Food and Drug Administration (FDA) able to "engage in ex ante regulation rather than relying on ex post judicial enforcement," "with a broad mandate to ensure that unacceptably dangerous algorithms are not released onto the market, rather than charged with the enforcement of piecemeal legislation and with "ultimate authority over algorithmic safety regardless of the type or kinds of products in which those algorithms are embedded". (Tutt 2017) Scope: Tutt indicates it "could act as a standards-setting body that coordinates and develops classifications, design standards, and best practices". It "could also nudge algorithm designers through soft- touch regulations. That is, it could impose regulations that are low enough cost that they preserve freedom of choice and do not substantively limit the kinds of algorithms that can be developed or when or how they can be released." Further, the agency "could act as a hard-edged regulator that imposes substantive restrictions on the use of certain kinds of machine-learning algorithms, or even with sufficiently complex and mission-critical algorithms, act as a regulator that requires pre-market approval before algorithms can be deployed." (Tutt 2017)
3. Purpose/objective/what need does the option fulfil?	To ensure that algorithms are safe and effective.
4. What gap does it address?	Gaps in remedies offered by tort and civil law. Tutt outlines, "For consumers, tort and criminal law are unlikely to efficiently counter the harms from algorithms. Harms traceable to algorithms may frequently be diffuse and difficult to detect. Human responsibility and liability for such harms will be difficult to establish. And narrowly tailored usage restrictions may be difficult to enforce through indirect regulation. For innovators, the availability of federal pre-emption from local and ex-post liability is likely to be desired." (Tutt 2017)

Option: An FDA for algorithms Proposer: **Andrew Tutt** Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994 Assessed by: TRI Date of assessment: 13 November 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
5. What added value does it have?	Tutt outlines that "A single highly-motivated regulator could develop comprehensive policy, could quickly respond to new products and practices, and could also ensure that consumers are adequately protected." Further Tutt outlines that "a new federal agency in this space could add significant value—in the form of centralized expertise—even if other agencies retained primary jurisdiction over specific technologies." (Tutt 2017)
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*	Limitations: resource constraints; too soft or too tough a mandate. Risks: Per Tutt, "those who favor free markets may think a federal regulatory agency is too radical and more than is necessary at this early stage." (Tutt 2017) Challenges: these include determining what is excessive and/or insufficient regulation and excessive regulatory authority, addressing any internal knowledge gaps.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	Tutt has outlined the powers of such a body (see above Q1). Tutt has discussed how it could act as a standards setting body (covering aspects such as classification, performance standards, design standards, and liability standards), how it could act as a soft-touch regulator or hard-edged regulator. Tutt has also looked at other regulatory options and their inadequacy (i.e., state regulation, federal regulation by other subject-matter agencies, and presents the case for a central federal agency covering its complexity, opacity complexity, dangerousness and how it might work.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Tutt inter alia suggests the agency "could act as a hard-edged regulator that imposes substantive restrictions on the use of certain kinds of machine- learning algorithms, or even with sufficiently complex and mission- critical algorithms, act as a regulator that requires pre-market approval before algorithms can be deployed. That pre-market approval process could provide an opportunity for the agency to require that companies substantiate the safety performance of their algorithms. The agency could work with an applicant to develop studies that would prove to the agency's satisfaction that the algorithm meets that performance standard. Algorithms could also be conditionally approved subject to usage restrictions—for example, a self-driving car algorithm for cruise control could be approved subject to the condition that it is only approved for highway use. Off-label use of an algorithm, or marketing an unapproved algorithm, could then be subject to legal sanctions" (Tutt 2017).

Option: An FDA for algorithms Proposer: **Andrew Tutt** Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994 Assessed by: TRI Date of assessment: 13 November 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Citizens: none Public administrations: cost and implementation burdens connected with establishing a new agency. Businesses and particularly SMEs: regulatory compliance burdens.
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Open source community, commercial firms, to customers, to potential victims.
11. Whose rights and/or interests does this option neglect?	-
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	It does not discuss human rights.
13. How does it address ethics and ethical principles? Which ones?	Transparency is covered (in context of algorithmic disclosures) Accountability.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	Not elaborated.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not elaborated.

Option: An FDA for algorithms Proposer: **Andrew Tutt** Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994 Assessed by: TRI Date of assessment: 13 November 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
16. What provisions are there for regular review and update?	Not elaborated.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	The US FDA is part-funded by federal budget authorization and the other part is paid for by industry user fees. Sustainability of the proposed (new) FDA for algorithms will have to be similarly ensured and guaranteed. It is susceptible to policy changes (e.g., deregulation) and the restriction of its powers by changes to policy/legislation.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Potentially, it might be an obstacle. Tutt recognises that "there are legitimate concerns that regulation stifles innovation and impedes competition. Those who favor free markets may think a federal regulatory agency is too radical and more than is necessary at this early stage." (Tutt 2017)
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	Not applicable – this is could be a model for a national regulator.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	-
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	3
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Groth, Nitzberg and Russell outline that "An oversight body would need to carry the authority of a government agency like the FDA, but also employ the depth of technical know-how found at existing technology-focused governing bodies like ICANN. It would need to house a rich diversity of expertise to grasp the breadth of society, seating psychologists and sociologists alongside programmers and economists. Because not every piece of code needs tight oversight, it would need distinct trigger points on when to review and at what level of scrutiny, similar to the ways the FDA's powers stretch or recede for

Option: An FDA for algorithms Proposer: **Andrew Tutt** Reference/link to relevant document: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994 Assessed by: TRI Date of assessment: 13 November 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	pharmaceuticals versus nutritional supplements." (Groth, Nitzberg and Russell 2019). These should be taken into account. The US FDA has developed a new regulatory framework specifically tailored to promote the development of safe and effective medical devices that use advanced artificial intelligence algorithms, so might it be possible to extend the agency's scope?
References consulted	Food and Drugs Administration, Statement from FDA Commissioner Scott Gottlieb, M.D. on steps toward a new, tailored review framework for artificial intelligence-based medical devices, 2 April 2019. https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-steps-toward-new-tailored-review-framework-artificial Groth, Olaf J., Mark J. Nitzberg, Stuart J. Russell, "AI Algorithms Need FDA-Style Drug Trials" *WIRED opinion*, 15 August 2019. **https://www.wired.com/story/ai-algorithms-need-drug-trials/#** Tutt, Andrew, An FDA for Algorithms (March 15, 2016). 69 Admin. L. Rev. 83 (2017). Available at SSRN: https://ssrn.com/abstract=2747994 or http://dx.doi.org/10.2139/ssrn.2747994

4.28. US Federal Trade Commission to regulate robotics

Option: Proposal for US Federal Trade Commission to regulate robotics Proposer: Various, Woodrow Hartzog Reference/link to relevant document: https://digitalcommons.law.umaryland.edu/mlr/vol74/iss4/4/ Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	Currently robotic devices in the US are subject to a patchwork of federal regulation, depending on their function. For example, autonomous aircraft are regulated by the Federal Aviation Administration, AI-based diagnostic systems are regulated by the Food and Drug Administration, autonomous land-based vehicles are under review by the National Highway Traffic and Safety Administration, and any robotic toys would fall under the jurisdiction of the Consumer Product Safety Commission. This assessment reviews a recommendation[1] by Prof. Woodrow Hartzog for the Federal Trade Commission ("FTC") to be given primary

Option: Proposal for US Federal Trade Commission to regulate robotics Proposer: Various, Woodrow Hartzog Reference/link to relevant document: https://digitalcommons.law.umaryland.edu/mlr/vol74/iss4/4/ Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
	responsibility for overseeing regulation of autonomous systems, under its jurisdiction to regulate unfair and deceptive trade practices. Note: This recommendation answers *who* should regulate autonomous systems, not *how* they should be regulated.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: This proposal is for the consolidation of regulation of autonomous systems from multiple federal agencies to the FTC, based on its existing mandate to regulate unfair and deceptive trade practices. Nature: Binding Scope: National (autonomous systems manufactured, offered, sold or used in the US)
3. Purpose/objective/what need does the option fulfil?	Consolidation of federal regulation of autonomous systems to one primary agency (the FTC)
4. What gap does it address?	Autonomous systems are currently regulated (inconsistently) by multiple federal agencies based on their function.
5. What added value does it have?	This approach could build a rich cross-industry knowledge base and experience base for regulation of a wide spectrum of autonomous systems, avoiding knowledge "silos".
6. What are the limitations, risks and challenges?	Autonomous systems are not used in a vacuum. Regulation of autonomous systems in highly specialised environments (such as medical uses) or in environments presenting risk of injury or death to bystanders (drones and autonomous vehicles) may require specialized knowledge that is already in place in other agencies. The FTC's jurisdiction does not extend to federally regulated financial institutions, common carriers, or non-profit organisations. The FTC does not have the power to approve or certify medical devices, passenger vehicles or aircraft.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	Implementation would require: • legislation • clarification of the respective responsibility of various agencies for areas currently being regulated by other agencies (e.g. drones, medical devices and autonomous vehicles), and • coordination of efforts among agencies for areas outside the scope of the FTC.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	The FTC is currently subject to administrative oversight by the Office of the Inspector General, as well as legislative and judicial oversight by Congress and the federal court system.

Option: Proposal for US Federal Trade Commission to regulate robotics Proposer: Various, Woodrow Hartzog Reference/link to relevant document: https://digitalcommons.law.umaryland.edu/mlr/vol74/iss4/4/ Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Not applicable; no specific new regulations were proposed.
10. Which stakeholders would benefit most from the use of this option?	Unclear. Implementation could provide clarity for developers, though some developers who currently know how to work with other agencies would have to learn how to work with the FTC. It's unclear whether the proposal would result in different rights for consumers than are currently provided through the patchwork of agencies.
11. Whose rights and/or interests does this option neglect?	None identified
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	No. The FTC's jurisdiction is to regulate unfair and deceptive trade practices, not necessarily to support or advocate for human rights.
13. How does it address ethics and ethical principles? Which ones?	The FTC's focus is on fairness and avoiding deception.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	The FTC is already a federally funded agency. If the FTC takes over work that had been performed by other agencies, a reallocation of budget resources would be needed.
16. What provisions are there for regular review and update?	None identified.

Option: Proposal for US Federal Trade Commission to regulate robotics Proposer: Various, Woodrow Hartzog Reference/link to relevant document: https://digitalcommons.law.umaryland.edu/mlr/vol74/iss4/4/ Assessed by: UCLANCY, date assessed 17 Nov 2019 Stakeholder(s) consulted in option assessment:	
Criteria/touch point	**Assessment**
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	sible: Yes (the FTC already regulates a wide variety of businesses) tainable: Yes, supported by policy ure-proof: Possibly. Since its establishment in 1914, the FTC has shown the ability to adapt its regulatory approach to new technologies and new issues.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Insufficient information. New regulations could hinder innovation, but consolidation of regulation into a single lead agency could provide clarity that facilitates innovation.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The European Commission has the ability to regulate unfair trading practices (such as Directive (EU) 2019/633 of the European Parliament and of the Council of 17 April 2019 on unfair trading practices in business-to-business relationships in the agricultural and food supply chain (link)). However, most unfair trade practice legislation is at the Member State level.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	3 (neutral) (see #22 below)
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	This is primarily a political choice and will require political will to be adopted and implemented.
References consulted	1. Hartzog, Woodrow, "Unfair and Deceptive Robots", *Maryland Law Review*, Vol. 74, No. 4, 2015, pp. 785-829. https://digitalcommons.law.umaryland.edu/mlr/vol74/iss4/4/

4.29. Using anti-trust regulations to break up big tech and appoint regulators

Option: Using anti-trust regulations to break up big tech and appoint regulators to reverse illegal and anti-competitive tech mergers Proposer: Elizabeth Warren, US Senator. Reference/link to relevant document: https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c Assessed by: TRI with inputs from UCLANCY Date of assessment: 7 Nov 19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	Warren's proposes two actions 1. Passing legislation that requires large tech platforms to be designated as "Platform Utilities" and broken apart from any participant on that platform. 2. Appointing regulators committed to reversing illegal and anti-competitive tech mergers.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: Current antitrust laws and appointment of regulators. Nature: legislation supported by enforcement Scope: International with focus on large tech platforms to be designated as "Platform Utilities".
3. Purpose/objective/what need does the option fulfil?	Warren's proposal aims to "restore the balance of power in our democracy, to promote competition, and to ensure that the next generation of technology innovation is as vibrant as the last". It seeks to "promote healthy competition in the market — which will put pressure on big tech companies to be more responsive to user concerns, including about privacy." (Warren 2019)
4. What gap does it address?	Warren states, "big tech companies have too much power — too much power over our economy, our society, and our democracy. They've bulldozed competition, used our private information for profit, and tilted the playing field against everyone else. And in the process, they have hurt small businesses and stifled innovation." This is what is sought to be addressed.
5. What added value does it have?	Warren states "it allows us to make some progress on each of these important issues too. More competition means more options for consumers and content creators, and more pressure on companies like Facebook to address the glaring problems with their businesses." (Warren 2019). Indirectly it might also help generate greater transparency and oversight of the actions of dominant technological companies, which is something that has been identified many times as an issue.
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations*	Limitations: include limitations in antitrust enforcement officials' knowledge and the potential impact of ill-advised investigations and prosecutions on markets. (Cass 2012)

Option: Using anti-trust regulations to break up big tech and appoint regulators to reverse illegal and anti-competitive tech mergers Proposer: Elizabeth Warren, US Senator. Reference/link to relevant document: https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c Assessed by: TRI with inputs from UCLANCY Date of assessment: 7 Nov 19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).	Risks: might be similar to risks associated with regulation in general. Challenges: pertain to how the proposal will be implemented in practice. E.g. defining what conduct contravenes antitrust law (Cass 2012).
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	Not as currently outlined. It has been criticised for being too fuzzy. See https://www.economist.com/business/2019/10/24/dismembering-big-tech
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	As per the envisaged legislation in the proposal, "To enforce these new requirements, federal regulators, State Attorneys General, or injured private parties would have the right to sue a platform utility to enjoin any conduct that violates these requirements, to disgorge any ill-gotten gains, and to be paid for losses and damages. A company found to violate these requirements would also have to pay a fine of 5 percent of annual revenue." Warren's proposal also anticipates appointing "regulators committed to reversing illegal and anti-competitive tech mergers". (Warren 2019)
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	There will be implementation burdens on legislators (defining the letter and scope of the law) and on enforcement authorities (selecting appropriate targets for enforcement action and in making enforcement decisions).
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Warren expects the proposals will give a fair shot to small businesses to sell their products, be less smothered by competition from the likes of Google. They are also expected to benefit competing entrepreneurs and content creators (by helping them retain value of their content).
11. Whose rights and/or interests does this option neglect?	The proposal will definitely adversely affect big technological companies especially those listed in the proposal – Amazon, Google, Facebook. These businesses (if hit by such proposals – law and/or enforcement actions) would face increased (direct and indirect) costs of addressing these plus losing resources to facing such scrutiny. One news article (Economist 2019) outlines that "Break-ups could destroy value" - how they operate might be affected and this might compromise their ability to deliver their

Option: Using anti-trust regulations to break up big tech and appoint regulators to reverse illegal and anti-competitive tech mergers Proposer: Elizabeth Warren, US Senator. Reference/link to relevant document: https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c Assessed by: TRI with inputs from UCLANCY Date of assessment: 7 Nov 19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	offerings and maintain their competitive advantage. Antitrust lawsuits are both expensive and disruptive to business and might adversely affect innovation.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	It seeks to protect democracy and privacy by promoting healthy competition and forcing big tech companies to be more responsive to user concerns.
13. How does it address ethics and ethical principles? Which ones?	It does not address ethics and ethical principles.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not elaborated.
16. What provisions are there for regular review and update?	Not elaborated
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	The proposal has been subject to a lot of criticism – some economists say it will not solve problems that are of concern to lawmakers and citizens (e.g., privacy) nor is it feasible as such breaks-ups and unwindings are rare and hard to achieve due to their complexity (Matsakis 2019). Matsakis also suggests that given the high stakes, "if the government loses a case against one of the big tech companies, it could set a weaker precedent for antitrust enforcement in the future." (Matsakis 2019)
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Some say it might strip the company of its best prospects for growth. (Waters 2019). Sokol and Comerford underline that "using antitrust as a sword to address Big Data concerns risks reducing competition and innovation from new products." They further outline that "Antitrust intervention over market forces threatens consumer welfare, especially is fast moving markets, and proposed remedies, such as limiting the collection and use of Big Data or forcing large firms to share with rivals, are likely to harm competition and in- novation, and in fact may raise privacy concerns." (Sokol and Comerford 2016)

Option: Using anti-trust regulations to break up big tech and appoint regulators to reverse illegal and anti-competitive tech mergers Proposer: Elizabeth Warren, US Senator. Reference/link to relevant document: https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c Assessed by: TRI with inputs from UCLANCY Date of assessment: 7 Nov 19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The EU might not have direct jurisdiction on the matter of breaking up big tech companies. However, some other actions could be envisaged through competition law – e.g., disempowering through fines or mandating that some activities must be blocked as illegal. The EU and/or Member States could find business activity unlawful and pause the company. Views have been expressed that if this was possible it might only be used as a last resort option (given the lengthy court process involved) and the more preferred action might be to enable access to the data of such companies (see https://www.debatingeurope.eu/2019/06/26/should-the-eu-break-up-big-tech-companies/#.XcQngi2cbjB)
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	Experts highlight that anti-trust cases with sanctions of break-ups are expensive, and often produce disappointing results. (See expert views in Waters 2019) In addition, the proposal faces the certainty of court challenges, with appeals to the US Supreme Court. The US Justice Department has recently indicated reluctance to use antitrust laws for regulatory purposes and would likely oppose efforts to use antitrust laws in a regulatory manner. "Antitrust is law enforcement, it's not regulation. At its best, it supports reducing regulation, by encouraging competitive markets that, as a result, require less government intervention." (See remarks from Deputy Assistant Attorney General Barry Nigro on 13 Dec 2017, quoting Assistant Attorney General Makan Delrahim, at https://www.justice.gov/opa/speech/deputy-assistant-attorney-general-barry-nigro-delivers-remarks-capitol-forum-and-cqs). The US Supreme Court (most recently in a unanimous decision in *Verizon Communications, Inc. v Law Offices of Curtis V. Trinko* LLP, 540 US 398 (2004), available at https://www.law.cornell.edu/supct/html/02-682.ZO.html) has shown a reluctance to use antitrust laws to require market-dominant companies to assist their competitors. Therefore, the passage and enforcement of new legislation to enact this proposal would face significant opposition from the current US administration and reluctance from the current US Supreme Court. It may be more feasible for EU regulators to require the companies in question to share their datasets in order to reduce anticompetitive effects, based on existing EU competition law on

Option: Using anti-trust regulations to break up big tech and appoint regulators to reverse illegal and anti-competitive tech mergers Proposer: Elizabeth Warren, US Senator. Reference/link to relevant document: https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c Assessed by: TRI with inputs from UCLANCY Date of assessment: 7 Nov 19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	abuse of a dominant market position. For example, subsection (b) of Art. 102 of the Treaty on the Functioning of the European Union identifies "limiting production, markets or technical development to the prejudice of consumers" as an example of abuse of a dominant market position (Consolidated Version of the Treaty on the Functioning of the European Union [2016] OJ C202/89, available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C:2016:202:FULL&from=EN)
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	2
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	This proposal might answer a call for addressing the power imbalances (especially with regard to the control of personal data) with respect to big tech corporations versus the interests of smaller businesses and/or the public interest (protection of democracy). And while it is an easy proposal, some actions will be taken to address the issues (as indicative in the case of the US and the launch of antitrust investigations; also in Germany, France, the European Union, Israel, India, Singapore, Russia, Mexico and Australia) (Stoller 2019). The factors critical to its success are whether other legislative and regulative tools are able to address and/or redress the concerns of data power imbalances and whether further harms to consumer welfare result.
References consulted	Cass, Ronald A. "Antitrust for high-tech and low: regulation, innovation, and risk." *JL Econ. & Pol'y* 9 (2012): 16 Langlois, Richard N, "Hunting the big five: Twenty-first century antitrust in historical perspective." *Available at SSRN 3124356* (2018). Matsakis, Louise, "Break Up Big Tech? Some Say Not So Fast", *Wired*, 6 July 2019. https://www.wired.com/story/break-up-big-tech-antitrust-laws/ The Economist, "Breaking up is hard to do: Dismembering Big Tech", *The Economist*, 24 Oct 2019. https://www.economist.com/business/2019/10/24/dismembering-big-tech Sokol, Daniel D., & Roisin Comerford, "Antitrust and Regulating Big Data", 23 *Geo. Mason L. Rev*. 1129, 2016.

Option: Using anti-trust regulations to break up big tech and appoint regulators to reverse illegal and anti-competitive tech mergers Proposer: Elizabeth Warren, US Senator. Reference/link to relevant document: https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c Assessed by: TRI with inputs from UCLANCY Date of assessment: 7 Nov 19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	Stoller, Matt, "The great breakup of big tech is finally beginning", The Guardian, 9 September 2019. https://www.theguardian.com/commentisfree/2019/sep/09/the-great-break-up-of-big-tech-is-finally-beginning Warren, Elizabeth, "Here's how we can break up Big Tech", Medium, 8 March 2019. https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c Waters, Richard, "Three ways that Big Tech could be broken up", Financial Times, 7 June 2019. https://www.ft.com/content/cb8b707c-88ca-11e9-a028-86cea8523dc2

4.30. Three-level obligatory impact assessments for new technologies

Option: Three-level obligatory impact assessments for new technologies Proposer: Paul Nemitz Reference/link to relevant document: Nemitz P., "Constitutional democracy and technology in the age of artificial intelligence", *Phil. Trans. R. Soc. A* **376**: 20180089. http://dx.doi.org/10.1098/rsta.2018.0089 Assessed by: TRI Date of assessment: 11 November 19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	Nemitz proposes an obligatory three-level impact assessment for new technologies. First, the parliamentary technology impact assessment, at the level of policy making and legislation. Second, at the level of the developers and users of such technology. Third, individuals concerned by the use of AI should have a right, to be introduced by law, to an explanation of how the AI functions, what logic it follows, and how its use affects the interests of the individual concerned. (Nemitz 2018)
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g.,	Basis: proposed as an extension by law of being obligatory when AI processes personal data in the context of automated decision making to include all

Option: Three-level obligatory impact assessments for new technologies Proposer: Paul Nemitz Reference/link to relevant document: Nemitz P., "Constitutional democracy and technology in the age of artificial intelligence", *Phil. Trans. R. Soc. A* **376**: 20180089. http://dx.doi.org/10.1098/rsta.2018.0089 Assessed by: TRI Date of assessment: 11 November 19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
national or international, topic/domain/tech specific/general))?	aspects of democracy, rule of law and fundamental rights. Nature: proposed as obligatory. Scope: Scoped broadly in terms of for new technologies – especially high-risk technologies.
3. Purpose/objective/what need does the option fulfil?	It would potentially help "strengthen trust in technology in the age of Artificial Intelligence" (Nemitz 2018). It would more widely to contribute to the help shape policy and public opinion (and broaden knowledge base) on the impacts of science and technology.
4. What gap does it address?	Lack of transparency about capabilities and impacts of AI. It might help connect new technology to constitutional principles and human rights, especially where as Nemitz outlines, conflicts of interest cannot be solved by unenforceable ethics codes or self-regulation.
5. What added value does it have?	It could "help the corporations, their leaders and the engineers developing the new technologies and their applications to own up to the power they exercise. They would thus help to instil a new culture of responsibility of technology for democracy, rule of law and fundamental rights" (Nemitz 2018)
6. What are the limitations, risks and challenges? *(internal note: (look up research/policy documents that have analysed this option or its application in other areas. Limitations are what might restrict it; risks are potential or possible harms; challenges are difficulties it might face or be presented with).*	Limitations: these might include whether it is able to get a good grasp of medium to long-term impacts. Other limitations include time and resource constraints Risks: One key risk relates to how validity of impact assessments is affected by timing. Too premature or too late impact assessments carry inherent risks. Challenges: these will include getting institutional resistance and/or buy-in, adequate resourcing, mismatch in expectations of those conducting impact assessments.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	The overall vision has been expressed but finer details require elaboration. The connections between the three levels need further specification including what happens to the results of the assessment at each level. Further specification is also much desired as to how the assessment would evaluate compliance with the rule of law and democracy (especially at levels 2 and 3).
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Nemitz states, "Decisions as to the consequences to draw from the risk assessments carried out by experts are in the hands of governments and legislators, and on the EU level in the hands of the Commission and the

Option: Three-level obligatory impact assessments for new technologies Proposer: Paul Nemitz Reference/link to relevant document: Nemitz P., "Constitutional democracy and technology in the age of artificial intelligence", *Phil. Trans. R. Soc. A* **376**: 20180089. http://dx.doi.org/10.1098/rsta.2018.0089 Assessed by: TRI Date of assessment: 11 November 19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	Council and Parliament as co-legislators." Further, "the compliance with the standards for the impact assessment would have to be controlled by public authorities and non-compliance should be subject to sufficiently deterrent sanctions. In cases of AI to be used in the exercise of public power or in wide public use, the impact assessment would have to be made available to the public, and in high-risk cases, the public authority making use of AI would have to carry out its own complementary assessment and present a risk reduction and mitigation plan." (Nemitz 2018)
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Level 1 – there will be legislative change/amendments burden for policymakers and legislators, if new legislation is required (and this might add complexity depending on expectations) Level 2 and 3 – there will be some implementation burdens (resources, time and cost including on training, raising awareness) for developers and users of technology who have to carry out such assessments and/or be subject to such assessment.
10. Which stakeholders would benefit most from the use of this option? [*Developers/manufacturers/suppliers (industry); users; policymakers; regulators; civil society; individuals, others* (please specify)]	Individuals. Policy-makers.
11. Whose rights and/or interests does this option neglect?	Industry especially SMEs.
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	It explicitly supports human rights e.g., privacy, data protection.
13. How does it address ethics and ethical principles? Which ones?	Transparency.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No

Option: Three-level obligatory impact assessments for new technologies Proposer: Paul Nemitz Reference/link to relevant document: Nemitz P., "Constitutional democracy and technology in the age of artificial intelligence", *Phil. Trans. R. Soc. A* **376**: 20180089. http://dx.doi.org/10.1098/rsta.2018.0089 Assessed by: TRI Date of assessment: 11 November 19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	Not elaborated.
16. What provisions are there for regular review and update?	Not elaborated.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Its feasibility would depend on the political will to put this into action and institutional resistance and/or buy-in. Further it would also have to connect with other obligatory requirements such as data protection impact assessments under the GDPR.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	Yes, it will impact the ability of businesses and others to innovate in as much as it will promote responsible and lawful innovation in AI. We do not anticipate businesses will be restricted from innovating as such, but there might be some impact on smaller innovators.
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	This proposal is a good fit with the EU legal framework in as much as it would help Member States comply with the need to apply measures to protect the rule of law, democracy and human rights against AI-based infringements.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	-
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	4
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	There is a strong case to make impact assessments mandatory in line with that Nemitz proposes – given especially the high- risks connected to AI and its adverse impacts – they might help mitigate and address any adverse effects head-on and early. However, what needs clarification is the clear placement and connection the proposed three-level assessment with existing legislation (e.g., the GDPR), data protection impact assessments under the GDPR. Factors critical to its adoption/success include (in addition to some points already made before): a strong governance framework, stakeholder buy-in, transparency that facilitates some form of

Option: Three-level obligatory impact assessments for new technologies Proposer: Paul Nemitz Reference/link to relevant document: Nemitz P., "Constitutional democracy and technology in the age of artificial intelligence", *Phil. Trans. R. Soc. A* **376**: 20180089. http://dx.doi.org/10.1098/rsta.2018.0089 Assessed by: TRI Date of assessment: 11 November 19 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	external oversight and review to verify that such impact assessments are fit for purpose. Further as stated in the report of the *Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT), "* …in order for impact assessment approaches to provide real and substantive protection, it will be necessary to develop a clear and rigorous methodological approach that firms and other organisations are willing to adopt consistently and in ways that reflect a genuine commitment to identifying human rights risks, rather than merely regarding them as a bureaucratic burden resulting in 'ritual' displays of formal compliance without any genuine concern to respect human rights." (Council of Europe, 2018).
References consulted	Council of Europe Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence, *A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework*, 9 Nov 2018. https://rm.coe.int/draft-study-of-the-implications-of-advanced-digital-technologies-inclu/16808ef255 Nemitz P., "Constitutional democracy and technology in the age of artificial intelligence", *Phil. Trans. R. Soc.*, A 376: 20180089. http://dx.doi.org/10.1098/rsta.2018.0089

4.31. Regulatory sandboxes

Option: Regulatory sandboxes Proposer: European Commission, European Parliament, EC HLEG AI Reference/link to relevant document: See references below Assessed by: UCLANCY, date assessed 15 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
1. Outline its relevance/connection to AI and big data analytics (what does it regulate? Does it require specific features to be built in AI, such as transparency, robustness and security measures?) Give an application example)	Proposals to establish "regulatory sandboxes" for regulation of AI and autonomous systems have been suggested by: • the European Commission in the 2018 Coordinated Plan on Artificial Intelligence[1]; • the European Parliament in a 2019 resolution on a comprehensive European industrial policy on artificial intelligence and robotics[2]; and • the High-Level Expert Group on AI in its 2019 Policy and Investment Recommendations for Trustworthy AI[3]. A regulatory sandbox is a framework set up by a regulatory body to allow small scale, live testing of the distribution and use of innovations by stakeholders in a controlled environment under the regulator's supervision[4]. The regulator determines the criteria for participants, the scope and capacity of the sandbox, the testing parameters and conditions, the evaluation methodology and the exit criteria[5]. A regulatory sandbox is not a regulatory mechanism itself but is a way to allow regulators and other stakeholders to test proposed regulatory schemes for new technologies in a controlled manner.
2. What is its basis (on which the regulatory option is created - law? if yes which one), nature (e.g., is it binding?) and scope (e.g., national or international, topic/domain/tech specific/general))?	Basis: Administrative process to allow controlled testing Nature: Participation is voluntary, but the sandbox "rules" are binding on the participants during the sandbox exercise. Scope: Variable, depends on the parameters of the sandbox.
3. Purpose/objective/what need does the option fulfil?	Regulatory sandboxes allow testing of regulatory schemes for new technology in a more controlled environment, with close cooperation between regulators and stakeholders.
4. What gap does it address?	Regulatory sandboxes shorten the feedback loop for new regulatory schemes and new technologies, giving regulators and market participants the opportunity to "test-drive" how the regulations will work with the new technologies in a limited environment.
5. What added value does it have?	Closer communication and cooperation between regulators and stakeholders may help increase trust and confidence in public and private decision-making processes.
6. What are the limitations, risks and challenges?	Limitations, risks and challenges include:[4,5] • The sandbox process may give participants unfair competitive advantages both in regulatory advice and in being first to the market, particularly if the selection criteria are defined vaguely or there is a lack of transparency leading to selection bias or the appearance of selection bias.

Option: Regulatory sandboxes Proposer: European Commission, European Parliament, EC HLEG AI Reference/link to relevant document: See references below Assessed by: UCLANCY, date assessed 15 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
	• A sandbox needs transparency which must be balanced with classified and commercially sensitive information and trade secrets of participants. • With highly advanced technology, the regulator may be ill-equipped to select the most appropriate candidates to participate in the sandbox. • Liability issues in case of failed testing that resulted in harm to customers or other market participants, which may threaten the reputation of the regulator and trust of customers in the regulatory system. • If the regulatory scheme involves both EU-level and national regulation, then the sandbox would need to include regulators from both levels as well.
7. Is the option sufficiently clear, specific and able to be effectively and efficiently operationalised? If not, why**?**	Each regulatory sandbox is unique. In order to implement this proposal, the regulator needs to determine[4,5]: • objectives of the sandbox • eligibility to participate in the sandbox, both in terms of the qualifications of the market players and the qualifications of the innovative technology • rules for participants (transparency, accountability, oversight and assessment, allocation of risks, safeguards, and operational restrictions such as limits on the number and location of users, limits on types of uses, special testing requirements, etc.) • timing for applicants and sandbox tests • costs to the regulator and the sandbox entities • what happens to existing users when the sandbox period ends.
8. What explicit monitoring, oversight and enforcement mechanisms does the option include? Is there a gap/room for improvement?	Creation of each regulatory sandbox should include mechanisms for transparency, accountability, monitoring, oversight and assessment.
9. What implementation burdens (e.g., administrative or other burdens) might/does it create for: a. Citizens b. Public administrations c. Businesses and particularly SMEs?	Citizens: not applicable Public administrations: Regulatory sandboxes must be designed, implemented and closely overseen by regulators. Businesses and particularly SMEs: Administrative burdens on participating businesses will vary depending on the requirements of each sandbox.
10. Which stakeholders would benefit most from the use of this option?	• Regulators and policymakers receive feedback quickly and may be able to better tailor regulation to market needs; • Participating suppliers may be able to shorten the time to market for new products and to have stronger guarantees that their products are compliant with the existing legal requirements; and Participating users can get access to new technologies more quickly, with their rights being adequately protected.[5]

Option: Regulatory sandboxes Proposer: European Commission, European Parliament, EC HLEG AI Reference/link to relevant document: See references below Assessed by: UCLANCY, date assessed 15 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
11. Whose rights and/or interests does this option neglect?	-
12. Does it explicitly support or adversely affect human rights (if yes, which ones)? If not, how might it boost human rights?	No. A proposal to use regulatory sandboxes to test regulatory schemes for new technology is neutral with respect to human rights, but human rights impacts resulting from use of the technology could be built into the assessment process.
13. How does it address ethics and ethical principles? Which ones?	Depending how it is designed and managed, a regulatory sandbox can enhance transparency and trust.
14. Does it explicitly consider gender dimensions? How? E.g., in the composition of the agency/body, consideration of gender equality, gender neutrality.	No, but a sandbox can factor gender equality into the selection of participants.
15. Does it have a well-clarified source of funding, present and future, especially where the option is a body/agency/authority? Outline.	No funding source is identified. Sandboxes are typically funded by the regulatory agency overseeing the sandbox.
16. What provisions are there for regular review and update?	Sandboxes are for limited durations. Each regulatory sandbox should include mechanisms for regular review and assessment.
17. Is it feasible, sustainable (e.g., supported by policy and market incentives) and future-proof? Or might it be adversely affected by future developments e.g., technological, policy changes, social demands?	Feasible: Yes. Regulatory sandboxes have been used in more than 20 countries[162] to test regulation of new technologies. Sustainable: Yes. A well-designed sandbox benefits all of the participants (see #10 above). Future-proof: Yes. A regulatory sandbox is a time-limited testing mechanism with rapid feedback, allowing for rapid adaptation to future developments.
18. Will it adversely impact the ability for businesses and others to innovate? [elaborate, if yes]	No. Regulatory sandboxes facilitate innovation by allowing quicker market introduction in a controlled environment. They are a good, agile mechanism.

[162] Jenik, Ivo and Kate Lauer, Regulatory Sandboxes and Financial Inclusion, CGAP Working Paper, October 2017. https://www.cgap.org/sites/default/files/Working-Paper-Regulatory-Sandboxes-Oct-2017.pdf

Option: Regulatory sandboxes Proposer: European Commission, European Parliament, EC HLEG AI Reference/link to relevant document: See references below Assessed by: UCLANCY, date assessed 15 Nov 2019 Stakeholder(s) consulted in option assessment: -	
Criteria/touch point	**Assessment**
19. Outline its suitability/fit with the EU legal framework (assess against the powers and competences of the EU to implement these actions in accordance with the EU acquis)	The regulatory sandbox mechanism can fit within an existing regulatory mechanism at any level: EU, member state, or local.
20. Any other implementation challenges (especially those not covered above e.g., complexities)?	None identified
21. Based on this study, how likely is this option to succeed ? (1 – Extremely unlikely 2 – unlikely 3 – Neutral, 4 – likely 5 – Extremely likely)?	4 (likely), if the parameters of the sandbox are set thoughtfully (see #7 above). There is support for the general idea of regulatory sandboxes for regulation of artificial intelligence applications from both the European Parliament and the European Commission.
22. Overall conclusion (What are the factors critical to its adoption and/or success?)	Critical factors include: • Thoughtful design of the sandbox parameters (see #7) • Transparency in the design, operation and outcomes • Close communication and cooperation with stakeholders
References consulted	1. European Commission, Annex to the Coordinated Plan on Artificial Intelligence (COM(2018) 795) 7 December 2018. https://ec.europa.eu/digital-single-market/en/news/coordinated-plan-artificial-intelligence 2. European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics (2018/2088(INI)), item 32, 12 February 2019. http://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html 3. High-Level Expert Group on AI (AI HLEG), Policy and Investment Recommendations for Trustworthy AI, European Commission, item 29.2, 26 June 2019. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence 4. Jenik, Ivo and Kate Lauer, Regulatory Sandboxes and Financial Inclusion, CGAP Working Paper, October 2017. https://www.cgap.org/sites/default/files/Working-Paper-Regulatory-Sandboxes-Oct-2017.pdf 5. Yordanova, "Katerina, The Shifting Sands of Regulatory Sandboxes for AI", CiTiP blog of the KU Leuven Centre for IT & IP Law, 18 July 2019. https://www.law.kuleuven.be/citip/blog/the-shifting-sands-of-regulatory-sandboxes-for-ai/

5. Views of other stakeholders on AI and regulation

In this Annex, we present an analysis of other stakeholder perspectives, complementing Section 3. This analysis informed our analysis in Section 4 and is presented here as being of additional interest.

5.1 Academia

There is a substantial literature on AI and regulation, especially from the last three years. Below we provide examples of some of the main topics covered in the literature between 2017-2019. The overview is based on an analytic selection of papers, based on the search-string "AI+regulation" on SSRN and LawArXiv).

Many articles focus on human rights, especially rights defined in the EU *General Data Protection Regulation* (GDPR), and whether the GDPR provides appropriate protections of those rights, focusing on issues such as: a right to explanation[163]; a right to be forgotten[164]; freedom of expression[165]; a right to avoid automated decision-making[166]; and ownership of data[167].

There are also many articles discussing either sector- or technology-specific regulations, such as the financial sector and cryptocurrencies.[168] For example, Gal & Elin-Koren discuss the possibility of

[163] E.g., Casey, Bryan, Ashkon Farhangi, and Roland Vogl, "Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise", *Berkeley Technology Law Journal*, Vol. 34, Issue 1, 2019, pp. 143-188; Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation", *International Data Privacy Law,* Vol. 7, Issue 2, May 2017, pp. 76-99; Wachter, Sandra, Brent Mittelstadt, and Chris Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR", *Harvard Journal of Law & Technology*, Vol. 31, Issue 2, 2018, pp. 841-887; and Wachter, Sandra, and Brent Mittelstadt, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI", *Columbia Business Law Review*, Vol. 2019, No. 2, May 2019. https://journals.library.columbia.edu/index.php/CBLR/article/view/3424.

[164] E.g., Villaronga, Eduard Fosch, Peter Kieseberg, and Tiffany Li. "Humans forget, machines remember: Artificial intelligence and the Right to Be Forgotten". *Computer Law & Security Review*, Vol. 34, Issue 2, 2017, pp. 304-313. https://doi.org/10.1016/j.clsr.2017.08.007

[165] E.g., Balkin, J. M., "Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation", *UC Davis Law Review*, Vol. 51, Issue 3, 2018, pp. 1151-1210.

[166] E.g., Kuner, Christopher, Dan Jerker B. Svantesson, Fred H. Cate, Orla Lynskey, and Christopher Millard, "Machine learning with personal data: is data protection law smart enough to meet the challenge?", *International Data Privacy Law*, Vol. 7, Issue 1, 2017, pp. 1-2. https://doi.org/10.1093/idpl/ipx003; Veale, Michael and Lilian Edwards, "Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling", *Computer Law & Security Review*, Vol. 34, Issue 2, 2018, pp. 398-404. https://doi.org/10.1016/j.clsr.2017.12.002.

[167] E.g., Arrieta-Ibarra, Imanol, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier and E. Glen Weyl, "Should We Treat Data as Labor? Moving beyond 'Free'", *AEA Papers and Proceedings*, Vol. 108, 2018, pp. 38-42. https://ssrn.com/abstract=3093683; Determann, L., "No One Owns Data", *Hastings Law Journal*, Vol. 70, Issue 1, 2019, pp. 1-44. http://www.hastingslawjournal.org/wp-content/uploads/70.1-Determann.pdf.

[168] E.g., Brummer, Christ and Yesha Yadav, "Fintech and the Innovation Trilemma", *The Georgetown Law Journal*, Vol. 107, Issue 2., 2019, pp. 235-307, https://georgetownlawjournal.org/articles/298/fintech-and-the-innovation-trilemma/pdf; Gal, Michal S. and Niva Elkin-Koren, "Algorithmic Consumers", *Harvard Journal of Law & Technology*, Vol. 30, Issue 2, 2017, pp. 309-353; Omarova, S. T., "New Tech v. New Deal: Fintech as a

algorithmic consumption, in which an algorithm provides a consumer with services based on data (e.g., "When the pet food runs low, the algorithm automatically seeks the best deal and orders food of a kind which best fits your pet's needs").[169] As they argue, "These developments raise new and important conceptual and regulatory issues. Indeed, some of the most fundamental conceptions about how markets operate may need to be reevaluated."[170]

There are also discussions relating to medical technology.[171] For example, Chung & Zink argue that AIs used in medical practices (such as IBM's Watson), "warrants a unique legal status akin to personhood and is analogous to a medical resident", and that "liability for wrongful diagnosis by medical AI should attach on a medical malpractice basis rather than through a products liability or vicarious liability scheme".[172]

There are also discussions of machines as regulators[173] or as arbitrators[174]. For example, Alarie et al. argue that "Regulators – plagued by problems of resource constraints, the scarcity of human capital, and inconsistency – can use machine learning tools to provide faster advice and rulings to citizens. Machine learning tools can be also used to refine and improve laws, making them more applicable and relevant to the circumstances faced by citizens. Indeed, in the future, we may come to puzzle about how it came to be that 20th century regulation in an algorithmically limited environment was ever supposed to have worked at all."[175]

There are also more general overviews. For example, in a working paper, Petit proposes "to index the intensity of regulatory response upon the nature of the externality created by an AI application. When AI-generated externalities are discrete, social planners should defer to ex post litigation before courts. When AI-generated externalities are systemic, social planners should envision ex ante regulation, but carefully test and experiment."[176] Thierer et al. argue for a light-touch approach. Specifically, they argue that, "Unless a compelling case can be made that a new invention will bring serious harm to society, innovation should be allowed to continue unabated, and problems, if they develop at all, can be addressed later."[177]

Systemic Phenomenon", *Yale Journal on Regulation*, Vol. 36, Issue 2, 2019, pp. 735-793. http://yalejreg.com/articlepdfs/36-JREG-735-Omarova.pdf.

[169] Gal and Elkin-Koren, op. cit., p. 310.

[170] Gal and Elkin-Koren, op. cit., p. 311.

[171] E.g., Chung, Jason and Amanda Zink, "Hey Watson – Can I Sue You for Malpractice? Examining the Liability of Artificial Intelligence in Medicine", *Asia Pacific Journal of Health Law & Ethics,* Vol. 11, Issue 2, 2018, pp. 51-80; Nuffield Council on Bioethics, "Artificial intelligence (AI) in healthcare and research", Briefing Note, 2018. http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf; and Price II, W. N., "Artificial Intelligence in Health Care: Applications and Legal Implications", *The SciTech Lawyer*, Vol. 14, Issue 1, 2017, pp. 10-13. https://repository.law.umich.edu/articles/1932.

[172] Chung and Zink. op. cit., p. 51.

[173] E.g., Alarie, Benjamin, Anthony Niblett, and Albert H. Yoon, "Regulation by Machine", 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain. http://www.mlandthelaw.org/papers/alarie.pdf; Coglianese, Cary and David Lehr, "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era", *Faculty Scholarship at Penn Law*, 2017, 1734. http://scholarship.law.upenn.edu/faculty_scholarship/1734

[174] E.g., Sela, A., "Can Computers Be Fair? How Automated and Human-Powered Online Dispute Resolution Affect Procedural Justice in Mediation and Arbitration", *Ohio State Journal on Dispute Resolution*, Vol. 33, Issue 1, 2018, pp. 91-148.

[175] Alarie et al, op cit., pp. 5-6.

[176] Petit, N., "Law and regulation of artificial intelligence and robots: Conceptual framework and normative implications", Working Paper, 2017, p. 30. https://dx.doi.org/10.2139/ssrn.2931339

[177] Thierer, Adam, Andrea Castillo O'Sullivan, and Raymond Russell, "Artificial Intelligence and Public Policy", Mercatus Research, Mercatus Center at George Mason University, 2017. https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf, p. 2.

5.2 Media

This section surveys four media outlets by analysing a selected sample set of articles (politico.eu, reuters.com, euronews.com, observer.com); the selection was based on reflection following discussion with stakeholders. We used a site-relative search-string on google, using Chrome in Incognito-mode, searching for *AI+regulation*. Given that the total hits varied from 53 to 10,400, we selected a limited sample by performing monthly searches from September 2018 – August 2019, then we checked if the top hit matched our selection criteria,[178] if not, we moved further down the list. If none of the five top hits matched, we did not select a sample from that journal that month. After this, we summated the articles. Below, we present an analytic synthesis of that summation. Unsurprisingly, the articles presented a fairly diverse set of ideas, varying in form from news articles, columns, opinion pieces, and sponsored content.

A lot of content focuses on the work done by AI HLEG, and their work has been described as a "silver bullet in global AI battle".[179] But while there are articles on the value of ethically-regulated AI, including calls for regulations against mass surveillance[180], and regulatory gaps because privacy laws have failed to keep up with new technology, such as drones[181], there are also those who worry about over-regulation, including an EU Justice Commissioner[182], representatives for commercial interests,[183] and a member of the AI HLEG promoting self-regulation and *ex-post* regulation.[184]

While commercial interests aim to promote less regulation through "sponsored content"(i.e., purchased adds in the form of an article)[185], similar ideas are also being pushed in opinion pieces. In some opinion pieces, representatives from think tanks and non-profit organization, express a worry about regulations that will stall development. They argue that there are obvious limits of to EU's perceived reliance on ethical AI as a competing factor.[186] Instead, they suggest the EU should focus on competing with China. [187]

---

[178] The selection criteria were designed to be as inclusive as possible. We excluded articles that did not mention or discuss AI regulation (e.g., the hits may have been due to links on the website). We also excluded hits that were external material uploaded to the news outlet (e.g., the AI HLEG guidelines).

[179] Delcker, J., "Europe's silver bullet in global AI battle: Ethics", *Politico* (Article), 17 March 2019. https://www.politico.eu/article/europe-silver-bullet-global-ai-battle-ethics/

[180] Delcker, J., "AI experts call to curb mass surveillance", *Politico* (Article), 24 June 2019. https://www.politico.eu/article/eu-experts-want-curtailing-of-ai-enabled-mass-monitoring-of-citizens/

[181] Leon, H., "Top Secret Military-Grade Surveillance Drones Might Be Coming To Your Neighborhood", *Observer,* 28 June 2019, https://observer.com/2019/06/gorgon-stare-aerial-surveillance-drones/

[182] Delcker, op. cit., 24 June 2019.

[183] Koschwitz, L., "The copyright reform bug that risks derailing Europe's AI ambitions", *Politico* (Sponsored Content), 5 September 2018. https://www.politico.eu/sponsored-content/the-copyright-reform-bug-that-risks-derailing-europes-ai-ambitions/

[184] Delcker, J., "Europe's AI ethics chief: No rules yet, please", *Politico* (Article), 30 October 2018, https://www.politico.eu/article/pekka-ala-pietila-artificial-intelligence-europe-shouldnt-rush-to-regulate-ai-says-top-ethics-adviser/

[185] Koschwitz, op. cit. 2018.

[186] Castro, D. *("*the director of the Center for Data Innovation"), "Europe will be left behind if it focuses on ethics and not keeping pace in AI development", *Euronews* (Opinion), 7 August 2019, https://www.euronews.com/2019/08/07/europe-will-be-left-behind-if-it-focuses-on-ethics-and-not-keeping-pace-in-ai-development

[187] Chivot, E. ("a senior policy analyst at the Center for Data Innovation") and Daniel Castro ("the director of the Center for Data Innovation and vice president of the Information Technology and Innovation Foundation"), "The EU's 'softball' approach to Artificial Intelligence will lose to China's 'hardball'", *Euronews* (Opinion), 5 February 2019,

There are also similar themes describing national investments in AI and related policy goals, such as Germany's investments plans, which includes "loosen regulation"[188]. In separate articles officials implicitly promote a light-touch by arguing for the importance to "promote not impede" AI.[189] A Finnish plan (including co-operation with Sweden and Estonia), includes similar aims to lobby at the EU-level for loosening some regulations.[190] But contrary ideals are also being promoted, such as warnings against self-regulation[191] (i.e., promoting a heavier touch), and that surveillance and military applications imply a need for regulation.[192] We also see national trends which promote the need for stricter regulations, e.g., in the UK, calling for the creation of a new regulatory body, and protecting consumers against 'toxic content' online, holding companies responsible for content such as hate speech, misinformation or harmful information.[193]

5.3 Industry and professional associations

As part of this study we looked at computer, AI, or big data industry and professional associations. After a review, we selected the Big Data Value Association (BDVA), the Institute of Electrical and Electronics Engineers (IEEE), the Association for Computing Machinery (ACM), and Partnership on AI. The selection was based on reflection following discussion with stakeholders. We also looked at EurAI,[194] but it currently does not have any official policies suitable for analysis.

The BDVA ("industry-driven international not–for-profit organization")[195] wants to develop "the European AI Ecosystem" by promoting "adoption of AI technologies in all industrial sectors",[196] increasing access to data, and through public-private collaboration.[197] It envisions a need for a "firm, yet flexible" regulatory and legal system on AI that "delivers legal certainty and predictability".[198] It promotes a legal system based on general AI regulations, complemented by sector-specific amendments given the technologies' varied applications, especially in safety-critical domains (e.g., transport and healthcare).[199]

---

https://www.euronews.com/2019/02/05/the-eu-s-softball-approach-to-artificial-intelligence-will-lose-to-china-s-hardball-view

[188] Delcker, J. "Germany's €3B plan to become an AI powerhouse", *Politico* (Article), November 2018, https://www.politico.eu/article/germanys-plan-to-become-an-ai-powerhouse/

[189] Carrel, P., "Germany must close digital technology gap, Merkel ally says", *Reuters* (TECHNOLOGY NEWS), 8 November 2018. https://www.reuters.com/article/us-germany-tech/germany-must-close-digital-technology-gap-merkel-ally-says-idUSKBN1ND1W4

[190] Delcker, J., "Finland's grand AI experiment", *Politico* (Article), 2 January 2019. https://www.politico.eu/article/finland-one-percent-ai-artificial-intelligence-courses-learning-training/

[191] "Self-regulation of AI is 'dangerous': CognitiveScale CEO", *Reuters* (VIDEO), 23 January 2019. https://mobile.reuters.com/video/2019/01/23/self-regulation-of-ai-is-dangerous-cogni?videoId=506752793&videoChannel=118156

[192] Mak, R., "Breakingviews - Review: Why an AI apocalypse could happen", *Reuters* (BREAKINGVIEWS), 14 June 2019. https://www.reuters.com/article/us-tech-artificial-intelligence-breaking/breakingviews-review-why-an-ai-apocalypse-could-happen-idUSKCN1TF1FH

[193] Cao, S., "It's Serious This Time: Multibillion-Dollar Fines Could Hit Facebook & Google, UK Warns" *Observer*, 28 February 2019, https://observer.com/2019/02/facebook-google-face-multi-billion-dollar-fine-uk-content-regulator/

[194] European Association for Artificial Intelligence.

[195] BDVA, "Data-driven artificial intelligence for European economic competitiveness and societal progress", BDVA Position Statement, November 2018, p. 8, http://www.bdva.eu/sites/default/files/AI-Position-Statement-BDVA-Final-12112018.pdf

[196] BDVA, op. cit., p. 7.

[197] BDVA, op. cit., pp. 6-7.

[198] BDVA, op. cit., p. 8.

[199] BDVA, op. cit., p. 8

The IEEE wants to increase AI competence and support for R&D.[200] In virtue of supporting R&D it wants to remove legal impediments (or legal uncertainty) for reverse-engineering of AI with the purpose of promoting "third-party research on fairness and algorithmic bias, security, privacy, and social impacts of Artificial Intelligence systems".[201] The IEEE generally promotes an "appropriate mechanism" for the creation of AI regulations and AI coordination through stakeholder participation.[202] It wants to prioritize safety and "consider societal implications; public engagement; appropriate levels of public investment; economic and national security impacts; transparency, accountability and explainability; trust and safety assurance; ethical principles; and legal and regulatory compliance".[203] Furthermore, it wants to ensure that AI regulations comply with human rights laws (especially privacy), and that intellectual property laws and liability laws are adapted to the particular nature of AI technology.[204] Lastly, the IEEE also seeks to support and fund AI education and to "facilitate public understanding" of AI technology.[205]

As of yet, the ACM has no official overall position on AI regulations, but promotes through particular policy positions through various policy documents in its US or European policy committees. For example, the ACMEU and Informatics Europe has set out recommendations for technical, ethical, legal, economic, societal, and educational aspects of automated decision-making (ADM), in *When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making*. The recommendations includes value-sensitive design, privacy protection, and legal clarification of responsibility for ADM.[206]

The ACM US has responded to a call by the US Food and Drug Administration (FDA) for public response on *Proposed FDA Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning Based Software as a Medical Device Discussion Paper* (Docket No. FDA-2019-N-1185), and discussed the need for regulations "if a manufacturer proposes to use AI to dynamically change a device's behavior in the field without being subject to a regulated development and testing process". The ACM US Policy Committee notes that if the human in the loop is removed then there are various policy responses, ranging from banning such dynamic AI systems, requiring protections that limit the devices' behavioural changes, or mandating sharing of data to assure validity of that data. Furthermore, the committee also urges the FDA to:

- Employ and require outcomes-driven, automated and (where possible) deterministically reproducible testing outside of the vendors' own development laboratories;
- Require manufacturers to create a common pool of data for input to AI analyses, including both real-world deidentified data and synthetic data; and
- Foster the development of a common pool of varied simulation and other test environments using the deidentified and synthetic data endorsed above.[207]

---

[200] IEEE, "Artificial Intelligence", IEEE Position Statement, Approved by the IEEE Board of Directors, 24 June 2019. https://globalpolicy.ieee.org/wp-content/uploads/2019/06/IEEE18029.pdf
[201] IEEE, op. cit. The position statement has a large overlap with the previous IEEE-USA, Artificial Intelligence Research, Development and Regulation, IEEE-USA Position Statement, Adopted by the IEEE-USA Board of Directors, 10 February 2017, p. 3. https://globalpolicy.ieee.org/wp-content/uploads/2017/10/IEEE17003.pdf.
[202] IEEE, op. cit., p. 4.
[203] IEEE, op. cit., p. 4.
[204] IEEE, op. cit., p. 4.
[205] IEEE, op. cit., p. 5.
[206] Larus, James, Chris Hankin, Siri Granum Carson, Markus Christen, Silvia Grafa, Oliver Grau, Claude Kirchner, Bran Knowles, Andrew McGettrick, Damian Andrew Tamburri, and Hannes Werthner, "When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making", *Informatics Europe & EUACM*, 2018. https://www.acm.org/binaries/content/assets/public-policy/ie-euacm-adm-report-2018.pdf.
[207] Hendler, J. A. (Chair, ACM, US Technology Policy Committee), "Comments to Food and Drug Administration on AI-Augmented Software as a Medical Device Discussion Paper", ACM US Technology Policy Committee, 3

In a joint statement by the ACMEU and ACM US on *Algorithmic Transparency and Accountability* they list seven principles (Awareness, Access and redress, Accountability, Explanation, Data Provenance, Auditability, and Validation and Testing). The set of principles are, they note, "consistent with the ACM Code of Ethics", meaning that the ACM Code of Ethics may serve as a further basis for their policy position on these issues.[208]

The Partnership on AI (an industry consortium formed by Google, Facebook, Amazon, IBM, and Microsoft in 2016[209]) has eight tenets that "members believe in and endeavor to uphold". These tenets include: engaging people; increasing AI literacy; ensuring the positive outcome of AI; open research and dialogue; actively engaging shareholders; and addressing "the potential challenges of AI technologies". These challenges, in turn, include "privacy and security of individuals"; understanding the interests of all potentially impacted parties; "socially responsible, sensitive, and engaged" AI research; "[e]nsuring that AI research and technology is robust, reliable, trustworthy, and operates within secure constraint"; and opposing AI-technology "that would violate international conventions or human rights and promoting safeguards and technologies that do no harm") [210]. Given that the members of the Partnership are developers, it is fair to say that these are minimally, eight tenets of harmonised self-regulation.

5.4 Civil society

Various organisations in civil society aim to influence AI policy and regulations. In this section we focus on some of those organisations that propose policies for any kind of AI-application (rather than policies limited to a singular type of AI-application). We selected The Future of Life Institute, The Public Voice, and (jointly) Article 19 and Privacy International (the selection was based on reflection following discussion with stakeholders).

In 2017, The Future of Life Institute (a non-profit research and outreach institute) organised its second international conference on AI, where it discussed and collaborated on creating "a list of 23 principles, which "aim to reflect "what society should do to best manage AI in coming decades", ranging from research strategies to data rights to future issues including potential super-intelligence"[211]; this has since been signed by" 1521 AI/Robotics researchers and 3298 others."[212] The principles are divided into three sections (Research Issues, Ethics and Values, and Longer-term Issues) and cover a broad array of topics, including beneficial goal-setting, safety, human control and transparency in case of failures, and avoidance of an arms race of lethal autonomous weapons; legal transparency, and responsibility; alignment to values and a broad focus on aspects relating to human rights (i.e., dignity, freedom, cultural diversity, privacy, liberty), and distributional aspects (shared benefits and prosperity); as well as non-subversion of social and civic processes.

June 2019. https://www.acm.org/binaries/content/assets/public-policy/ustpc-comments-fda-software-based-device-safety-060319.pdf

[208] ACM US Public Policy Council, ACM Europe Council Policy Committee, "Statement on Algorithmic Transparency and Accountability", 2017. https://www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf

[209] Hern, A. "'Partnership on AI' formed by Google, Facebook, Amazon, IBM and Microsoft", *The Guardian*, 28 September 2016. https://www.theguardian.com/technology/2016/sep/28/google-facebook-amazon-ibm-microsoft-partnership-on-ai-tech-firms

[210] Partnership on AI, "Tenets". https://www.partnershiponai.org/tenets/

[211] Future of Life Institute, "A Principled AI Discussion in Asilomar". https://futureoflife.org/2017/01/17/principled-ai-discussion-asilomar/

[212] Future of Life Institute, "ASILOMAR AI PRINCIPLES". https://futureoflife.org/ai-principles/

Overall the principles are general, rather than sector-specific, and they adhere to a risk-based approach with a focus on "planning and mitigation efforts", especially for "catastrophic or existential risks".[213]

In 2018, The Public Voice, which cooperates with "ICDPPC, the OECD, UNESCO, and other international organizations, […] [to bring] civil society leaders face to face with government officials for constructive engagement about current policy issues",[214] proposed 'Universal Guidelines for Artificial Intelligence', which it wants incorporated at every level (in standards, national law, international agreement, and "built into the design of systems"). The aim is that the guidelines should "maximize the benefits of AI, to minimize the risk, and to ensure the protection of human rights." The guidelines consist of 12 principles and favour a risk-based approach (e.g., with a focus on public safety, cybersecurity, a required evaluation of the systems' purpose, benefits, and risks; a right to human decisions; and termination of the system if human control is not possible), with principle-based regulation (e.g., with strict prohibitions against secret profiling and government unitary scoring; and absolute requirements on human control). The principles also demand a right to transparent decisions, including the processes; a right to human decisions; required identification of the institution responsible for the system's decisions; fairness; accurate, reliable, and valid decisions; and requirements on relevant and qualitative data inputs.[215]

In 2018, Article 19 (a human rights organisation, registered as a charity, focusing on freedom of information and expression)[216], and together with Privacy International (a human rights organisation, registered as a charity, focusing on privacy)[217] released the report *Privacy and Freedom of Expression In the Age of Artificial Intelligence*.[218] The report calls for states to review "existing frameworks and regulations" for the special ethical and human rights concerns AI raise in different sectors (thus implicitly promoting sector-specific regulation), and in order to protect individuals against risk (i.e., their approach is partly risk-based), privacy and freedom of expression (which may be read as support for a principle-based approach).[219]

The report calls on both states and companies to ensure: 1) "protection of international human rights standards", which implies a principle-based approach and a need for harmonisation, given that it calls for "ensuring that laws and regulations, codes of conduct, ethical codes, and self-regulatory and technical standards meet the threshold set by international human rights"; and 2) accountability, transparency, and oversight (including required revisions of standards, regulations, and guidelines).[220]

Lastly, the report calls on civil society to, 1) further engage in the protection of fundamental rights; 2) "Collect and highlight case studies of 'human rights critical' AI"; and 3) to build coalitions and networks of expertise on AI through civil society (i.e., beyond academia and industry).

---

[213] Future of Life Institute, "Asilomar AI PRINCIPLES", op. cit.
[214] The Public Voice, "About Us". https://thepublicvoice.org/about-us/
[215] The Public Voice, "Universal Guidelines for Artificial Intelligence", 23 October 2018. https://thepublicvoice.org/ai-universal-guidelines/
[216] Article 19, "About us". https://www.article19.org/about-us/
[217] Privacy International, "About Privacy International". https://privacyinternational.org/about
[218] Privacy International and Article 19, "Privacy and Freedom of Expression in the Age of Artificial Intelligence", April 2018. https://privacyinternational.org/sites/default/files/2018-04/Privacy%20and%20Freedom%20of%20Expression%20%20In%20the%20Age%20of%20Artificial%20Intelligence.pdf
[219] Privacy International and Article 19, op. cit., p. 28.
[220] Privacy International and Article 19, op. cit., p. 28.

5.5 The public

As discussed below, various surveys show a public desire for regulation, with the strongest support for some form of co-regulation in which oversight is performed by a non-governmental agency. However, there are various potential methodological limits, such as the fact that most surveys we found surveyed residents of the USA, so there may be a cultural bias (e.g., as Zhang and Dafoe note, Americans are sceptical of public institutions[221]).

Various surveys show that a majority thinks that there is a critical need for regulation. For example, 60% of the general US population and 54% of US technology executives hold that "Regulation is critical and should be done by a public body to confirm safe development of AI", as compared to the industry self-regulating (15% and 17% respectively).[222] Other surveys support similar results. For example, there is majority support amongst Americans for both national and international regulations on artificial intelligence, and internationally for more regulations of both business and government use of AI.[223]

While a survey shows an overwhelming majority (88% general population and 94% technology executives respectively) thinks there is need for human oversight, most think that this oversight should be performed by technology companies (71/81%) as compared to the government (49/51%).[224]

As previously noted, "recent survey research suggests that while Americans feel that AI should be regulated, they are unsure who the regulators should be", which may "reflect Americans' general attitudes toward public institutions."[225] However, this uncertainty may also be an example of conflating regulations with oversight, given the previously mentioned support for regulation by a public body rather than self-regulation.

[221] Zhang, Baobao and Allan Dafoe "Artificial Intelligence: American Attitudes and Trends. Center for the Governance of AI", Future of Humanity Institute, University of Oxford, 2019. https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/

[222] Edelman, Edelman AI Survey Results Report, 2019, p. 28. https://www.edelman.com/sites/g/files/aatuss191/files/2019-03/2019_Edelman_AI_Survey_Whitepaper.pdf

[223] For national and international support, see: Morning Consult, "National Tracking Poll #170401", 30 March – 1 April 2017. https://morningconsult.com/wp-content/uploads/2017/04/170401_crosstabs_Brands_v3_AG.pdf, pp. 118-123. For international support for AI regulation, see: World Economic Forum, "Public Concern Around Use of Artificial Intelligence is Widespread, Poll Finds", World Economic Forum, 1 July 2019. https://www.weforum.org/press/2019/07/public-concern-around-use-of-artificial-intelligence-is-widespread-poll-finds

[224] Edelman, op. cit. 2019.

[225] Zhang and Dafoe, op. cit. 2019.