

# Towards Multi-Actor AI Governance

In Five Practical Steps

John Higgins

Paul MacDonnell

Global Digital Foundation

22nd October, 2020

[www.globaldigitalfoundation.org](http://www.globaldigitalfoundation.org)

<b>1.0 Introduction</b>	<b>1</b>
1.1 Preface	1
1.2 Background	1
1.3 Our approach	2
1.4 Our recommendation	3
1.5 Defining who is who in the AI regulatory and governance framework	4
<b>2.0 Why is an MAGF beneficial?</b>	<b>5</b>
<b>3.0 Step 1: Develop and document the regulatory principles of an MAGF</b>	<b>7</b>
3.1 Why the MAGF needs to rest on clear regulatory principles	7
3.2 Proposal	8
<b>4.0 Step 2: Evaluate systems for measuring trustworthy AI</b>	<b>9</b>
4.1 Introduction	9
4.2 Proposal	9
<b>5.0 Step 3 : Defining the profiles of the RGF members</b>	<b>11</b>
5.1 Introduction	11
5.2 Proposal	13
<b>6.0 Step 4 : Audit existing EU legislation affecting the use of AI</b>	<b>14</b>
6.1 Introduction	14
6.2 Proposal	14
<b>7.0 Step 5 : Draft the MAGF</b>	<b>15</b>
7.1 Introduction	15
7.2 Proposal	15
<b>Bibliography</b>	<b>17</b>

# 1.0 Introduction

## 1.1 Preface

This is a practical 5-step proposal to develop an inter-organisational governance framework (a Multi-Actor Governance Framework (MAGF)) for companies in the AI supply chain, including: chipset manufacturers, AI software developers, data providers, algorithm builders, technology integrators, cloud-service providers, and hardware / device manufacturers.<sup>1</sup>

**We are seeking industry partners to support and assist us in this work—which should be complete before 2022.**

## 1.2 Background

Fear of bad outcomes from AI technologies which are thought to be unpredictable is a recipe for political mistrust, regulatory repression, and consumer backlash. For the pessimist evidence is on offer. The recent use of an algorithm to grade public examinations in the UK, and questions about the fairness of systems used to support sentencing decisions for felons in the U.S. and Spain have already raised questions about the proper role of AI (BCS 2020; Larson J. et. al. 2016; Tolan et. al. 2019). The technology industry, other economic sectors, and policymakers and regulators acknowledge that AI technologies need protocols and, even, hard rules in both their design and implementation.

In response to anxiety about AI, bodies—including the EU, the White House, academic institutions such as Stanford and Oxford universities, industry groups such as the Global Partnership on AI, and technology companies such as Microsoft and Facebook—are working on rules and standards of ethics and safety. As with this proposal, a key objective of their work is to reduce mistrust in AI. However, a number of proposals for AI-specific regulatory and governance principles assume that AI's properties of unknowability demand new regulatory or governance paradigms—paradigms that are very different from those governing traditional organisations (Hallensleben S. et. al. 2020, p.2; European Commission, 2020 p.3; White & Case LLP 2020). This leads to proposals for regulatory and governance paradigms that are divorced from the fact that AI's use environment is often similar to that of existing digital technologies. This, in turn, leads to neglect of the need to calibrate what AI's real risks are. The assumption appears to be that 'AI risk' is analogous to 'nuclear risk'. Just as we regulate nuclear energy in a way that is distinct from non-nuclear energy so, it is suggested, AI risk presents extraordinary—even exponential—dangers that need a special approach. The net result is that authors of proposals for AI ethics intuitively place too much reliance on the precautionary principle—*what we don't know might harm us* (Kuziemski M. 2018). This proposal argues that we need to start with the fact that most uses of AI are likely to be in contexts that are sufficiently, or even exactly, similar to those of traditional technologies and decision-support systems. Hence we do not need a new precautionary

---

<sup>1</sup> We would like to thank Charles Radclyffe and Sebastien A. Krier who provided important guidance on a number of points in this proposal. Responsibility for any opinions expressed or errors made are, of course, the authors' alone.

paradigm to manage AI risk (MacDonnell P. 2020 pp.4-5; Castro D. and McLaughlin M. 2019; Turner D. and Hartzell L. 2004).

### 1.3 Our approach

Instead of intuitively relying on the precautionary principle and reaching for a new regulatory-governance paradigm, our intuition is to look to Europe's existing legal framework of product liability, social equality, and data protection law.<sup>2</sup> We propose to address the need for a governance framework for the AI supply-chain through the application of existing legal principles and on a sector or context-specific basis. Limiting the scope of our work to the AI supply-chain makes it more achievable and will enable us to provide a platform that demonstrates how appropriate AI governance should be developed in a way that is universally applicable. This will serve as an example for AI governance within AI adopters—organisations that use or will use AI. Our proposal follows a review of documents published by the following: the EU (European Commission 2020 2019a); the White House (White House Management and Budget Office 2020; see also Fitch A. 2020); OECD (2019); Global Partnership on AI (GPAI) (OECD 2020); World Economic Forum (2019); the Government of Singapore (SG:D, IMDA, PDPC 2020); the Partnership for AI (2020), the IEEE (n.d.); the Software & Information Industry Association (2017); Bertelsmann (Hallensleben S. et. al. 2020); the Centre for the Governance of AI, Oxford (Dafoe A. 2018); and the Stanford Institute for Human-Centered Artificial Intelligence (HAI n.d.).

Our review of this documentation and of the AI regulatory landscape leads us to the following conclusions. A number of initiatives that chart a path to give practical effect to AI ethics and safety principles take as their starting point that AI is qualitatively and categorically different from other technologies (Hallensleben S. et. al. 2020 p.2; European Commission, 2020 p.3; White & Case LLP, 2020). By focusing their attention on the new and unusual nature of AI technologies the authors of these initiatives often ignore the striking similarities between the use of AI to the use of other technologies. For the most part AI is simply another—albeit cheaper and faster—way to get from A to B. Hence our starting assumption is that there should be no need to invent a new regulatory paradigm that focuses on the technology itself. Legislators and regulators are, rightly, concerned with the application of technologies and its implications for fairness and safety. But to focus regulatory attention on the technology itself is to unconsciously acquiesce to the anthropomorphic idea that AI is truly 'intelligent'. However, people who do bad things with AI can be just as easily held to account as people who do bad things with spreadsheets or, for that matter, pen and paper. That's why our starting assumption is that, from a governance and regulatory perspective, AI is, for the most part, similar to other technologies. We conclude, therefore, that:

1. The current and foreseeable dangers arising from AI mostly overlap with the dangers of other organisational decision-making—whether supported with ordinary digital technologies such as spreadsheets or, even, a pen and paper.

---

<sup>2</sup> See Bibliography, Legislation.

2. The EU's existing product liability, equality, and data protection legislation is more than adequate to support at least a foundational governance framework for interactions *between* AI development companies.
3. Because the U.S (*Wall Street Journal* 2020) is all but certain to substantially align AI rules to existing legislation on product liability, equality, and data protection, the EU is likely to take a similar course with regard to current EU legislation. In other words, legislation, when it comes, will, for the most part, constitute the provision for new contexts in which established and known legal principles are upheld.

**Our starting point, therefore, is that—with few, context-dependent, exceptions—AI technology will be little different, for regulatory purposes, from other digital technologies. AI could eventually require what future historians might call a ‘paradigm shift’ in our view of economic, and regulatory policy (Kuhn T.S. 1962). But to attempt to create an anticipatory new paradigm in regulatory or governance structures for AI now is both impractical and unnecessary.**

We do not deny the likelihood that rules governing the specific use of certain AIs will be necessary. But such rules will likely always be context-dependent.

The starting point for the proper regulation of AI, therefore, is law and regulations that already exist. And if the starting point for the regulation of AI is law and regulations that already exist, then the starting point for an MAGF must also be law and regulations that already exist.

## 1.4 Our recommendation

In contrast to proposals for AI governance within organisations, we propose an MAGF—that is, a framework to operate *between* rather than within the organisations that make up the AI supply-chain.

**Our programme will support the development, adoption and implementation of an MAGF in the form of common standards—notably standards of safety and ethics aligned with the EU's Ethics Guidelines for Trustworthy AI—across the AI supply-chain (European Commission 2019a).**

The programme comprises five components which will lead to the development of an MAGF. These are:

1. Develop and document the regulatory principles of an MAGF.
2. Evaluate systems for measuring trustworthy AI.
3. Define and distinguish the profiles of the individual .
4. Audit existing EU legislation affecting the use of AI.
5. Draft the MAGF.

## 1.5 Defining who is who in the AI regulatory and governance framework

The MAGF is a key part of a much broader regulatory and governance framework which we call the *AI Regulatory and Governance Framework* (RGF). The RGF comprises all organisations and individuals with any material interest in, ability to have an effect on, or any propensity to be affected by, AI. The RGF's members are:

<b>Regulatory and Governance Framework (RGF) Members</b>	
<i>The AI Supply-Chain</i> (organisations / MAGF actors)	<p>Chipset manufacturers, AI developers, Data providers, Algorithm builders, Technology integrators, Cloud-service providers, Hardware / device manufacturers.</p> <p>They may be members of consortia and associations considering AI governance.</p> <p><b>These will be the co-creators and actors within the Multi-Actor Governance Framework.</b></p>
<i>AI adopters</i> (organisations)	<p>Organisations that make use of AI technologies to conduct their business. They include: B2B, B2C, Healthcare providers, Police, Education bodies, and Government service providers (e.g. local government water supply services).</p> <p>They may be members of consortia and associations considering AI governance.</p> <p>(Some of them may also be members of the AI supply-chain, i.e. actors of an MAGF)</p>
<i>Professional users</i> (individuals)	<p>Individuals who use AI technologies for any professional purpose, e.g. Professional users of tools such as AI ethics tools. They include developers and creators of algorithms and datasets.</p> <p>NB. Users can be found anywhere in the <i>AI Supply-Chain</i> or amongst <i>AI Adopters</i>.</p>
<i>Consumers / Citizens</i> (individuals and groups)	<p>Private individuals who are domestic consumers of any service. In their relation to the business of public administration they are described as <i>citizens</i> and in relation to services provided by any organisation they are described as <i>consumers</i>. This group also includes civil society organisations.</p>
<i>Legislators / Policymakers</i> (organisations and individuals)	<p>Elected officials, government departments, members of revising / 2nd legislative chambers, permanent officials (civil servants and advisors).</p>
<i>Regulators</i> (organisations)	<p>Regulatory organisations with supervisory authority over any organisation or individual who develops or uses AI technologies.</p>

## 2.0 Why is an MAGF beneficial?

A world without an MAGF will present increasing barriers to trade and cooperation. This is because the AI supply-chain for most enterprises is likely to include actors based in multiple jurisdictions. Not only will each jurisdiction's overall consideration of AI regulation and governance proceed at different speeds, its evaluation of the urgency of—and therefore its emphasis on—the distinct risks within AI (e.g. loss of privacy, bias, and physical safety) will differ from other jurisdictions. The result is likely to be an heterogeneous global regulatory environment that will generate considerable legal uncertainty for enterprises within the AI supply-chain, making it more difficult to bridge the valleys of trust between regulators, AI supply-chain actors, and consumers.

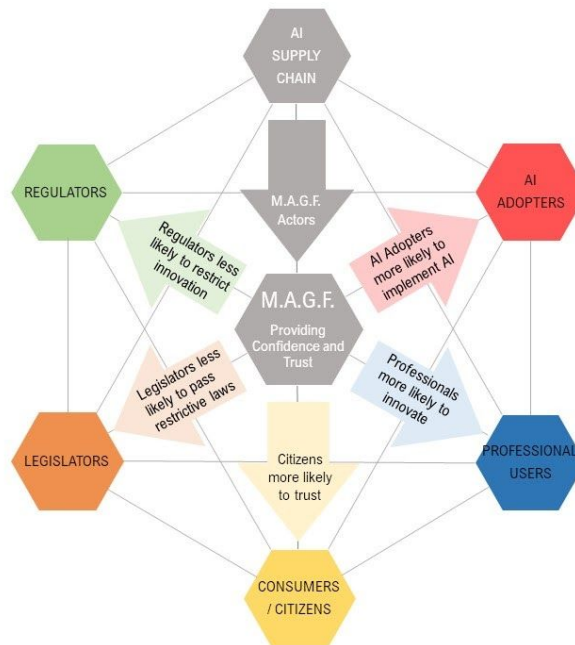
So that it can be developed and used to its full potential it will be necessary for any jurisdiction to eliminate uncertainty about AI's legal and regulatory status within different contexts. For example, the application of privacy regulations in the UK could prevent the export of important medical AI technology from Germany. A law requiring *ex ante* validation that a dataset reflect particular demographic and ethnic characteristics could prevent the import of technology to support the marketing of financial services products in France. Without an MAGF comprising multiple actors in multiple jurisdictions different jurisdictions are more likely to make new rules aimed at local *uncertainty* about AI. In other words without an MAGF individual jurisdictions are more likely to conclude that they are operating in an AI 'wild west' and, thus, adopt the precautionary principle (European Parliament : Think Tank 2015). This view has already been expressed in China (Kuziemski M. 2018). The EC has already adopted the precautionary principle in ruling on such issues as GM foods, the environment, and in its white paper on AI where it proposes regulating all AI as 'high risk' (European Commission 2017; European Commission 2020; MacDonnell P. 2020; CEC European Managers 2019; and Chivot E. 2020).<sup>3</sup> While it is constructive to view this as a natural characteristic of early thinking about a subject where risks are not well understood we should bear in mind that the EU has, since its foundation, believed in the virtues of the precautionary principle, in particular with regard to environmental protection (Treaty on the Functioning of the EU, Article 191(2)1957). The problem with applying the precautionary principle to AI is that it places a potentially infinite value on avoiding all harm while shifting the entire burden of proving that AI is safe onto AI developers and users. This is too heavy a burden for any innovative technology to bear.

If actors within the AI supply-chain agree a framework that applies existing or expected rules concerning ethics and safety to AI then national regulators will develop a better sense of AI risk as an *operational* matter that is, therefore, more easy to quantify. This will lessen their temptation to apply the precautionary principle and increase the likelihood that any new AI rules will be no more than is necessary to protect consumers and citizens.

---

<sup>3</sup> Though in applying the precautionary principle to AI the EU's white paper appears to be addressing what it sees as perception of AI risk rather than its own evaluation of AI's real risks.

The MAGF can be understood as the centrepiece of the RGF. And the RGF can, in turn, be understood as a multi-sided platform with each of its members forming one of the six sides (Evans D.S. Schmalensee R. 2017). The RGF platform's six sides are, therefore: The AI Supply-Chain (MAGF actors); AI Adopters; Professional Users; Consumers / Citizens; Legislators / Policymakers; Regulators.



On a multi-sided platform benefits to one side increase its economic / social value to each of the other sides. The benefits of the RGF platform accrue in the form of *trust and confidence* between its members, the increase of which promotes the maximisation of the development and application of AI technologies. Each of the six sides benefits the others in the following ways. 1) The AI supply chain (i.e. MAGF actors) apply a governance framework to AI which gives confidence and promotes trust in AI technology among the other five sides of the platform; 2) AI adopters are more likely to implement AI if they have assurances about its efficacy and its ethical standing; 3) Professional users are more likely to innovate and experiment with AI technologies if they are assured that they are doing so within a governance framework that provides guard-rails preventing harm; 4) Consumers and citizens are more likely to trust AI as technology that will enhance their lives if they know that high standards of protection are being applied to its development and application, and that they have avenues for redress; 5) Legislators are less likely to enact punitive legislation if they see that the public has justified confidence in AI; and 6) Regulators are less likely to impose damaging restrictions on the development and application of AI if they believe that members of the AI supply-chain (MAGF actors) are adhering to high ethical and safety standards.



## 3.0 Step 1: Develop and document the regulatory principles of an MAGF

### 3.1 Why the MAGF needs to rest on clear regulatory principles

In order for the MAGF to be viable—that is to say, attractive to actors in the AI supply-chain, and credible to policymakers and regulators—it will need to rest on the following regulatory principles:

- A. The MAGF should be grounded on the principle that the development and application of AI governance should apply *sufficient* standards to be effective. In other words, if existing legislation and regulations are sufficient to meet the EU’s ethical principles, then the MAGF should not, in-effect, gold-plate them by going beyond what they require. That legislation and regulations should not be gold-plated has been articulated and is supported by the EU (Atthoff K. and Wallgren M. 2012; European Commission 2019b). This principle concerns the proper economic management of risks and it should also apply within governance environments.
- B. The MAGF should not duplicate but, rather, should adopt and apply *existing* regulatory injunctions aimed at protecting the safety and rights of citizens and consumers which are in force or due to come into force within each actor’s jurisdiction.<sup>4</sup> In other words where existing legislation and regulations are fit for purpose the MAGF should not attempt to ‘reinvent the wheel’ but should treat its rules as the implementation of this, already existing, legislation and regulations.
- C. The MAGF should distinguish between: 1) risks posed by a malfunctioning AI; 2) risks created or aggravated by the use of properly-working AI; and 3) AI decisions or recommendations made by a properly-working AI that may look unfair in themselves but which reflect underlying social inequalities. The MAGF needs to reflect awareness that, while 3) is a matter for legitimate intervention by elected legislators, poorly-designed unilateral actions by its actors to address social inequalities could cause unintended consequences leading, for example, to the unjust treatment of individuals.

---

<sup>4</sup> ‘Consumers’ in this document means domestic consumers of products and services that use AI. For examples of legislation in force in the EU see the bibliography.

## 3.2 Proposal

Based on the above we will:

1. Circulate draft principles of, and proposals for, an MAGF to representatives of the global AI community for discussion and feedback;
2. Directly engage with regulatory professionals, governance and legal specialists to evaluate;
3. Host seminars and workshops for RGF members to discuss.<sup>5</sup>
4. Revise and finalise these principles as a first chapter in the final MAGF proposal.

---

<sup>5</sup> Regulators' independence prevents us from inviting them to these meetings.

## 4.0 Step 2 : Evaluate systems for measuring trustworthy AI

### 4.1 Introduction

In addition to considering and determining its regulatory principles an MAGF could also benefit from an agreed risk-indication system that could apply to AI technologies. As the currency of the MAGF will be trust and confidence such a shared system of risk identification and indication could enable its actors to share their understanding of risk *probability* and, should that risk materialise, its *impact*.

A risk-indication system could:

1. Support MAGF members' adherence to their legal and regulatory obligations, in particular it can support transparency obligations by indicating the likelihood and impact of AI risk in relations to specific components and deployment scenarios for AI.
2. Build trust amongst RGF members.

A number of system types and approaches are available for consideration. These include risk labelling, ratings systems, and independent auditing. Such systems could be implemented singly or together. However we do not believe that more is, necessarily, better. The ultimate purpose of a risk-indication system is to enhance the effectiveness of MAGF members' roles as managers of AI risk and as builders of trust. Some system that indicates risk is likely desirable.

### 4.2 Proposal

Based on the above we will:

1. Review literature on risk labelling, ratings systems, and independent auditing.
2. Assess existing ratings systems (e.g. those used in financial services, or chemical-plant safety) whose features may bear an isomorphic relationship to risks in the AI supply-chain. In particular we will review literature on risk labelling, ratings systems, and independent auditing—identifying the possible strengths and weaknesses of each in the context of AI development and use.
3. Author a proposal that recommends a specific risk indicator system for an MAGF.
4. Circulate this proposal to representatives of the global AI community for discussion and feedback;

5. Directly engage with regulatory professionals, insurance experts, governance and legal specialists to evaluate (for the potential contribution of insurance experts see Shankar R. et. al. 2020);
6. Host seminars and workshops for RGF members to discuss.<sup>6</sup>
7. Revise and finalise options for an implementation path as the 2nd chapter in the final MAGF proposal.

---

<sup>6</sup> Regulators' independence prevents us from inviting them to these meetings.

## 5.0 Step 3 : Define the profiles of the RGF members

### 5.1 Introduction

RGF members are MAGF actors, AI adopters, Professional users, Consumers / citizens, Legislators / Policymakers, and Regulators. Each of these has a particular profile that comprises:

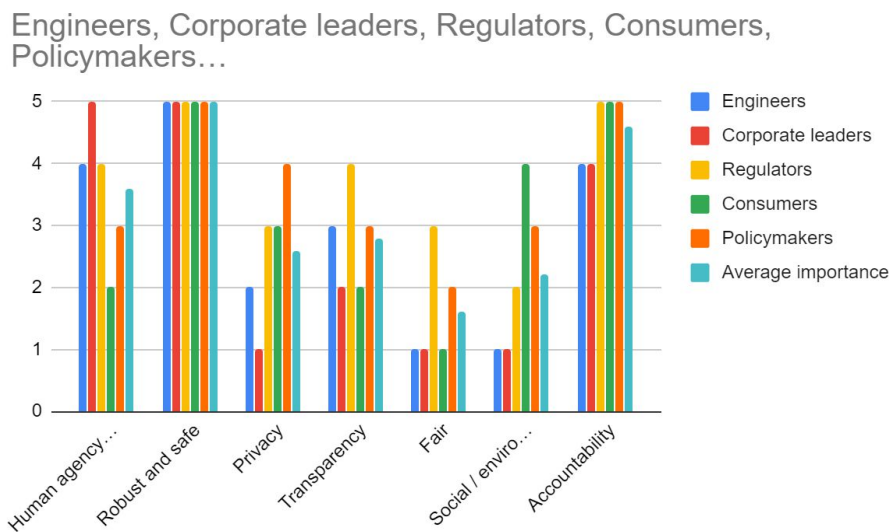
1. Their role-specific objectives in interacting with other members of the RGF and
2. The non-role related incentives that influence them as they interact
3. Their level of information about the other RGF members with whom they interact

If we take, for example, the different responses to driverless cars we will likely see that different RGF members accord different priorities to values such as privacy and safety.

A hypothetical analysis of the RGF members' ethical considerations when thinking about driverless cars could look something like this.

#### What are the important ethical considerations in AI for driverless cars?

1-5 from least to most important : a hypothetical illustration.

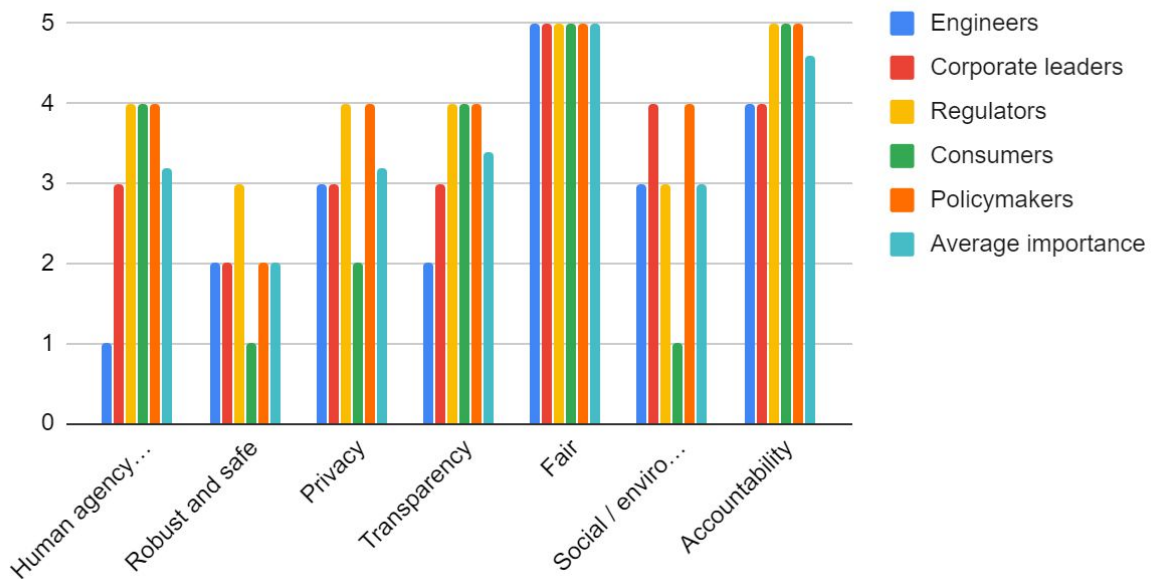


Whereas their response to AI that supports job recruitment could look something like this.

### What are the important ethical considerations in AI for recruitment?

1-5 from least to most important : a hypothetical illustration.

Engineers, Corporate leaders, Regulators, Consumers, Policymakers...



Non role-related incentives are also likely to be a factor in the responses of members of the RGF. These include both economic and informational considerations.

Hence we propose a market-for-regulation perspective of the RGF to indicate that not all incentives to regulate are made with consumer / citizen protection in mind. It is well known that consumers and businesses organisations can be influenced by economic, political, and psychological factors that may have a negative effect on parties with whom they interact. Furthermore policymakers and regulators are also subject to economic, political, and psychological pressures and incentives (Stigler 1985; Peltzman 1976, 1989; Becker 1983; McChesney 1987).

For example members of the RGF could be subject to influences as follows :

- A company may seek regulation of its industry to create barriers to entry to its market, thus preventing competition.
- Legislators and policymakers may take advantage of consumers' anxiety to gain political benefits by overstating the risks of AI.

- Consumers may overestimate the risks of AI due to asymmetry of information between them and other members of the RGF and / or basic anxiety about the unknown and, hence, they may be reluctant to trust AI (Akerlof 1970).

## 5.2 Proposal

We propose constructing an indicative profile of each member of the RGF that will enable us to build out detailed governance chapters in the MAGF including:

1. A summary of their goals when interacting with other members of the RGF.
2. A summary of how they are likely to view the risks inherent within transactions in the RGF.
3. A summary of non role-specific incentives that are likely to influence their motivations in interaction with other members of the RGF.
4. A summary of their expected governance obligations under an MAGF.

We will:

4. Circulate this summary to representatives of the global AI community for discussion and feedback;
5. Directly engage with regulatory professionals, and economics experts to evaluate;
6. Host workshops for RGF members to discuss.<sup>7</sup>
7. Revise and finalise for insertion as the 3rd chapter in the final MAGF proposal.

---

<sup>7</sup> Regulators' independence prevents us from inviting them to these discussions.

## 6.0 Step 4 : Audit existing EU legislation affecting the use of AI

### 6.1 Introduction

The EU white paper, together with much thinking by think tanks, academic institutions, and industry consortia has focused on the ‘uniqueness’ of AI in a way that has distracted from the fact that most AI risk can be addressed through existing EU legislation.

The EU’s record in implementing, over past decades, an extensive body of legislation, covering product liability, equality, and privacy has left a considerable legal foundation upon which to begin constructing an MAGF—even in the absence of specific legislation on the development and use of AI technologies (Access Now 2018).

The next year presents an opportunity to set the agenda by beginning the construction of an MAGF along the lines of in-force product, privacy and equality legislation.

### 6.2 Proposal

1. We will conduct an audit of existing and proposed EU-level regulations and legislation (both human rights and product-safety/liability) that applies to AI. We will examine work already undertaken and take into account fresh EU thinking to build our understanding of the emerging regulatory environment.
2. We will submit this audit to regulatory-legal experts for review.
3. We will finalise this audit and it will form Chapter 4 of the final MAGF proposal.



## 7.0 Step 5 : Draft the MAGF

### 7.1 Introduction

In the fifth and final chapter of the MAGF document we will evaluate different models and stakeholder groupings, including supply chain ecosystems and use / application cases. We will identify organisations which could act as conveners for different groupings.

Guided by the Regulatory Principles we will draw on white papers, statements of principles, and other published documents to outline the working processes of a MAGF. Beginning with an AI ethics charter, the framework will describe the decision processes to be applied when an organisation interacts with another organisation at different levels and in different contexts.

The AI Ethics Charter will be a synthesis of the EU's *Ethics Guidelines* (European Commission 2019a) and the regulatory principles that will be the output of 3.0 above. We envisage that the framework will be confirmed / checked by use of the EU's ALTAI assessment [tool](#). In other words, the level of acceptance in the MAGF proposal will be informed by the output of 3.0 above while the ALTAI tool will be used to confirm it's adherence to the EU ethics guidelines. Hence the tool will be indicative only.

### 7.2 Proposal

We will:

1. Identify the governance responsibilities of different business types and individuals within the multi-actor model.
2. Identify the internal participation in a governance model that is appropriate for each type of business (e.g. a business that is primarily interested in content, data, and marketing will have different departmental governance considerations to one which is concerned with hardware manufacturing).
3. We will identify practical considerations that need to be addressed in a company's internal governance structure that will help make the MAGF work. Furthermore, we will draw out differences in such structures applicable to different types of business organisations.
4. Finalise and publish the framework.
5. We will circulate the draft MAGF to a specialist panel of regulatory and legal experts for discussion.

6. Following feedback we will finalise the MAGF and use it as a tool with which to engage with representatives of organisations participating in existing governance initiatives, as well as representatives from the EU, the GPAI, think tanks, trade associations, consumer groups, and leading technology and AI-user corporations.
7. In particular, we propose to host, in partnership with DIGITALEUROPE's member associations, panel discussions on our global framework in European capitals, including: Paris, Berlin, Madrid, London, Oslo, Stockholm, Copenhagen, The Hague, Helsinki, Dublin, and Athens. We will invite participants in groups concerned with global standards, national and EU policymakers, policy-analysts, regulators, and corporate representatives, to these discussions.

End

## About Global Digital Foundation

Global Digital Foundation is a Europe-based think tank that offers policymakers a framework for dialogue with stakeholders and with their counterparts in other countries so that they can develop a shared understanding of the challenges of policy in a digital world. Global Digital Foundation is also an international network of policy expertise providing a base for interdisciplinary research and a forum for three-way learning interaction between policymakers, stakeholders and scholars.

## Contacts

Executive Director: Paul MacDonnell: +44 (0) 7534 580976; Email: [paul.macdonnell@globaldigitalfoundation.org](mailto:paul.macdonnell@globaldigitalfoundation.org).

Chairman: John Higgins: + 44 (0) 7966 136638; Email: [john.higgins@globaldigitalfoundation.org](mailto:john.higgins@globaldigitalfoundation.org).

# Bibliography

## Articles, Books, Reports, Papers, and Press Releases

Access Now (2018), *Mapping Regulatory Proposals for AI in Europe*, Vodafone Institute for Society and Communications, November, 2018, available at:  
[https://www.accessnow.org/cms/assets/uploads/2018/11/mapping\\_regulatory\\_proposals\\_for\\_AI\\_in\\_EU.pdf](https://www.accessnow.org/cms/assets/uploads/2018/11/mapping_regulatory_proposals_for_AI_in_EU.pdf).

AI4People (2018), 'AI4People: responsible use of artificial intelligence', video, 11th September, 2018. Available:  
<https://www.audi-mediacycenter.com/en/audiomediacycenter/video/ai4people-responsible-use-of-artificial-intelligence-4343>.

Akerlof G. A. (1970), 'The Market for "Lemons": Quality Uncertainty and the Market Mechanism,' *The Quarterly Journal of Economics*, Oxford University Press, vol. 84(3), pages 488-500. Available: <https://ideas.repec.org/a/oup/qjecon/v84y1970i3p488-500..html>.

Atthoff K. and Wallgren M. (2012), *Clarifying Gold-Plating: Better Implementation of EU Legislation*, Stockholm, Swedish Better Regulation Council. Available:  
<https://www.regelradet.se/wp-content/uploads/2012/11/Clarifying-Gold-Plating2.pdf>.

Becker G. S (1983), 'A Theory of Competition Among Pressure Groups for Political Influence', *Quarterly Journal of Economics*, 98, August, 371-400

BCS (2020), *The Exam Question: How do we Make Algorithms do the Right Thing?* Swindon, BCS. Available: <https://www.bcs.org/media/6135/algorithms-report-2020.pdf>.

Castro D. and McLaughlin M. (2019). 'Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence', ITIF, 4th February, 2019. Available: [shorturl.at/IKOT4](http://shorturl.at/IKOT4).

CEC European Managers (2019) 'AI Ethics guidelines, Draft CEC feedback on the draft proposal of the High-Level Expert Group on Artificial Intelligence', CEC website, January, 2019. Available at:  
<https://www.cec-managers.org/wp-content/uploads/2019/01/CEC-contribution-to-AI-ethical-guidelines.pdf>.

Chivot E. (2020), 'Response to the European Commission's Consultation on the White Paper on Artificial Intelligence', Washington DC: ITIF, 14th June, 2020. Available:  
<http://www2.datainnovation.org/2020-eu-ai-whitepaper-response.pdf>.

Dafoe A. (2018), 'AI Governance: A Research Agenda', Oxford, Future of Humanity Institute. Available: <https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf>.

European Commission (2017), Future Brief: The precautionary principle: decision-making under uncertainty, in *Science for Environment Policy*, September 2017, Issue 18. Available: [https://ec.europa.eu/environment/integration/research/newsalert/pdf/precautionary\\_principle\\_decision\\_making\\_under\\_uncertainty\\_FB18\\_en.pdf](https://ec.europa.eu/environment/integration/research/newsalert/pdf/precautionary_principle_decision_making_under_uncertainty_FB18_en.pdf).

European Commission (2019a), *Ethics Guidelines for Trustworthy AI*, High-Level Expert Group on Artificial Intelligence (2019), Brussels, European Commission, Available: <https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf>.

European Commission (2019b), 'Better Regulation principles: at the heart of the EU's decision-making process', Press Release, Brussels, European Commission, 15th April, 2019. Available: [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_19\\_2117](https://ec.europa.eu/commission/presscorner/detail/en/IP_19_2117).

European Commission (2020), *White Paper on Artificial Intelligence: a European approach to excellence and trust*, European Commission, 19th February, 2020.

European Parliament : Think Tank (2015), The precautionary principle: Definitions, applications and governance, website. Available: [https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS\\_IDA\(2015\)573876](https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_IDA(2015)573876).

Evans D.S., Schmalensee R. (2017) Multi-sided Platforms. In: Palgrave Macmillan (eds), *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, London. [https://doi.org/10.1057/978-1-349-95121-5\\_3069-1](https://doi.org/10.1057/978-1-349-95121-5_3069-1). Available: [https://link.springer.com/referenceworkentry/10.1057%2F978-1-349-95121-5\\_3069-1](https://link.springer.com/referenceworkentry/10.1057%2F978-1-349-95121-5_3069-1).

Facebook (Facebook (2020), Collaborating on the future of AI governance in the EU and around the world, webpage, 15th June, 2020. Available: <https://ai.facebook.com/blog/collaborating-on-the-future-of-ai-governance-in-the-eu-and-around-the-world/>

Fitch A. (2020), White House Nears New Rules on Artificial Intelligence, *Wall Street Journal*, 21st October, 2020. Available: <https://www.wsj.com/articles/white-house-nears-new-rules-on-artificial-intelligence-11603300366>.

Hallensleben S. et. al. (2020), *From Principles to Practice : an interdisciplinary framework to operationalise AI ethics*, VDE, Bertelsmann Stiftung.

IEEE (n.d.), 'The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems'. Available: szcHD92zszSE970b19GKVP2fl20-fOc.

Kuziemski M. (2018), 'We need a precautionary approach to AI', *China Daily*, 8th May, 2018.

Kuhn T.S. (1962), *The Structure of Scientific Revolutions*, 4th ed. Chicago: University of Chicago Press.

Larson J. et. al. (2016), 'How We Analyzed the COMPAS Recidivism Algorithm', *ProPublica*, 23rd May, 2016, available:  
<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>

MacDonnell P. (2020), *Making Rules for AI: The errors and fallacies regulators must avoid*, Brussels: Global Digital Foundation. Available: <http://www.globaldigitalfoundation.org/>.

McChesney F. S. (1987) 'Rent Extraction and Rent Creation in the Economic Theory of Regulation', *Journal of Legal Studies*, 16, January, pp. 101-118.

OECD (2019), 'Artificial intelligence > OECD Principles on AI', webpage. Available: <https://www.oecd.org/going-digital/ai/principles/>.

OECD (2020) 'OECD to host Secretariat of new Global Partnership on Artificial Intelligence', website, 15th June, 2020. Available:  
<https://www.oecd.org/going-digital/ai/oecd-to-host-secretariat-of-new-global-partnership-on-artificial-intelligence.htm>.

PAI (2020), webpage, updated 14th October 2020. Available:  
<https://www.partnershiponai.org/>.

Peltzman (1976), 'Towards a More General Theory of Regulation', *Journal of Law and Economics*, 19, August, pp. 211-40.

Peltzman S..(1989), 'The Economic Theory of Regulation After a Decade of Deregulation', in Martin Neal Baily and Clifford Winston (eds), *The Brookings Papers on Economic Activity*, Washington, Brookings Institution, pp. 1-41.

Responsible AI (n.d.), Microsoft. Available:  
<https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6>.

SG:D, IMDA, PDPC (2020), *Model Artificial Governance Framework*, 2nd ed., paper published at the 2020 World Economic Forum Annual Meeting, Davos, Switzerland.

Shankar R., Kumar S. and Nagle F. (2020), 'The Case for AI Insurance', *Harvard Business Review*, 29th April, 2020. Available: <https://hbr.org/2020/04/the-case-for-ai-insurance>.

SSIA (2017), 'SIIA Releases Ethical Principles for Artificial Intelligence and Data Analytics, with Support from the Future of Privacy Forum, and the Information Accountability Foundation', Press Release, 7th November, 2017. Available:  
<https://www.siiia.net/Press/SIIA-Releases-Ethical-Principles-for-Artificial-Intelligence-and-Data-Analytics-with-Support-from-the-Future-of-Privacy-Forum-and-the-Information-Accountability-Foundation>.

Stanford Institute for Human-Centered Artificial Intelligence (n.d.) webpage, Stanford University. Available: <https://hai.stanford.edu/>.

Stigler G. J. (1985), 'The Origin of the Sherman Act', *Journal of Legal Studies*, 14, 1-12.  
Tencent Research Institute (2020), 'Tencent AI Whitepaper', trans. Caroline Meinhardt and Jeffrey Ding. Google Drive. Available at [https://docs.google.com/document/d/1IRHsKWGmu-dJuH2C8t4kjkeeD\\_StNYxynx0-zHxzHqo/edit](https://docs.google.com/document/d/1IRHsKWGmu-dJuH2C8t4kjkeeD_StNYxynx0-zHxzHqo/edit).

Tolan et. al. (2019), 'Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia', International Conference on AI and Law, Montreal, QC, Canada, 17–21 June, 2019, available: [https://chato.cl/papers/miron\\_tolan\\_gomez\\_castillo\\_2019\\_machine\\_learning\\_risk\\_assessment\\_savry.pdf](https://chato.cl/papers/miron_tolan_gomez_castillo_2019_machine_learning_risk_assessment_savry.pdf).

Turner D. and Hartzell L. (2004). 'The Lack of Clarity in the Precautionary Principle', *Environmental Values*. Vol. 13. No. 4. November, 2004, pp. 449-460. Available at: <https://www.jstor.org/stable/30302021>.

Wall Street Journal (2020) 'White House to Release New AI Regulations', video, *Wall Street Journal*, 21st October, 2020. Available: <https://www.wsj.com/video/white-house-to-release-new-ai-regulations/1727C02D-FC00-4C92-BCCF-0C4BC0CC35B3.html>.

White & Case LLP (2020). 'Regulation of Artificial Intelligence in Europe and Japan'. webpage, lexology.com, 24th August, 2020. Available: <https://www.lexology.com/library/detail.aspx?g=3f873d98-cf3b-4fe9-b26e-39fb74dfbf8e>.

White House Management and Budget Office (2020), *Request for Comments on a Draft Memorandum to the Heads of Executive Departments and Agencies, "Guidance for Regulation of Artificial Intelligence Applications"*, Federal Register, 13th January, 2020. Available: <https://www.federalregister.gov/documents/2020/01/13/2020-00261/request-for-comments-on-a-draft-memorandum-to-the-heads-of-executive-departments-and-agencies>.

World Economic Forum (2019), *Responsible Use of Technology*, Geneva, World Economic Forum, August, 2019. Available: [http://www3.weforum.org/docs/WEF\\_Responsible\\_Use\\_of\\_Technology.pdf](http://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology.pdf).

## Legislation

Council of the European Union (1985), *Directive on Product Liability*, Directive 85/374/EEC, 25th July, 1985, available:

<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31985L0374>.

Council of the European Union (2000), *The Employment Equality Directive* (2000/78/EC).

Council of the European Union (2000), *The Racial Equality Directive* (2000/43/EC).

Council of the European Union (2001), *Directive on General Product Safety*, Directive 2001/95/EC, 3 December 2001, available:

<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32001L0095>.

Council of the European Union (2006), *Directive on Machinery*, Directive 2006/42/EC, 17th May 2006, and amending Directive Directive 95/16/EC (recast), available:

<https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1579870568589&uri=CELEX:02006L0042-20190726>.

Council of the European Union (2014), *Directive on Radio Equipment* (2014/53/EU), 14th April, 2014, available:

<https://eur-lex.europa.eu/search.html?qid=1579875189091&text=Radio%20equipment%20directive&scope=EURLEX&type=quick&lang=en>.

European Commission (2016), *General Data Protection Regulation (GDPR)*, Official Journal of the European Union, 27th April, 2016, Brussels, available: <https://gdpr-info.eu/>.

Treaty on the Functioning of the EU (1957). Available:

<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012E/TXT&from=EN>.