

## **The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence**

*A continuous journey towards an appropriate governance framework for AI*

*Nathalie A. Smuha\**

*As part of its European strategy for Artificial Intelligence (AI), and as a response to the increasing ethical questions raised by this technology, the European Commission established an independent High-Level Expert Group on Artificial Intelligence (AI HLEG) in June 2018. The group was tasked to draft two deliverables: AI Ethics Guidelines and Policy and Investment Recommendations. Nine months later, its first deliverable was published, putting forward a comprehensive framework to achieve “Trustworthy AI” by offering ethical guidance to AI practitioners. This paper dives into the work carried out by the group, focusing in particular on its AI Ethics Guidelines. First, this paper clarifies the context that led to the creation of the AI HLEG and its mandate (I). Subsequently, it elaborates on the Guidelines’ aim and purpose (II), and analyses the Guidelines’ drafting process (III). Particular focus is given to the questions surrounding the respective role played by ethics and law in the AI governance landscape (IV), as well as some of the challenges that had to be overcome throughout the process (V). Finally, this paper places the Guidelines in an international context, and sets out the next steps (VI) ahead on the journey towards an appropriate governance framework for AI (VII).*

### **I. Introduction**

Artificial Intelligence (AI) systems are becoming increasingly widespread. From the spam filter in our inboxes, to the recommender system in our music apps; from the navigation systems on our smart phones to our daily-visited social media platforms; and from the headlines we read in newspapers to the small talk we have with colleagues. It can even be found in the descriptions of start-ups that do not appear to use the technology at all.<sup>1</sup>

---

\* Lawyer, Researcher and Assistant Lecturer at the KU Leuven Faculty of Law (beneficiary of the FWO - Flanders Research Foundation Grant). The author was the coordinator of the High-Level Expert Group on Artificial Intelligence at the European Commission, DG Connect. The author writes in her own name. Correspondence address: [nathalie.smuha@kuleuven.be](mailto:nathalie.smuha@kuleuven.be).

<sup>1</sup> A survey conducted by MMC Ventures, a London-based venture capital firm, indicated that for 40% of European start-ups that are classified as AI companies, no evidence of the use of AI can be found. The report is accessible at: <https://www.mmcventures.com/wp-content/uploads/2019/02/The-State-of-AI-2019-Divergence.pdf>

AI systems – comprising a range of different yet related applications<sup>2</sup> – are not a new phenomenon.<sup>3</sup> However, the scale on which they are being applied has reached an unprecedented level,<sup>4</sup> owing to advances in research, easier access to computing power and the ever-wider availability of big data. AI systems are being used in numerous domains of our lives, optimising processes not only in the public or professional sphere, but also in the private sphere. Advances in the field of AI have not merely increased its attractiveness and pervasiveness, but have also led to a revival of ethical concerns around AI systems. Whereas AI ethics – a subfield of applied ethics – was once a rather obscure field of study for sci-fi loving philosophers, it has now become virtually mainstreamed, occupying the mind of the public at large.

## 1. The Revival of AI Ethics

It is becoming ever more clear that, contrary to the hope some may hold, AI systems are not a miracle solution to all our problems. Rather, much like any type of technology, they represent a double-edged sword that can be used both in a beneficial and in a harming way. To offer but a few examples; AI systems can be used to help us achieving more objective decision-making, yet they can also perpetuate and even exacerbate unfair biases. AI systems can be used to offer us more personalised and qualitative services, yet they can also hamper our decisional autonomy. And while AI systems can be used to enhance our security, they can also be deployed as a means of illegitimate surveillance and stifle our freedom.

The multifaceted impact that AI systems (along with other modern ICTs) have on our lives – which is not only ethical, but also legal, social, economical, political, cultural and psychological in nature – enables those systems to shift more than one paradigm in our society. The transformative force they exert on the world may not be as brisk and theatrical as some movies portray, but is rather expected to generate changes in more slow and subtle ways, which nevertheless have the potential to alter our habits, processes and lives profoundly.

Governments and policy-makers around the world are starting to wake up to these important concerns. Besides adopting national strategies to boost the development and uptake of AI systems to reap their benefits, they are also assessing the technology’s potentially harmful impact and exploring which policy measures are best suited to address it. This occurs not only at national, but also at European level. Indeed, given that the concerns raised by AI do not stop at national borders, coordinated action is needed to tackle them.

---

<sup>2</sup> In this regard, see *A definition of AI – Main Capabilities and Disciplines*, published by the High-Level Expert Group on AI on 8 April 2019, accessible at: <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>.

<sup>3</sup> The term Artificial Intelligence was famously coined in 1956 at a workshop held in Dartmouth College. Since then, the field of research of AI has known several waves of success, and consecutive “winters” during which the interest and investment in AI research declined. For a historical overview, see for instance N. J. Nilsson, *The Quest for Artificial Intelligence*, Cambridge University Press, 2009.

<sup>4</sup> See for instance “Notes from the AI Frontier: Modelling the Impact of AI on the World Economy”, McKinsey Global Institute (2018); “Artificial Intelligence in Europe”, Outlook for 2019 and Beyond”, EY (2018); PwC’s “Global Artificial Intelligence Study: Exploiting the AI Revolution” (2017), and “Artificial Intelligence Innovation Report,” Deloitte (2018).

## 2. Europe's Response

In April 2018, following an invitation by the European Council to put forward a European approach to AI<sup>5</sup>, the European Commission presented its strategy for AI in the Commission Communication “Artificial Intelligence for Europe”.<sup>6</sup> The strategy evolves around three pillars, aiming to tackle AI's challenges and opportunities through a holistic approach: (1) boosting investments in AI, (2) preparing for socio-economic changes and (3) ensuring an adequate ethical and legal framework. At the same time, the strategy emphasised the need for coordination with Member States, so as to leverage joint efforts and maximise the impact of individual states' action. This resulted in the adoption of a second Communication in December 2018, containing a Coordinated Plan on AI that sets out around 70 actions that Member States and the Commission have committed to undertake in order to implement the European strategy.<sup>7</sup> To ensure continuity of action, the Plan will be updated on an annual basis.

Both of the Commission's Communications stressed the need to assess the adequacy of existing legal rules to protect individuals from potential adverse effects of AI systems, and to establish ethical guidance to ensure that the development, deployment and use of AI systems do not occur at the expense of fundamental rights and European values.

It is against this background that in June 2018 the High-Level Expert Group on Artificial Intelligence (AI HLEG) was established. Bearing in mind the holistic approach of the European strategy, the AI HLEG was tasked with the preparation of two deliverables: (1) AI Ethics Guidelines and (2) AI Policy and Investment Recommendations.<sup>8</sup> The documents are meant to complement each other. While the Guidelines provide guidance to AI practitioners<sup>9</sup> on how they can bring their AI systems in line with ethical principles (on top of any legal obligations they already have), the Recommendations are addressed to European institutions and Member States, offering them advice on which policy measures to adopt in order to ensure Europe's competitiveness in AI as well as an adequate regulatory framework, beyond ethical guidance.

On 8 April 2019, the AI HLEG published its Ethics Guidelines for Trustworthy AI. On the same day, the Commission adopted its third Communication on AI within a year, in which it welcomed the work of the AI HLEG and expressed its support for the ethical framework it put forward.<sup>10</sup>

This paper focuses on the creation of the AI HLEG “Ethics Guidelines for Trustworthy AI”, starting with a contextualisation of their aim and purpose (II.). Next, some insights on their drafting process are provided, which spanned not only expert group deliberations, but also over 500 comments received through an open consultation (III.). Particular attention is devoted to

---

<sup>5</sup> Accessible at: <http://data.consilium.europa.eu/doc/document/ST-14-2017-INIT/en/pdf>.

<sup>6</sup> COM(2018) 237 final, Brussels, 25.4.2018.

<sup>7</sup> COM(2018) 795 final, Brussels, 7.5.2018.

<sup>8</sup> More information on the group is available at: <https://ec.europa.eu/digital-single-market/en/news/call-high-level-expert-group-artificial-intelligence>.

<sup>9</sup> AI practitioners are defined in the Guidelines as “all individuals or organisations that develop (including research, design or provide data for), deploy (including implement) or use AI systems, excluding those that use AI systems in the capacity of end-user or consumer”.

<sup>10</sup> COM(2019) 168 final, Brussels, 8.4.2019.

the questions on the respective role played by ethics and law (IV.), and to the challenges that had to be overcome throughout the drafting process (V.). Subsequently, the Guidelines' next steps are set out (VI.), and their relevance against a background of international AI policy-making is examined (VII.). Finally, some concluding remarks are offered (VIII.).

## **II. The Guidelines' Aim and Purpose: Beyond Principles**

The Commission's AI HLEG was not the first to work on an ethical framework for Artificial Intelligence. On national level, a few Member States already started similar exercises within the context of their own national strategies, most notably Denmark<sup>11</sup>, Finland<sup>12</sup>, the UK<sup>13</sup> and Germany.<sup>14</sup> Also outside the EU, countries such as Singapore, Japan, Canada and Dubai made the development of ethical principles for AI an important element of their strategies. Furthermore, at international level, a number of governmental<sup>15</sup> and non-governmental<sup>16</sup> organisations started to undertake similar efforts.

In other words, a proliferation of ethical principles and guidelines at various levels is taking place. And while this is good news for the stimulation of the ethical debate on this important topic, it is less helpful for organisations operating within the European Union and looking for guidance as to which ethical rules they should respect when developing, deploying or using AI systems. In the context of a single European market, and the aspiration of a single European AI market<sup>17</sup>, such uncertainty and confusion is usually more of a foe than a friend. The Commission's aim was therefore to task an independent expert group with the preparation of European ethical guidelines for AI, with the request to take inspiration from, build on, and rise above more abstract existing guidance, in order to come up with an overall ethical framework that could be used by AI practitioners.<sup>18</sup>

---

<sup>11</sup> In March 2018, the Danish Ministry of Industry, Business and Financial Affairs established an expert group on data ethics, tasked to draw up a set of data ethical recommendations.

<sup>12</sup> The Finnish Ministry of Economic Affairs and Employment of Finland launched an AI Ethics Challenge, encouraging companies to publicly subscribe to ethical principles for AI, as part of its AI strategy.

<sup>13</sup> In November 2018, the UK's department for Digital, Culture, Media and Sport launched a new Centre for Data Ethics and Innovation, tasked to look into ethical issues relating to data and AI. It is but one of the various governmental initiatives around the subject.

<sup>14</sup> Both the German Parliament and the German Federal Government set up an AI Ethics Commission in late 2018, each of which are expected to deliver ethical guidance.

<sup>15</sup> These include for instance the OECD, UNESCO and the Council of Europe, the work of which is touched upon in Chapter VII of this paper.

<sup>16</sup> These include for instance the Asilomar Principles for AI, the work carried out by the Partnership for AI, the Montreal Declaration, and the AI4People initiative.

<sup>17</sup> See in this regard also the second deliverable of the High-Level Expert Group on Artificial Intelligence, which was published on 26 June 2019 and contains a recommendation to establish a single European Market for Trustworthy AI. This document is available at: <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>.

<sup>18</sup> See COM(2018) 237 final, Brussels, 25.4.2018, at pages 16-17. See also the Call for Experts, setting out the AI HLEG's Mandate, published on 9 March 2018 and accessible at <https://ec.europa.eu/digital-single-market/en/news/call-high-level-expert-group-artificial-intelligence>.

## 1. From Principles to Practice

The statement issued by the European Group on Ethics in Science and New Technologies (EGE) – a Commission advisory group – on AI, Robotics and Autonomous Systems<sup>19</sup> offered a first step towards the preparation of such ethical framework. Consisting of academics specialised in ethics of science and new technologies, their statement was the AI HLEG’s first input<sup>20</sup> of theoretical nature when starting the drafting process, and was further built on throughout the group’s work process.

The AI HLEG’s task however consisted not only in bringing together various ethical principles in a comprehensive manner, but – more importantly – to bring those principles “to the ground”. In order to offer concrete and practical guidance for AI practitioners, the AI HLEG had to set foot on the so far uncharted territory of operationalisation: what concrete operational steps should an AI developer take to ensure ‘human autonomy’ when building an AI system? What practical measures should an AI deployer put in place to ensure ‘transparency’ when deploying an AI system?

Any attempt to such operationalisation necessitates a multi-disciplinary and multi-stakeholder model, bringing together not only ethicists and philosophers, but also people with other relevant expertise such as computer scientists, engineers and lawyers. And bringing together not only academics and researchers, but also organisations developing and deploying AI systems in different sectors, as well as civil society organisations representing various interest groups, from workers to consumers.

## 2. A Trustworthy AI Guide for Practitioners

With as many perspectives as experts – 52 in total<sup>21</sup> – the AI HLEG set out to undertake a mapping of ethical issues and practical solutions thereto from a variety of angles. The explicit aim was thus to go beyond the drafting of yet another document listing ethical principles, in order to avoid that adhering to these ethics guidelines – by nature “voluntary” only – could turn into an empty PR-exercise without concrete action. Rather, the purpose was to provide guidance and tools that could help actually implementing those principles in practice.

These efforts resulted in the publication of Ethics Guidelines in April 2019, which start from a fundamental rights-based approach and set out a comprehensive framework to achieve “Trustworthy AI”. The notion of Trustworthy AI is described as a foundational ambition and

---

<sup>19</sup> Accessible at: [http://ec.europa.eu/research/ege/pdf/ege\\_ai\\_statement\\_2018.pdf](http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf).

<sup>20</sup> During the first meeting of the AI HLEG which took place in June 2018, the EGE representative attending the AI HLEG meetings presented the EGE statement to the expert group members, which was further discussed and elaborated on in subsequent discussions.

<sup>21</sup> The Commission aimed to select a diverse set of experts, in line with the selection criteria set out in the call for experts, accessible at: [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=5025](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=5025). Based on a careful consideration of the criteria, and the requirement to “seek a balance in terms of gender, age and geographic origin”, ultimately 52 experts were selected from a diverse range of backgrounds and interest groups, spanning 28 countries and counting 44% female experts. See also section V.1. on the group’s composition as a challenge.

comprises three components, entailing that the actors and processes involved in AI systems (including their development, deployment and use) should be:

- (1) *lawful* - complying with all applicable laws and regulations,
- (2) *ethical* - ensuring adherence to ethical principles and values, and
- (3) *robust* - both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm.

#### *a) Four Ethical Principles*

Guidance is provided in three chapters, from the most abstract to the most concrete. In the first chapter, drawing on the Charter of Fundamental Rights of the EU and international human rights law more broadly, the Guidelines prescribe four principles which should be considered as ethical imperatives in the context of AI: *respect for human autonomy, prevention of harm, fairness and explicability*.

#### *b) Seven Key Requirements*

From those principles, the second chapter derives seven more concrete requirements that AI systems should take into account: (1) *human agency and oversight* (including fundamental rights, human agency and human oversight), (2) *technical robustness and safety* (including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility), (3) *privacy and data governance* (including respect for privacy, quality and integrity of data, and access to data), (4) *transparency* (including traceability, explainability and communication), (5) *diversity, non-discrimination and fairness* (including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation), (6) *societal and environmental wellbeing* (including sustainability and environmental friendliness, social impact, society and democracy) and finally, (7) *accountability* (including auditability, minimisation and reporting of negative impact, trade-offs and redress).

#### *c) Practical Assessment List*

In a third chapter, the Guidelines offer an assessment list that operationalises each of the seven requirements through concrete questions aimed at guiding AI practitioners towards Trustworthy AI in practice. Finally, a last section of the Guidelines lists some examples of beneficial opportunities raised by AI, as well as critical concerns linked to the technology that need to be carefully considered, including for instance covert AI systems, the tracking of individuals, and the development of lethal autonomous weapons.

Rather than elaborating on the content of the Guidelines, the following sections of this paper will dive deeper into its drafting process and its role in the broader context of establishing an appropriate governance framework for AI.

### III. The Guidelines' Drafting Process: from 52 to over 500 contributors

On 18 December 2018, a first draft of the Ethics Guidelines was published. The document underwent an open consultation for a period of six weeks through the European AI Alliance<sup>22</sup>, during which over 500 commenters provided their feedback through an online form.<sup>23</sup> A meeting with Member States representatives<sup>24</sup> on 22 January 2019 provided the group with further feedback, and additional points were also raised by the members of the expert group itself during subsequent AI HLEG meetings.

Overall, the draft was deemed to constitute a good start for the creation of an ethical framework for AI and perceived as a welcome development, yet various aspects for improvement were pointed out. A document summarising those aspects (and the AI HLEG's way of dealing therewith) was made available on the platform European AI Alliance<sup>25</sup> simultaneously with the publication of the Guidelines' revised version on 8 April 2019.

#### 1. Irreconcilable Opinions

Interestingly, though not surprisingly, a number of topics received contradictory or conflicting comments, indicating a lack of consensus – or even a strong division – within the AI community. The most blatant example was the inclusion (or not) of concerns relating to “Artificial General Intelligence”<sup>26</sup> and “Artificial Moral Agents”<sup>27</sup> in the document, and the manner in which these concerns should be described and addressed. The conflicting views by members of the European AI Alliance merely reflected a similar divide of views among the AI HLEG members. The same was true for the document's “tone” when discussing AI systems, which was perceived by some to be too negative, and by others too optimistic.

---

<sup>22</sup> The consultation was openly accessible for members and non-members of the European AI Alliance alike.

<sup>23</sup> This feedback can be found here (to the extent commenters consented to the publication of their feedback on an open or anonymised basis): [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=57590](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57590).

<sup>24</sup> The European Commission established a Member States Group on Artificial Intelligence and Digitising European Industry, consisting of Director-Generals from the various EU member states as well as Norway and Switzerland, to prepare a Coordinated Plan on AI and align actions and strategies at EU level. The Member States Group met the AI HLEG in October 2018 at the margin of the AI Forum which took place in Helsinki, and again in January 2019 to discuss Member States' feedback on the first draft of the Guidelines published in December 2018.

<sup>25</sup> Accessible at: [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=58480](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58480).

<sup>26</sup> As opposed to (currently existing) systems of *Artificial Narrow Intelligence*, denoting AI systems that are only capable of carrying out a domain-specific task, *Artificial General Intelligence* is used to describe (currently non-existing) machines capable of understanding or learning any intellectual task that a human being can. A number of research labs within and outside Europe are aiming to develop this type of AI and – while it is widely disputed whether or not such systems are capable of being created – several researchers nevertheless consider it primordial to raise attention to the potential ethical consequences that such systems may generate.

<sup>27</sup> Defined in the draft Guidelines as systems that (a) autonomously arrive at normative judgments and conclusions, and (b) autonomously act on the basis of such self-generated judgments and conclusions.

## 2. Acceptable Trade-Offs

On some issues, however, comments were more converging, and hence easier to mold into an improved version of the document. Several commenters thus flagged that the Guidelines' different chapters would benefit from better integration and from a more extensive emphasis on the context-specificity of AI systems' ethical challenges, which the AI HLEG attempted to remedy. Moreover, guidance was requested for those situations where the simultaneous application of different ethical principles or requirements might raise tensions, which led to the addition of a section dedicated to "trade-offs" in the final document. At the same time, it was stressed that this does not mean that these principles can or should always be traded-off and balanced against each other. Indeed, the final document states "*there may be situations, however, where no ethically acceptable trade-offs can be identified. Certain fundamental rights and correlated principles are absolute and cannot be subject to a balancing exercise (e.g. human dignity).*"<sup>28</sup> Consequently, "*in situations in which no ethically acceptable trade-offs can be identified, the development, deployment and use of the AI system should not proceed in that form.*"<sup>29</sup>

## 3. AI For Europe

In addition, the objective to work towards Trustworthy AI "made in Europe", echoing the words of the Coordinated Plan of the Commission and Member States, was deemed too narrow. In fact, it was found that any ethical framework for AI should apply to all AI applications in Europe, regardless of whether they were "made" within or outside its borders. This push for a wider geographical scope is interesting to bear in mind in the context of existing or future EU regulation applying to AI systems. Akin to the logic behind the extraterritorial application of the GDPR, the comments received indicate the desirability to ensure that obligations relating to the protection of individuals from AI's potential adverse impact should also apply to non-European actors when affecting individuals in Europe.

## IV. The role of Ethics and Law: Both Necessary, Neither Sufficient

Besides the material issues discussed in section III. above, it is worth elaborating on another particular point raised by numerous commenters, as it is likely to continue shaping the debate on Trustworthy AI in the months and years to come. It was claimed that the Draft Guidelines at some instances conflated ethics and law, which at best created confusion for the document's addressees, and at worst seemed to suggest a so-called "ethics-washing"<sup>30</sup> approach. To remedy this perceived concern, the group set up a small "legal task force" with the specific assignment of going through the document and eliminating such risk.

---

<sup>28</sup> Guidelines, at page 13.

<sup>29</sup> Guidelines, at page 20.

<sup>30</sup> This term is used to describe the phenomenon whereby adherence to ethical principles is used as a "substitute" of regulation, with the specific aim to avoid the creation of binding legal rules. See e.g. Wagner, B. (2018), "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?", in M. Hildebrandt (ed.), *Being Profiling. Cogitas ergo sum*, Amsterdam University Press.



## 1. Textual Clarifications

This led to a number of clarifications in the text, and an increased emphasis on already existing sentences which excluded such reading of the document. For instance, the first of the formerly two components of “Trustworthy AI” – which used to consist of “respecting fundamental rights, applicable regulation and ethical principles” on the one hand, and “being technically robust” on the other – was disentangled, clarifying that a difference exists between legal obligations and ethical obligations, and that AI systems could only be deemed trustworthy<sup>31</sup> if both (1) *lawful* and (2) *ethical*, on top of being (3) *robust*.

In the same spirit, potential confusion created by the status of fundamental rights – sometimes referred to as special moral entitlements inherent to human beings (i.e. “ethical”), sometimes referred to as legally enforceable rights (i.e. “lawful”) – was cleared up. Moreover, the non-binding nature of the guidelines was re-emphasised, as well as the fact that many of its ethical principles are already enshrined in existing regulation, compliance with which is a prerequisite that in no way can be escaped. And to further eliminate any doubt, it was restated that “*these Guidelines do not intend to substitute any form of current or future policymaking or regulation, nor do they aim to deter the introduction thereof*”.<sup>32</sup>

Criticism was however voiced not only on the Draft’s seeming conflation between ethics and law, but also on the document’s inherent non-binding nature, and hence on the group’s mandate itself. It should, however, be noted that the group was explicitly tasked with the drafting of non-binding Guidelines, hence rendering criticism on their nature towards the AI HLEG<sup>33</sup> appears misplaced or, at the very least, wrongly addressed. Moreover, such criticism seems to forego the crucial complementarity between law and ethics, and falls for the very fallacy of which it accuses its opponent to make: conflating ethics and law.

## 2. A Symbiotic Relationship

Both ethics and law have their role to play within a framework for AI governance, both necessary, and neither sufficient. Moreover, neither can substitute the other. Rather, they act in a complementary way, and are able to inform and inspire each other. Undeniably, however, the reach and limits of their respective roles and territories form the subject of a heated debate.<sup>34</sup>

---

<sup>31</sup> Another criticism pointed to the fact that the term “trust” is reserved for interpersonal relationships, rendering any attribution of such feature to a machine moot. The revised version of the Guidelines now includes a paragraph which explicitly states that, when mentioning “Trustworthy AI”, reference is made not just to the system as such, but rather to the qualities of the socio-technical systems involving AI applications, comprising also the human beings or actors developing, deploying and using them. See the Guidelines at page 5: “*It is not simply components of the AI system but the system in its overall context that may or may not engender trust. Striving towards Trustworthy AI hence concerns not only the trustworthiness of the AI system itself, but requires a holistic and systemic approach, encompassing the trustworthiness of all actors and processes that are part of the system’s socio-technical context throughout its entire life cycle.*”

<sup>32</sup> See the Guidelines at page 3.

<sup>33</sup> It should in this regard also be stressed that expert groups do not have the power to create any binding legal rules.

<sup>34</sup> See for instance G. C. Shaffer and M. A. Pollack, “Hard vs. Soft Law: Alternatives, Complements, and Antagonists in International Governance”, *Minnesota Law Review* 84, 2010, pp. 706-799; R. De George, Law and Ethics in the Information Age, *Business and Professional Ethics Journal*, Volume 20, Issue 3/4,

And while this debate is far from new, it has received renewed attention in the context of policy-makers' explicit exploration of new governance models for AI.

The Commission's decision to task an independent expert group with the drafting of non-binding ethics guidelines for AI has further fuelled this debate. Yet focusing on that aspect alone would be like walking in a flowering meadow and only looking at one flower – it entirely ignores the fact that (i) countless EU regulations already apply to AI systems today (both at horizontal and at sectoral level), (ii) that EU regulation is automatically reviewed and updated periodically<sup>35</sup>, (iii) that the Commission is speeding up a number of those reviews specifically to ensure their adequacy for – and spur their potential adaptation in light of – the emergence of new technologies such as AI systems<sup>36</sup> and (iv) that no moratorium on new regulation was ever pronounced.<sup>37</sup>

### 3. Policy Recommendations for Trustworthy AI

Perhaps even more importantly, it should be noted that the Commission also mandated the same expert group to prepare broader Policy and Investment Recommendations on AI (Recommendations), which include advise on the necessity to update existing regulation or adopt new binding rules. In their Recommendations, which were published in June 2019 – just three months after the publication of the Guidelines – the AI HLEG explicitly opted to maintain the term “Trustworthy AI”, denoting that both documents are heavily linked. Indeed, as the introduction to the Recommendations states, the Guidelines “*constituted a first important step in identifying the type of AI that we want and do not want for Europe, but that is not enough to ensure that Europe can also materialise the beneficial impact that Trustworthy AI can bring*”. The Recommendations therefore dive into the policies and measures – beyond voluntary guidance – that European institutions and Member States should adopt in order to ensure such beneficial impact. This encompasses both measures to boost Europe's development of and competitiveness in AI so as to secure the benefits its brings, and measures to empower and protect individuals and society in the face of the risks raised by AI.

Besides references to such protective measures being made throughout various other sections of the document<sup>38</sup>, an entire section is dedicated to “*establishing an appropriate governance*

---

Fall/ Winter 2001; R. Hagemann, J. Huddleston and A.D. Thierer, *Soft Law for Hard Problems: The Governance of Emerging Technologies in an Uncertain Future*, Colorado Technology Law Journal, February 2018, available at SSRN: <https://ssrn.com/abstract=3118539>.

<sup>35</sup> This is also part of the European Commission's Better Regulation initiative, as set out in the Commission Communication COM(2015) 215 final.

<sup>36</sup> In this regard, it can be noted that the European Commission also established an expert group focused on liability and new technologies in 2018, tasked inter alia with assessing the need to revise the Product Liability Directive. The same is not unlikely to happen as concerns other legal domains or instruments.

<sup>37</sup> To the contrary, the Commission's new President, Ursula von der Leyen, who will be heading the EU Institution responsible for drafting legislation between 2019 and 2024, has already declared her intention to propose legislation for a coordinated European approach on the human and ethical implications of AI systems within her first 100 days of office, thus building further on the focus given to this topic by the previous Commission. See: [https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission\\_en.pdf](https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf).

<sup>38</sup> Under Chapter I, dealing with the empowerment and protection of humans and society, a number of proposals for potential regulatory action are made, for instance regarding the self-identification of AI systems and the

*and regulatory framework*". In that section, the AI HLEG not only calls for the exploration of binding regulation to tackle the "critical concerns" mentioned in the Guidelines, but it also puts forward the need to consider inter alia mandatory traceability, auditability and ex-ante oversight obligations (and hence the translation of the Guidelines' requirements into binding regulation) for AI systems that can potentially have a significant impact on human lives.

#### **4. A Comprehensive Approach**

In other words, the two deliverables of the AI HLEG are closely intertwined and should not be seen in isolation. The experts of the AI HLEG acknowledged the important role that both ethics and law have to play and described their symbiotic relationship<sup>39</sup>, without being blind to their respective limitations. They carried out their mandate as prescribed by the European Commission which established the group, and ensured that the documents they delivered together offer a comprehensive approach for AI policy-making, acknowledging the need for both an ethical and regulatory framework in Europe, together ensuring an appropriate governance approach for AI.

### **V. A Journey Not Devoid of Challenges**

While the purpose of the AI HLEG's task might have been clear, the road to achieve it was not devoid of challenges – and that is an understatement. The two main challenges that the AI HLEG had to overcome in order to deliver its Guidelines were (i) the group's composition and (ii) the timeline.

#### **1. Diverse Composition**

The group's main strength – its diversity – also constituted a significant hurdle. With a subject as "personal" as ethics, and a group comprising members from very different horizons, uniting such diversity of opinions was not a given. The quest for unity was not merely a lofty goal set

---

surveillance of individuals through such systems. Chapter III, dealing with Europe's public sector, likewise points out a number of such areas, for instance concerning citizen scoring and the right to information about AI-informed governmental decisions.

<sup>39</sup> See in this regard also page 2 of Guidelines, which elaborates on the three different components of Trustworthy AI (law, ethics and robustness): "*Each component in itself is necessary but not sufficient for the achievement of Trustworthy AI. Ideally, all three components work in harmony and overlap in their operation. If, in practice, tensions arise between these components, society should endeavour to align them.*" Moreover, on page 6 the link between ethical and lawful AI is further developed: "*Achieving Trustworthy AI requires not only compliance with the law, which is but one of its three components. Laws are not always up to speed with technological developments, can at times be out of step with ethical norms or may simply not be well suited to addressing certain issues. For AI systems to be trustworthy, they should hence also be ethical, ensuring alignment with ethical norms.*" This relationship is also mentioned in L. Floridi, "Establishing the rules for building trustworthy AI", *Nature Machine Intelligence*, volume 1, pages 261–262 (2019).

by the group, but specifically foreseen in the AI HLEG’s Rules of Procedure, stipulating the adoption of its documents by consensus.<sup>40</sup>

The Commission explicitly opted for a multi-stakeholder approach, in line with the criteria set out in the call for experts.<sup>41</sup> This meant that not only academics were included (covering disciplines such as ethics, philosophy, law, computer science and engineering, all relevant when it concerns practical guidelines), but also civil society representatives (ranging from consumer organisations and NGOs to trade unions). This also entailed the inclusion of industry representatives (from different sectors and with broad practical experience of AI use cases), as being those meant to actually implement the guidelines on a voluntary basis. Critically, the fact that adoption of the document had to occur by consensus rather than by vote, meant that not one stakeholder group could impose its views upon others.<sup>42</sup> Moreover, pursuant to the Rules of Procedure, every individual member of the group had the option to not sign the document and instead write a dissenting opinion, explaining his or her dissenting views.<sup>43</sup>

Despite the short time frame set out by the European Commission in the group’s mandate, the AI HLEG managed to meet 9 times<sup>44</sup> in person for in-depth discussions, in addition to the many online interactions, before adopting the Guidelines by consensus and delivering them to the Commission. Over time, those 52 diverse experts progressively gained the ability to find a shared language<sup>45</sup> and reach convergence on an ever-increasing number of subjects. While some topics remained controversial until the very last<sup>46</sup>, the consecutive reiterations of the document and continued fine-tuning of wording allowed the framework to take shape based on commonly found ground.

An important step towards such common ground was taken with the decision of using a fundamental rights-based approach to frame the document. This is not surprising, as in the quest

---

<sup>40</sup> This is in line with the standard Rules of Procedure of Commission expert groups, which contains the default rule that only when consensus cannot be reached the group shall proceed to a vote. See also the Rules of Procedure of the AI HLEG point 6, accessible at the Register of Commission Expert Groups, <http://ec.europa.eu/transparency/regexpert/>. Note however the statement at the document’s first page, pointing out that not every member of the group necessarily agrees with every single statement of the document. This denotes that the Guidelines’ adoption occurred by consensus on the document as a whole.

<sup>41</sup> Available at: [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=5025](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=5025).

<sup>42</sup> It should moreover be noted that, when it came to the most controversial points discussed by the group, such as for instance the inclusion of Artificial General Intelligence as a concern, or the inclusion of the “principle of beneficence”, the dividing lines between members’ opinions often did not at all reflect the stakeholder groups.

<sup>43</sup> See also the Rules of Procedure of the AI HLEG point 6, accessible at the Register of Commission Expert Groups, <http://ec.europa.eu/transparency/regexpert/>. In this regard, it can be noted that each of the AI HLEG members decided to sign the document, without opting to make use of their right to write a dissenting opinion. The document was hence adopted by consensus.

<sup>44</sup> Meetings were held on 28 June 2018, 20 September 2018, 8 and 9 October 2019, 8 November 2018, 13 and 14 December 2018, 22 and 23 January 2019, 14 February 2019, 28 February 2019 and 18 and 19 March 2019 before the Guidelines were adopted.

<sup>45</sup> The word “bias”, for instance, has a different connotation for a computer scientists and an ethicist.

<sup>46</sup> As mentioned above, no point sparked such discussion as the ethical risks posed by the potential emergence of Artificial General Intelligence and Artificial Moral Agents, which was debated within the AI HLEG until the last hour of the last meeting before the Guidelines’ adoption. While the importance of assessing long-term risks in itself was widely accepted, the question of whether the aforementioned phenomena could actually be considered as such risk proved highly controversial.

for moral universalism, the human rights *acquis* is often put forward as one of the most viable conceptual candidates, hence increasing the potential for the framework’s acceptability also beyond Europe.<sup>47</sup>

## 2. Strenuous Timeline

The second obstacle to overcome consisted of the group’s timeline. The deadline to complete a first draft of the Guidelines was set to the end of 2018, or six months from the group’s very first meeting, with a revised version due three months later in March. While not unusual for expert groups, this 9-month deadline was very short considering the fact that the group was tasked with not one but two reports<sup>48</sup>, and given its vast size<sup>49</sup> – vaster than most Commission expert groups. Though already announced in the call for experts, this timeline was heavily criticised from numerous angles, not least from within the AI HLEG.

At the same time, it could not be denied that speed was of the essence in view of the continuous and rapid developments taking place in the field. Not only did the global AI ethics boom herald a race for ethical leadership, but the increasing reports of ethically problematic behaviour by AI practitioners – often also occurring involuntarily or in ignorance – stressed the urgent need for both awareness and guidance. Moreover, the AI HLEG’s Guidelines solely presented a starting point for debate, constituting a living document rather than an unchangeable truth. While sacrificing quality over speed could not be accepted, it would also have been a mistake to believe that more time would allow for a “perfect” or “finalised” document, this being terminology that simply does not belong to the sphere of ethics.<sup>50</sup>

While the first two chapters of the document had benefitted from substantial discussion and feedback from the public at large during the 9-month preparation period, the AI HLEG remained sceptical about releasing the Guidelines’ third chapter, the practical assessment list that aims to operationalise the framework, by the deadline. The group’s original plan was to test out this list in practice across a few use cases<sup>51</sup>, enabling “bottom up” practical feedback on how it works before releasing the final document. That aspiration, however, proved impossible within the given time line.

---

<sup>47</sup> See for instance the International Bill of Human Rights, of which the Universal Declaration of Human Rights, the International Covenant of Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights form part. While a large number of countries spanning different world regions may be recognising the rights enshrined in those documents, it should be noted that these rights are not universally nor equally enforced.

<sup>48</sup> The second deliverable of the AI HLEG, the AI Policy and Investment Recommendations, were due before the summer of 2019 and were published on 26 June 2019.

<sup>49</sup> See for instance the Commission’s High-Level Expert Group on the impact of digital transformation on EU labour markets (HLG-FoW), with a time line spanning nine months: <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-impact-digital-transformation-eu-labour-markets>. Important detail: this group however consisted of 9 experts instead of 52.

<sup>50</sup> For a definition of “ethics” as understood by the AI HLEG for the purpose of its work, reference can be made to the Guidelines’ glossary, at page 37.

<sup>51</sup> This was also mentioned in the draft version of the Guidelines published on 18 December 2018, at page 28.

Ultimately, this obstacle was likewise overcome. *First*, through a great amount of hard work and dedication from the group members, who strived to deliver something of quality despite the short time frame. *Second*, through the introduction of a unique solution: a piloting phase. Indeed, insisting that the assessment list would need to be tested across various use cases and organisations to ensure practical relevance in different contexts, the AI HLEG proposed the European Commission to set up a piloting process. While the AI HLEG reached consensus to put forward a horizontal framework of principles and requirements applicable to all AI systems, for the framework's concrete implementation, a one-size-fits-all approach was not deemed appropriate, particularly given the context-specificity of AI's ethical issues. The Commission's agreement to launch the piloting phase, hence, bought the expert group additional time to gather feedback and prepare a more tailored assessment list in 2020.

## **VI. The Guidelines' Next Steps – An innovative Piloting Phase**

The European Commission counts around 700 active expert groups.<sup>52</sup> While those groups are usually tasked to draft an opinion or report advising the Commission on a particular subject, their input is not binding. The Commission remains entirely independent in the way it takes into account the groups' advice and expertise. Reports prepared by expert groups often serve as an important source of inspiration for EU policy-making, yet certainly do not get translated into EU policy on an automated basis, and only rarely form the object of a Commission Communication.

For the Commission to adopt a Communication – or any other policy instrument for that matter – agreement is needed from the entire college, and hence from all cabinets and Directorate-Generals (DG's) constituting the institution, whether or not the subject matter is directly relevant to them. The achievement of such agreement is arguably a good guarantee for a nuanced and balanced approach to EU policy, since every cabinet and DG has a different area of focus and expertise, and can, hence, ensure that their angle is properly considered – without trumping the interests of others in a disproportionate manner. Yet reaching such agreement can often be a lengthy process.

### **1. Welcome by the European Commission**

Interestingly, when the AI HLEG presented its finalised Guidelines to the European Commission in March 2019, agreement to welcome the document by means of a Commission Communication was relatively swiftly reached. Published on 8 April 2019 – just a few weeks after the Guidelines' finalisation, the Communication *Building Trust in Human-Centric AI*<sup>53</sup> not only stated its support for the seven key requirements, but also encouraged all stakeholders to implement them when developing, deploying or using AI systems. This is, however, not very

---

<sup>52</sup> The Register of Commission expert groups contains further information on those different groups, including the Commission department running the group, its tasks and membership, as well as relevant documents. The Register is accessible at: <http://ec.europa.eu/transparency/regexpert/index.cfm>

<sup>53</sup> COM(2019) 168 final, Brussels, 8.4.2019. Accessible at: <https://ec.europa.eu/digital-single-market/en/news/communication-building-trust-human-centric-artificial-intelligence>.

surprising, as the Guidelines already reflected – in light of the multi-disciplinary and multi-stakeholder composition of the AI HLEG – many of the different angles around the table of the Commission’s college. Moreover, the Communication pre-empted the undoubtedly numerous questions that the Commission would otherwise need to face concerning the Guidelines’ status, as the Guidelines after all remain a document prepared by an independent expert group, and not by the Commission itself.

## **2. Launch of the Piloting Phase**

The Communication also announced the group’s suggestion to launch the assessment list’s piloting phase, acknowledging the need to ensure its practical and sectoral relevance. This launch formally took place on 26 June 2019 at the first Assembly of the European AI Alliance<sup>54</sup>, and the piloting phase runs until 1 December 2019. The aim is to gather practical feedback from an as large and diverse as possible set of organisations (both from the private and public sector, spanning different operation domains and applications, and different sizes of organisations). Feedback is collected through an online survey, through in-depth interviews with AI practitioners, and through the European AI Alliance. This should result in a revised assessment list which is expected to be published in the first half of 2020.

Setting aside the ultimate end result, it is particularly the process set up to achieve it – a “testing phase” akin to the idea of “policy sandboxing” – that is interesting from a policy-making perspective. Rather than rushing into a regulatory process to tackle some of the identified challenges posed by the technology, the European regulator – backed by EU Member States who likewise rejoiced in the concept of the piloting phase – established a space to test out how a novel and all-purpose technology such as AI best be governed, and how such challenges best be addressed. The data that will be collected from the piloting phase may offer Europe a competitive advantage when it comes to being the “smartest” regulator, a feature not unimportant for a technology as crucial as AI, for which the arena of regulatory competition is already starting to get crowded.<sup>55</sup> Furthermore, this exercise can also help the EU defining its position amidst the discussions on ethical AI taking place at international level.

---

<sup>54</sup> The European AI Alliance is a multi-stakeholder platform at EU level, hosted on an online platform and engaged in a broad and open discussion of all aspects of AI development and its impacts. Members of the Alliance can interact with the experts of the AI HLEG (which acts as its steering group) and with other members in a forum-style setting. Discussions hosted on the platform contribute to the European debate on AI, and feed into the European Commission’s policy-making in this area. Today, the Alliance counts over 3500 members. Anyone can register to become a member. For more information, see: <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>.

<sup>55</sup> This relates not only directly to AI as a technology (as for instance announced by the European Commission’s president-elect following the recommendations of the AI HLEG – see supra footnote 37), but also indirectly to the building blocks of AI such as data (see infra footnote 66). For a recent analysis, see for instance: <https://www.ft.com/content/0c91b884-92bb-11e9-aea1-2b1d33ac3271>.

## VII. The Guidelines beyond Europe: Towards a Global Governance Framework

Europe has not been alone in making strides towards addressing the ethical concerns raised by AI systems – and that is a good thing. Although “*Europe has taken a strong initiative to lead the global debate on the applied ethics of AI*”<sup>56</sup>, the ethical impact of AI reaches beyond Europe, and, hence, global solutions must be found to address the global challenges that AI brings forth.<sup>57</sup> Discussions with other countries and regions, taking place at global fora such as the G7 and the G20, and within other organisations such as the OECD, the Council of Europe and UNESCO, are essential to work towards a much-needed global governance framework for AI.

### 1. Promising Initiatives

In its Communication of 8 April, the Commission acknowledged the importance of international cooperation in this field, stating that it will “*continue its efforts to bring the Union’s approach to the global stage and build a consensus on a human-centric AI*”, adding that “*the European Union can have a leadership role in developing international AI guidelines and, if possible, a related assessment mechanism*”.<sup>58</sup>

Various initiatives are aiming to materialise that aspiration. In December 2018, the Council of Europe’s Commission for the Efficiency of Justice (CEPEJ) adopted an ethical charter for the use of AI in judicial systems.<sup>59</sup> Whilst limited in scope to the judicial context, the principles of the CEPEJ’s charter largely overlap with the Guidelines of the AI HLEG, which instead are meant to apply to all AI systems. Moreover, in May 2019, the Foreign Ministers of the Council of Europe member states agreed to examine the creation of a legal framework for AI<sup>60</sup>, hence asserting the importance of international governance mechanisms also beyond voluntary guidance. Simultaneously, UNESCO’s Secretary General has been very vocal about the institution’s role in the international AI governance sphere, and announced the plan to develop an ethical framework for AI with its members over the next few years.<sup>61</sup>

Besides the European Union, no international institution has been as advanced on developing such framework as the OECD. In an exercise that started by gathering independent experts as well as the OECD’s member states representatives, the OECD published its ethical principles for AI in May 2019, just a month after the AI HLEG Guidelines were presented. Interestingly, the principles of the OECD to a very large extent converge with the framework embraced by the European Commission – and this is no surprise. In fact, a number of experts of the AI HLEG

---

<sup>56</sup> See the Recommendations, at page 24.

<sup>57</sup> While the call for global solutions does not need to imply a curtailing of culturally diverse views, it does imply working towards a common understanding of values that are shared cross-culturally, and that require global cooperation in order to be preserved.

<sup>58</sup> See COM(2019) 168 final, Brussels, 8.4.2019, at page 8.

<sup>59</sup> The Charter drafted by CEPEJ is accessible at: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.

<sup>60</sup> For more information, see <https://www.coe.int/en/web/artificial-intelligence/-/foreign-ministers-legal-framework-for-artificial-intelligence-is-a-priority>.

<sup>61</sup> See for instance at: <https://en.unesco.org/courier/2018-3/towards-global-code-ethics-artificial-intelligence-research>.



were also part of the expert group of the OECD, and the Commission itself was one of the OECD's expert group members. Moreover, the OECD decided to use the same concept of "Trustworthy AI" which was put forward by the AI HLEG, likewise rendering its principles to broadly coincide with those of the Guidelines.

## **2. The Guidelines as Inspiration**

This apparent spread of the AI HLEG's Guidelines concepts is a welcome development, as it confirms the soundness of the framework proposed by the group, and facilitates the achievement of a global consensus on such framework, given the fact that 42 countries<sup>62</sup> adopted the OECD principles. Those principles also formed the basis of the discussions that took place during the G20 in June 2019, where a declaration<sup>63</sup> was made that expressed support for ethical principles drawn therefrom, including by a number of countries which were not part of the OECD exercise.<sup>64</sup> On the other hand, it should be noted that the adopted OECD principles are relatively abstract, omitting to go into the details of what adherence thereto requires in practice. It hence remains to be seen if and how their implementation will occur by the subscribing countries.

What sets the Guidelines of the AI HLEG apart from the OECD principles, is the fact that the Guidelines are more elaborate, more concrete and more far going (for instance requiring the auditability of AI systems with a certain level of risk). Moreover, as previously mentioned in section II.1 above, the AI HLEG explicitly decided to move beyond the listing of ethical principles and devised an assessment list that operationalises the framework, allowing it to be used by AI developers, deployers and users in practice. The piloting phase, which encourages organisations to test out the assessment list on concrete AI applications, further strengthens this practical nature. In addition, the decision to anchor the Guidelines in a fundamental rights-based approach, offers a strong signal on the necessity to link back to existing and foundational legal instruments, which in the EU concern the Treaties and Charter of Fundamental Rights. More generally, the Guidelines and its concrete requirements can not only complement, but where needed also provide inspiration to adapt and improve the existing regulatory framework for AI in Europe and beyond.

## **VIII. Concluding Remarks**

With the establishment of the High-Level Expert Group on AI and the delivery of the Ethics Guidelines for Trustworthy AI in April 2019, Europe has taken an important leap towards a common ethical framework for AI, which has resonated through other initiatives. Three aspects of Europe's approach are worth highlighting in this concluding section, as they are bound to set

---

<sup>62</sup> The ethical principles for AI were adopted by the OECD's 36 member countries, along with Argentina, Brazil, Colombia, Costa Rica, Peru and Romania.

<sup>63</sup> See the G20 Ministerial Statement on Trade and Digital Economy of June 2019, accessible at: [https://g20trade-digital.go.jp/dl/Ministerial\\_Statement\\_on\\_Trade\\_and\\_Digital\\_Economy.pdf](https://g20trade-digital.go.jp/dl/Ministerial_Statement_on_Trade_and_Digital_Economy.pdf).

<sup>64</sup> Including for instance two major players in the field, China and Russia. This also indicates support for those principles by non-Western countries.

the tone for future policy-making around AI systems: (i) *inclusivity*, (ii) *agility* and (iii) *globality*.

## 1. Inclusivity

First, the framework for Trustworthy AI – stooled on a fundamental rights-based approach and consisting of four ethical principles, seven key requirements and an operationalising assessment list – was achieved through an inclusive process. The Commission brought together not only 52 experts from multiple countries, multiple disciplines and multiple stakeholder groups, but also ensured the inclusion of society at large through establishing and consulting with the broader European AI Alliance, a multi-stakeholder platform which at the time of writing counts over 3,500 members. As the impact of AI is multi-fold, only an approach that can map those impacts and their various challenges through different angles and points of view, can aspire to be successful in adequately addressing those challenges. As AI systems are increasingly being used in domains that affect society at large, finding novel and accessible methods to ensure society’s awareness and participation in the policy-making debate is crucial, and will only grow in importance.<sup>65</sup>

## 2. Agility

Second, the setting-up of the Guidelines’ piloting phase indicates an openness for agile and evidence-based policy-making, with due attention to the specificities of different applications. Despite the cliché that “*China and the US innovate, while Europe can only regulate*”, the EU should not shy away from regulating issues that require regulatory solutions, and has shown to be on the winning end with its regulatory approach to data protection (the GDPR now for instance being a major export product<sup>66</sup>). However, Europe at the same time also shows acknowledgment of the importance to regulate the right issues, in the right way, at the right time, and ensuring that new regulation does not unnecessarily hamper beneficial innovation.

Policy-makers are only beginning to understand the many benefits that AI systems can generate, and the ways in which they can pose risks – beyond the risks of other already existing ICTs. The piloting process, hence, offers an interesting tool to increase awareness around those risks, encourage stakeholders to put in place mechanisms to tackle them, and at the same time learn what the most efficient mechanisms for this purpose are – in a tailored and risk-based manner. In other words, it fosters regulatory humility, focusing on learning rather than rushing. Yet by learning to tackle those challenges even just an inch faster than other regions, Europe can gain

---

<sup>65</sup> See Kaeseberg, The Code-ification of Law and Its Potential Effects, CRi 2019, ■■ (in this issue).

<sup>66</sup> Indeed, various countries outside the European Union have adopted similar data protection rules, akin to or inspired by the GDPR, which is increasingly being recognised for its high standard-setting in data protection. The European Commission has also recognised a number of countries to have “adequate” data protection rules in place (based on the standard of the GDPR) that can enable the free flow of data to such countries based on an adequacy decision. So far, Andorra, Argentina, Canada (commercial organisations), Faroe Islands, Guernsey, Israel, Isle of Man, Japan, Jersey, New Zealand, Switzerland, Uruguay and the United States of America (limited to the Privacy Shield framework) were recognised as providing such adequate protection.

a major advantage, and find the right balance between ensuring adequate protection from AI's adverse impact, while enabling the beneficial development and use of the technology to flourish. It, however, remains to be seen whether the European Commission will deliver on the promise that the piloting phase holds, and make appropriate use of that process to inform its policy-making.

### **3. Globality**

Finally, in each of the three AI-related Communications published by the Commission, international outreach – and the striving towards global solutions – was deemed essential. Although it was made clear that discussions with other countries should not lead to compromising the protection of fundamental rights and other core values enshrined in the EU Treaties and Charter, the Commission acknowledged the necessity to build international consensus around the governance of AI. This indicates that the ethical issues raised by AI will become increasingly relevant not only when it comes to policy-making in Europe, but also in the EU's external relations. The focus on globality, however, also strengthens the need for a unified position at European level that can be backed by all EU Member States. To the extent that some of AI's ethical issues may touch upon traditionally national competences, consensus building at EU level may bring forth challenges of its own. The strong push given by the AI HLEG's Recommendations to work towards a "Single European Market for Trustworthy AI" only increases the urgency of reaching a common European understanding – yet in this regard the support expressed by Member States for pursuing the AI HLEG's work is encouraging.<sup>67</sup>

### **4. Human Intelligence for Appropriate Governance**

Artificial Intelligence does not overcome us. It concerns nothing more than an application, system or tool, developed by human beings. As most tools developed by humans, it can be used in a way that causes harm, it can be used in a way that causes unintentional harm, it can be used in a way that counters harm, and it can be used in a way that causes good. If we are intelligent enough to create AI systems, we must be intelligent enough to ensure appropriate governance frameworks to harness the good use of those systems, and avoid those that lead to (un)intentional harm.

The Ethics Guidelines for Trustworthy AI constitute a step towards such framework, explicitly identified by their authors as a starting point for discussion rather than a final truth. The Guidelines provide a helpful tool for AI practitioners to assess and improve their AI systems, and may serve as a source of inspiration for future (binding) governance approaches. However, they form part of a continuous journey towards an appropriate governance framework for AI and are but a small piece in the large tapestry of AI policies currently being woven by policy-makers and stakeholders around the world. These Guidelines - and any other instrument - will ultimately only be as effective as we will allow them to be.

---

<sup>67</sup> See in this regard the European Council's Conclusions adopted earlier this year, available at: <https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/en/pdf>.