Feedback on EC White Paper on Artificial Intelligence

Leuven.AI & Centre for IT & IP Law (CiTiP)
University of Leuven (KU Leuven)

1. INTRODUCTION

This report contains the feedback of **Leuven.AI**[1] and the KU Leuven **Centre for IT & IP Law (CiTiP)**[2] on the White Paper on Artificial Intelligence (AI) issued by the European Commission in February 2020. Leuven.AI brings together the wealth of expertise on AI that exists at the KU Leuven, spread over multiple faculties and departments. CiTiP is a research centre at the Faculty of Law of the KU Leuven with a staff of over 50 researchers specialised in legal and ethical aspects of IT innovation and intellectual property

Leuven.AI and CiTiP **welcome the publication of the White Paper** on AI as well as the opportunity to give feedback on it. The White Paper contains several interesting perspectives and ideas such as a risk-based approach to the regulation of AI. We also welcome its reliance on the Ethics Guidelines for Trustworthy AI developed by the High Level Working Group on AI (AI HLEG) including human oversight and accountability. The White Paper aims to enhance technological innovation without imposing excessive regulatory compliance requirements upon companies and organisation in the European Union (EU). However, several concerns and problems remain as well. These, for instance, relate to the governance of AI, the terminology that is used and the implementation/ feasibility of some proposals.

Researchers at CiTiP organised the online feedback session on Wednesday 29 April 2020. The wider research community at the KU Leuven working on AI was also involved. Each part of the White Paper was thoroughly discussed and analysed by experts from various fields and with different academic backgrounds. Against this background, we gathered some **general feedback** (part 2), **feedback on the ecosystem of excellence** (part 3) and **feedback on the ecosystem of trust** (part 4). Some **concluding remarks** are provided as well (part 5).

2. GENERAL FEEDBACK

- **Dual use AI-systems**: footnote 7 mentions that "Although further arrangements may need to be put in place to prevent and counter misuse of AI for criminal purposes, this is outside the scope of this white paper". It remains unclear whether there will be additional documents on the dual use of AI-systems (civilian/military) and on Lethal Autonomous Weapons Systems (LAWS). The White Paper should make this more clear and the Commission may even consider to address these topics more thoroughly in its revised version.

[1] See: https://ai.kuleuven.be/.
[2] See: https://www.law.kuleuven.be/citip/en.

- **Definition of AI:** the White Paper relies on several definitions/concepts of AI (compare p. 2 & 16 with footnotes 46 & 47). We recommend to be less ambiguous and use a single definition of AI. The idea of AI as "data", "algorithms" and "computing power" is linked to computer sciences. It may be an option to rely on and use the definition of AI as developed by AI HLEG in its Ethics Guidelines for Trustworthy AI.

- **Consensus definition AI:** closely related to the previous point is the importance to reach an international consensus on the definition of AI. In this regard, it may be examined how the EU, the United States and China define AI, and what the possible consequences of different conceptions may be.

- **Pre- and post-COVID 19:** the White Paper was issued in February 2020. This was before the entire EU was affected so hard by the COVID-19 pandemic. Arguably, the priorities and content of the White Paper may need to be revised taking into account this new reality and issues that arose (e.g. use of AI for tracing/contact applications,…). The revised White Paper may, therefore, include an additional part 'AI and COVID-19' or should at least mention how the pandemic influenced its content.

- **Language and terminological issues:** the White Paper should use clear and well-defined terms and concepts. For instance, the subject-matter of future AI regulation is referred to as "AI-based products and services", "AI applications" or "AI systems", which obscures the understanding of the document. Moreover, some additional refinements are necessary as well. The White Paper, for instance, stipulates that "Although Europe currently is in a weaker position in consumer applications and on online platforms, which results in a competitive disadvantage in data access, major shifts in the value and re-use of data across sectors are underway" (p. 4). It remains unclear towards whom the EU may have such a weaker position (e.g. China, the United States,…). The following parts will provide other examples that may require more clarity. Additionally, clarification would be welcome on whether it is the EU as an legal order or companies established in the territory of the EU Member States, that is/are 'lagging behind'.

- **More good practices and use cases:** the revised White Paper may include more use cases, examples and good practices to illustrate the development and design of AI-systems, especially for those with a high risk.

- **AI for good?:** The White Paper seems to assume that AI could lead to 'good' while it also poses risks that regulation should therefore tackle. AI can certainly do many things in a great array of sectors, but nothing demonstrates that AI would 'do good' *per se.* It takes context-specific research to make such an evaluation. For instance, in the mobility sector, it has been shown that the benefits of autonomous vehicles for capacity optimisation and for the environment are doubtful. Such benefits would be obtained only provided a large number of conditions are met, many of which imply strong channelling by the law. This reinforces the need for more use cases and concrete examples.

- **Trust and AI**: The White Paper refers to the need for trust in order to enable the deployment of AI. For instance, it stipulates that "lack of trust is the main factor holding back a broader uptake of AI" (p. 9). In a democratic society, bringing trust should however not be viewed as a means to nudge public acceptance. Contrarily, it is the democratic character of our societies

and our shared 'European values' that should serve as guiding principle why we aim at bringing trust to citizens.

3. FEEDBACK ON THE 'ECOSYSTEM OF EXCELLENCE'

- **Scope of the White Paper – ethical and societal perspective:** the White Paper stresses ethical and philosophical values/concerns when developing and deploying AI-systems. In creating an ecosystem of excellence, an additional research centre (solely) devoted to philosophical and ethical aspects of AI may be considered. The revised White Paper may specifically refer to/mention it. Ethical and societal concerns should thus already be mentioned in the part dealing with the efforts of the research and innovation community (p. 4-5).

- **Scope of the White Paper – scientific perspective:** we welcome the idea of a regulatory framework for high-risk AI-systems and especially the working assumption that it applies to products and services relying on AI. However, the establishment of a regulatory framework should not create obstacles for scientific research on AI, especially not for fundamental, methodological and algorithmic research, in line with the objective set by the White Paper to foster European research on AI. Just like in other areas, a differentiated approach might be necessary between research and commercial applications so that AI regulation does not hinder fundamental research on AI.

- **Different approaches to AI-research**: the revised White Paper may take into account the differences and complexities of/between academic AI research such as top-down approaches (i.e. from methods to applications) and bottom-up approaches (i.e. from applications to theories and methods). It has already been mentioned that concepts used in the revised White Paper should be clearly defined to avoid any confusion and/or misinterpretation. For instance, the concept 'robustness' can have a different meaning according to the field/sector in which it is used (e.g. robust loss functions in machine learning, adversarial robustness, robust statistics, robust control theory, robustness related to security aspects, robustness and reliability, etc.).

  The White Paper points to the need for a European "lighthouse centre of research, innovation and expertise that would coordinate [the efforts of the current fragmented landscape of centres of excellence] and be a world reference of excellence in AI" (p. 6). While acknowledging the need for increased impact of European research, we recall that pluralism in research should also be preserved as a core driving force of scientific knowledge.

  More clarity on international aspects: it is stipulated that the "EU will continue to cooperate with like-minded countries, but also with global players, on AI, based on an approach based on EU rules and values" (p. 8-9). The question remains whether there are any like-minded countries. Criteria are needed to determine when and which countries are 'like-minded', which necessarily imply to clarify what is meant by 'European values'. The importance of finding a consensus on a proper and single definition of AI is stressed again (also see *supra* part 2). As many international companies are involved in the development of AI-systems, more cooperation at an international level regarding the governance, regulation and supervision of the development and use of AI may be required (cf. Global AI Partnership).

4. FEEDBACK ON THE 'ECOSYSTEM OF TRUST'

- **Link between ecosystems of excellence and trust**: it remains unclear how the ecosystem of trust can be linked to the ecosystem of excellence. Thus, there is a need for a better/closer connection between both ecosystems. They can be linked/bridged by adopting new ways and types of regulation (e.g. regulatory sandboxes, Digital Innovation Hubs, living labs,…). Research and innovation should not be disconnected by regulation but instead be connected by using innovative ways of regulation. Those new types of regulation may take into account the needs from the research and innovation community. Governance of AI and procedural issues can be relevant as well as to link excellence and trust (e.g. deciding whether everyone may have access to an AI-system or do we need specific restrictions on the use of a specific AI-system, cf. comment *infra*).

- **European Data Strategy**: there is a complementary relationship between the White Paper and the European Data Strategy. This may be considered as parts and parcels of a whole. Some also argue that the European Commission may consider to first implement the European Data Strategy and create the ecosystem of excellence before addressing the need and establishment of a regulatory framework. Considering the reliance of AI-systems on data, it seems appropriate to first create a single market for data. This market will allow data to flow freely within the EU and across sectors for the benefit of businesses, researchers and public administrations.

- **Linguistic and terminological concerns**: we would once more like to emphasise the importance of using clear language and unambiguous concepts in the revised White Paper.

  For instance, it remains unclear to which extent the ecosystem of 'trust' is the same as/or refers to 'trustworthy' AI (cf. AI HLEG). The requirement of 'transparency' remains unclear as well (p. 9) as its scope will (also) depend upon which specific party is using the AI-system. It should be made clear what exactly is understood under the requirement that an AI-system has to be transparent. The 'Black Box' problem can be addressed from different perspectives and the expected level of transparency may need to be considered/varies per AI-application and per users group (different groups of users, developers and researchers have different backgrounds and may perceive transparency in different ways). The expression 'Trustworthy AI' remains questionable, as trust is traditionally given to persons and institutions and not to 'things'.

- **Risk-based approach to AI**: the European Commission rightly stresses that any regulatory intervention should be targeted and proportionate. That is why it does not aim to regulate all AI-systems but only those with a high risk. Such systems may be subjected to specific (additional) requirements. The existing regulatory framework will continue to apply to all AI-systems regardless of the risks they entail.

  Two essential questions will be to determine (1) what 'high-risk' means and (2) which AI-systems qualify as high-risk and, therefore, subject to additional requirements.

  (1) Although based on the notion of (high-)risk, the White Paper does not properly define risk, for the purpose of the future regulation. The revised White Paper will, therefore, need to

answer questions such as: what is at risk? Are we talking about legal entitlements such as fundamental rights or, broader than that, about 'European values' and ethics values? It should also be clarified who is at risk. Clarifying the notion of risk may therefore lead to clarifying the policy purpose of the AI regulation.

(2) The distinction between high-risk and low-risk AI systems should be further clarified. The use of the two cumulative criteria (i.e. use and sector) may imply that not all AI systems in high-risk sectors will be regulated and AI systems that may be regulated in a certain high-risk sector may not be regulated in another sector. The Commission also acknowledges that the use of AI-systems may in exceptional circumstances be considered as a high-risk by itself, irrespective of the sector where it is deployed. Taking into account the (different) assessment criteria, a situation of legal uncertainty might arise as organisations will probably argue that their AI-systems(s) do not qualify as high-risk. Clear guidelines on the application of these criteria is necessary in the revised White Paper.

The list of sectors will need a careful impact assessment on behalf of the EC. In this respect, the list indicated in the White Paper seems incomplete. To give a taste of other candidates, one may wonder why media and social media are not included, given the risk that some AI applications pose to freedom of expression, which is indeed mentioned in the White Paper. Financial markets may also be sensitive to AI applications, which could in particular lead to systemic risks. Agriculture, and especially data-driven or 'smart' farming, may also be considered a sensitive sector, as it is about food resilience.

More attention should also be given to the difference between intentional and non-intentional risks. A designer may apply all best efforts when developing an AI-system but non-intentional/intended risks cannot be entirely/totally prevented or excluded. AI-systems or methods can be used for 'good' purposes but, nevertheless, contain a risk when deployed differently. Mechanisms could be established allowing the developer of an AI-system to already indicate for which specific purposes the system can and especially cannot be used. It is conceivable that a designer may not want to have its algorithm/AI-systems being used for specific purposes. Sanctions may be imposed when users deploy the AI-system it in another way as was envisaged by the developer. The possibility for developers to impose restrictions on the specific use of AI-systems may be added to the illustrations (p. 18) in the revised White Paper.

The provision that "remote biometric identification" and other "intrusive surveillance technologies" (p. 18) will always be considered as high risk undermines the premise that technology is neutral when being developed but that its use/application makes it 'good' or 'bad'. There is thus a danger of generalizing the misuse of an AI-systems to give a 'bad name' to a specific technology. This may be considered in the revised White Paper and is actually closely related to the possibility of the developer to influence the use and application of its AI-system (cf. *supra*). Some argue to make a distinction between "good"/"bad" on the one hand and "high-risk" on the other. These may be two different things. High-risk AI-systems are not necessarily "bad"

Several other issues remain as well with regard to the risk-based approach. For instance, it is unclear who and at which level it will be determined whether an AI-system is low-risk or high-risk. Will this be decided by the European Commission or by Member States? It is also not always possible to anticipate the risks of an AI-system in advance. Whereas it can be foreseen

that some AI-systems will incur a high(er) risk for users, this not always is the case for other applications. Additionally, the high-risk may depend upon the environment in which the AI-system is deployed, embedded and used. In this respect, having AI-systems as the subject-matter of an AI regulation may lead to underestimate additions of small risks, which could altogether lead to high risks in certain environments. Specific safeguards should be included in order to set the right incentives and take into account such risks.

- **Adapting the legal framework in context AI**: it has already been mentioned that the existing regulatory framework will continue to apply to all AI-systems regardless of the risks they entail. The applicable legal framework may, however, need some adjustments to sufficiently address the risks created by AI systems. The legal framework may be improved, for instance with regard to the uncertainty of the allocation of liability between different economic actors, the distinction between services and products or the changing functionality of AI-systems.

  This means that it should first be established which regulations already apply to AI-systems such as the General Data Protection Regulation (GDPR), product safety/liability legislation (cf. *de lege lata* analysis). It remains unclear which criteria will be used to determine whether the existing framework sufficiently addresses the risks created by AI systems. In other words, one needs to find and establish specific evaluation criteria to assess whether the existing legal framework is able to sufficiently cover the many consequences of AI and if this is not the case, which criteria can be used to provide regulatory solutions and overcome existing gaps (cf. *de lege ferenda* research).

- **Liability:** the White Paper and its accompanying report on liability rightly acknowledge the limitations of current liability frameworks raised by AI-systems and applications (p.15). It insists on the fact that some of AI's features may render it difficult for the victim to identify the liable party and effectively claim compensation. It rightly considers possible amendments to the Product Liability Directive as well as "targeted harmonization of national liability rules" (p. 15). With regards to this latter element, the White Paper would gain more clarity if it detailed a bit more what national liability rules it refers to and in which context. Indeed, notwithstanding general fault-based liability that can be found in almost every country, different sectoral liability regimes may exist (for instance in road traffic).

- **Types of requirements for high-risk systems**: it should be taken into account that a trade-off may need to be made between 'robustness' and 'accuracy'. When making an AI-system more robust, accuracy may lower. Increasing robustness may thus reduce accuracy. There is a risk that an AI-system which is easy to understand may not be the 'best' performing one, even though developers have a moral duty to make methods and AI-systems as 'good' as possible. It also remains unclear how far the obligation of explainability (transparency) goes and towards which group of users it applies. This should be thoroughly addressed in the revised White Paper as it will determine/influence the level of explainability (p. 18). The Commission should consider whether the obligation of transparency for AI-systems is a 'hard' or 'soft' one, and especially provide guidance on the balance between transparency and robustness and its consequences (p. 20). These issues may call for context-specific answers, which may be a limitation to the horizontal character of the AI regulation. The same can be said about the human involvement. The White Paper rightly categories different types of human involvement,

which may have to be tailored according to the expected purpose of the AI application and to the sector.

Regardless of the technical discussion and required level of transparency and explainability, it should be made clear how decisions taken by AI-systems affect users. Fundamental rights should be taken into account when developing AI-systems. If someone develops/deploys AI systems that may take decisions affecting human rights, users should be entitled to a proper explanation.

When referring to "human oversight", the EC also refers to "ensuring appropriate involvement by human beings in relation to high-risk AI applications" (p. 21). It remains uncertain how "appropriate" should be defined. The White Paper may expand the requirement to a "meaningful human control" for AI-systems used in high-stakes situations. In other words, the White Paper should specify that the human oversight should not be limited to just validating the decisions by AI-systems, but it should rather be ensured that the human-in-the-loop is in a position to appreciate the limits of the system (to mitigate the risk of over-reliance on the system's decisions) and also is aware of the importance of his/her intervention in the AI-assisted decision-making (for which she/he can/will be held accountable).

- **Risk of fragmentation**: it is acknowledged that if "the EU fails to provide an EU-wide approach, there is a real risk of fragmentation in the internal market, which would undermine the objectives of trust, legal certainty and market uptake" (p. 22). Although this underlines the importance of a supra- and international approach towards AI, the division of competences between the EU and its Member States has to be taken into account as well. In this respect and given the horizontal character of the AI regulation, it should be clarified that it would consist in minimal harmonisation, so that Member States will remain allowed to enact further legislation in order to deal with *specific* challenges of AI applications.

- **Prior conformity assessment**: The European Commission envisages the creation of prior conformity assessments in order to ensure compliance of high-risk AI-systems with the listed mandatory requirements. These prior assessments could include procedures for testing, inspection or certification. The introduction of certification schemes for AI-systems should be carefully addressed: what exactly can and will be certified (process vs. system), who will certify (cf. public/private bodies), impact on market access for new players and on competition,…

  Moreover, the legal consequences and the (intended) impact of a certificate given to an AI-system remain unclear. The revised White Paper should clarify the legal value of such certificates. Would they guarantee compliance with the AI regulation provisions? Would they even consist in a form of prior authorisation before deployment or placing on the market? Or would they consist in a presumption of conformity, similarly to the regulatory system in place in technical harmonisation of products where compliance with EU norms grant a presumption of conformity with legal essential requirements?

  Certification of AI-systems should not become a lucrative business for companies. Certification of AI-system needs be effective and adequate with sufficient controls and safeguards on certification entities. Although we endorse testing and certification schemes, the certification process should not become an expensive/'fake' bureaucratic process. Experiences from other

sectors in which certifiers provide their services also illustrate that several (legal) challenges remain (cf. immunity, liability, public role,...). These challenges can have an impact on the quality of the certification process. Regulators should thus be aware of these (legal) challenges when adopting a certification scheme in the context of AI.

- **Voluntary certification**: voluntary certification can have several benefits, both for purchasers of the certified AI system as well as for its producer. Such certification increases the confidence of users in AI systems as it indicates the producer's commitment towards higher safety and quality standards. At the same time, however, voluntary certification should be carefully addressed in the revised White Paper as it can result in a meaningless label, and even increase non-compliant behaviour when no proper verification mechanisms are established. Control by authorities should still be organised on voluntary labelling schemes (cf. cybersecurity scheme).

- **Practical concerns:** several practical concerns remain as well such as the temporal application of the regulatory framework that may be adapted for high-risk AI systems (cf. retroactivity). This issue may need to be addressed in the revised White Paper.

5. CONCLUDING REMARKS

In this report, CiTiP and Leuven.AI provided feedback on the White Paper on AI issued by the Commission in February 2020. We would like to stress again that the **EU's overall approach to AI is positive and a step in the right direction**. We believe that a **coordinated supranational approach** to the many benefits and challenges created by AI-systems is necessary. We look forward to the publication of the revised White Paper on AI and hope to **further collaborate** on establishing a proper (regulatory) framework on AI, taking into account all interests including those of the research communities.