Kenniscentrum Data & Maatschappij

FEEDBACK ON EC WHITE PAPER ON ARTIFICIAL INTELLIGENCE

Knowledge Centre Data & Society

June 2020

1. INTRODUCTION

This report contains the feedback following a consultation organised by the **Flemish Knowledge Centre Data & Society** (KCDS) on the White Paper on Artificial Intelligence (AI) issued by the European Commission (EC) in February 2020. The KCDS focuses on the interplay between data, AI and society. It enables socially responsible, ethical, and legally appropriate implementations of AI in Flanders. The KCDS aims to enable Flemish companies, policymakers, regulators, and citizens to achieve the greatest social and/or economic benefits of AI. It brings together diverse representative organisations/ stakeholders and facilitates the creation of tools, advice, and recommendations. It comprises three existing research centers: imec-SMIT (Vrije Universiteit Brussel), imec-MICT (Ghent University) and the Centre for IT & IP Law (KU Leuven). The Flemish Department on Economy, Science and Innovation funds the initiative.

The KCDS organised an online feedback session on Thursday 4 June 2020. The following stakeholders and persons participated: Rik Hendrix (VITO Data Science Hub), Jelle Hoedemaekers (Agoria) Elisabeth Geenen (Vlaams ABVV), Lode Lauwaert (KU Leuven Institute of Philosophy), Erik Mannens (Ghent University IDLab), Astrid Eichstädt (Unia), Guido Van Humbeeck (VDAB), Dieters Somers (VOKA), Astrid Eichstadt (Unia), Jan De Bruyne (KU Leuven CiTiP) and Rob Heyman (KCDS). Against this background, the KCDS 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

- **Military purposes**: the White Paper stipulates that it does not address the development and use of AI for military purposes (p. 1). It remains unclear why this topic is not addressed in the White Paper. The EC may consider addressing issues related to dual use of AI-systems (civilian/military) and Lethal Autonomous Weapons Systems (LAWS) more thoroughly in its revised version. However, we are aware that this is a complex and important topic, which may also/even be addressed more thoroughly in other (policy) documents.

- **Language and terminological issues:** the White Paper should use clear and well-defined terms and concepts. Some additional refinements are necessary as well. For instance, it is stipulating that "Europe's current and future sustainable economic growth and societal wellbeing increasingly draws on value created by data" (p.1). It remains unclear what is meant with societal wellbeing (e.g. does it relate to health or social equality?). Likewise, the White Paper mentions that it is "vital that European AI is grounded in our values and fundamental rights such as human dignity and privacy protection" (p. 2). Again, it remains unclear what is meant with the concept 'human dignity' as it can have different meanings/interpretations.

- **Definition of AI:** the White Paper mentions that "AI is a collection of technologies that combine data, algorithms and computing power". This seems to be an appropriate approach as it is wider/broader than the definition used in the documents issued by the High Level Expert Group on AI. The latter especially focuses on machine learning, whereas a focus on semantics and AI is relevant as well. At the same time, however, some problems remain. For instance, it is not a workable/operational definition because this definition is quite broad, so broad that regular software could also fall under its scope. It may make more sense to talk about scale, actions and consequences than infrastructure if we want an AI definition that can hold up for a longer time in the future. The EC also relies on other definitions/concepts of AI in subsequent parts in the White Paper (compare p. 2 & 16 with footnotes 46 & 47).

- **Costs related to the use of AI:** according to the White Paper, an AI ecosystem will have advantages for services of public interest, for example by reducing the costs of providing services (p.2). However, the focus may also be more oriented towards efficiency and/or sustainability created by AI-systems, the need for a proper legal framework, and data and computing infrastructure (and not only on costs reduction).

- **European Strategy for Data:** 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.

  Moreover, it is stipulated that the Strategy aims to enable Europe to become the most attractive, secure, and dynamic data-agile economy in the world. It should be ensured that this transformation results in sustainable jobs, taking into account the needs of employees as well. Data should also be made available to/shared with more organisations and/or to the larger public. Data cannot solely remain under control of the 'traditional' big tech companies. In this digital transformation, socio economical as well as legal and ethical aspects should also be considered.

  The White Paper also seems to suggest that European AI algorithms should be based on European data. European data however is not necessarily/sufficiently representative, accurate or technically robust. Setting it as a benchmark endangers the objectives of fairness and diversity. The cost of retraining algorithms created elsewhere in the world on EU data can also be significantly high, especially for SMEs. It will also hamper trade relations and could lead to more protectionist measures.

- **International cooperation**: the White Paper mentions that "Harnessing the capacity of the EU to invest in next generation technologies and infrastructures, as well as in digital competences like data literacy, will increase Europe's technological sovereignty in key enabling technologies

and infrastructures for the data economy" (p. 3). The need for international cooperation should be stressed as well, together with Europe's ambitions regarding the development of AI. 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.

- **Actions:** it is unclear why specific actions are not included in points G-F (p. 8-9) and in the part on the ecosystem of trust (p. 9 and further). One reason may be that the listed actions are currently already being pursued, while establishing specific actions for regulatory aspects is innovative and, therefore, more challenging and complex. Nevertheless, the need for a proper risk-based regulatory framework for AI is emphasised at several occasions in the White Paper. As such, some specific actions are welcome in the revised White Paper. It seems that the White Paper pursues a twofold approach. On the one hand, it wants to enhance the uptake and promotion of AI within the ecosystem of excellence ('increasing speed'). On the other hand, it seems to slow down that idea by the establishment of a proper regulatory framework in the ecosystem of trust ('hitting the brakes'). The link between the two parts should be made clear and more explicit in the revised version (cf. also comment infra).
- **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 (e.g. more need for international cooperation and social innovation).

3. FEEDBACK ON ECOSYSTEM OF EXCELLENCE

- **Skills (part C):** the White Paper seems to especially focus on developing skills within universities and higher education institutes. However, it should be ensured that workers and employees should also acquire the necessary skills in the evolution towards a digital society and workplace. The White Paper rightly mentions that "workers and employers are directly affected by the design and use of AI systems in the workplace. The involvement of social partners will be a crucial factor in ensuring a human-centred approach to AI at work" (p. 7). Nevertheless, we take this opportunity to stress the need to include employees in the digital transformation and pro-actively focus and think of their position in the automatisation of many processes (e.g. via retraining programs, lifelong learning, focusing on social innovation and other ways of meaningful work,…). Likewise, it should be ensured that officials working at public bodies and/or regulatory authorities possess the required skills to assess compliance of the development and use of AI-systems with the applicable (regulatory) requirements. More generally, social dialogue will be crucial to ensure that employees and workers are properly informed as well as trained to successfully face and overcome digital challenges. We further suggest increasing digital literacy for those who are affected by AI. More in particular, skills and education should educate people about their rights. On education – equally important to address also the "supply side" so that developers and deployers of AI are

trained on understanding equality and human rights standards and this way will be knowledge to "supply" AI systems, which are compliant with these standards. Equality bodies could be key "educators" for AI suppliers in both the private and public sector.

- **Focus on SMEs (part D):** SMEs will be the most affected by new rules or legislation. They have limited resources and often no legal expertise/knowledge. The EC should focus more on how these firms could be supported and how the Digital Innovation Hubs can play an instrumental role here (or other initiatives that will be launched under the umbrella of Digital Europe or Horizon Europe).

- **Societal aspects and Sustainable Development Goals (SDGs):** the revised White Paper should focus more on the societal impact of AI research and AI-related projects. If innovation should have an ethical/trustworthy element, the societal impact of AI should be taken into account as a KPI in allocating public funding for projects. The White Paper rightly stresses that "It is also clear that the responsible development and use of AI can be a driving force to achieve the Sustainable Development Goals and advance the 2030 Agenda" (p. 9). However, the importance of the SDGs and the way in which AI can contribute to achieving them should be given a far more prominent role in the White Paper. Likewise, the impact of AI on achieving the Green Deal may be explicitly acknowledged and addressed. Instruments and practical guidelines are necessary to show how AI can contribute to the SDGs. An idea could be to adopt a checklist for AI-related projects, which indicates compliance with/contribution to a (specific) SDGs. More generally, the White Paper may identify the priorities and goals for which AI can be deployed ('dreams'). Funding can then be given to those AI-related projects and research programs that contribute to these priorities and goals. Social partners should be involved when determining the priorities and goals. Future advances in Research & Development and scientific innovation, which underpin and enable AI development, should be based a clear agenda or program, which is led by equality and fundamental rights considerations. In practice, this means that finance for research and innovation in AI goes only for uses which provenly have no negative implications for equality. Financial incentives through increased investment should be created for AI-related research and science for which there is evidence that it will improve the current situation with respect to equality & fundamental rights protection

- **Securing access to data (part G)**: although improving access to and the management of data is indeed fundamental, a digital divide should be prevented. The White Paper should propose actions to reduce the gap between individuals and/or big tech companies who have access to modern information and communication technology and those who lack access including some SMEs. Although it is addressed more thoroughly in the European Strategy for Data, the White Paper may explicitly acknowledge this ambition as well.

4. FEEDBACK ON 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).

- **General approach regulation AI**: it is recommended to focus first on existing regulation instead of creating additional legislation, which might result in more red tape as well. The focus should first be placed on updating and clarifying existing legislative documents and directives. To avoid red tape, the use of pragmatic tools and conformity checks is favoured.

- **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, or what, is at risk. And lastly who assesses the risks. 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. AI risk is not binary and as such it is very likely that a large number of moderately risky AI applications will fall in the 'high-risk category' and will be subject to disproportionate requirements. Clear guidelines on the application of these criteria is thus necessary in the revised White Paper. An option may be to include some specific scenario's, examples or use cases that would qualify as high risk.

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? Who will assess these risks and also what will be done if an ethical board identifies risks, but the board of directors decides to ignore these? We propose to have a dispute committee for these challenges. 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. A first suggestion would be to focus more on the ultimate context of an AI-system instead of the current focus on the model generated through machine learning.
The high number and wide diversity of situations where the use of AI systems has been identified as problematic by the Report suggests an crucial consideration in response to the above question: with respect to equality and fundamental rights, a risk-based approach strictly limited to particular sectors and AI uses will not provide adequate protection against potential rights violations by AI systems.
With regard to the above, our position on risk is that some differentiation is both necessary and inevitable for both economic (burden to businesses) and practical (feasibility, high/growing number of AI uses) reasons, but the exemption threshold for excluding AI applications from scrutiny should be set high (i.e. only the most patently harmless, e.g. speed regulation of manufacturing belt lines, as they produce beyond reasonable doubt no negative effects on equality, and other fundamental rights). Thus, when it comes to equality and AI, the widest possible range of AI applications should be subject to compliance checks and only the most "conspicuously innocuous" AI uses should be exempted from such control.

Next to a risk-based approach, it is imperative that a longer "testing period" should be afforded at the users' end, allowing those affected by AI systems to "test' the adequacy of existing equality and fundamental rights legislation by claiming their rights and seeking redress. That is why, the Commission's strategy on AI should be adjusted to place greater emphasis on those potentially affected by AI-enabled products and services and more specifically, on enhancing their ability to identify and seeks redress against breaches of equality and other fundamental rights.

- **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).
The approach currently suggested in the White Paper risks to imply an untenable hierarchy of rights with some rights, such as the rights to respect for private life and protection of personal data, "ranking" higher in their entitlement to common EU legal protection against AI-specific risks than others. Instead, we suggest that a common European approach on the ethical and human implications of AI should level up existing common minimal standards for the protection of equality and fundamental rights under EU law, thereby imposing AI—tailored regulatory requirements on all uses of AI systems with relevant impact on equality and fundamental rights.

- **Limitations of scope of existing EU legislation:** the White Paper focuses on EU product safety legislation and points out that this legislation is limited to products and not services. Our panel wants to see a clear recommendation on the level of product and/or service safety so that the following question might be answered: What is it we need to prove in order to show that we are compliant with product safety legislation for AI?

- **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 system**s: the question arises whether (additional) horizonal legislation including requirements for high-risk AI-systems is required. Such requirements may already exist in the applicable framework. Moreover, a distinction can be made between the training of an AI-system on the one hand and the use of the system on the other hand. An AI-system that complies with privacy/data protection requirements may need a training of data that by itself does not necessarily comply with privacy/data protection requirements. The training of an AI-system may thus to a certain extent be disconnected from its deployment. The self-learning capacities of AI-systems especially occurs during the training stage. Concerns with regard to privacy/data protection may not be decisive in the training stage of a system, especially if it pursues another goal (e.g. ensuring the realisation of one of the SDGs). More generally, it should be examined which (regulatory) obstacles may impede the development of AI-systems and what remedies are possible. The General Data Protection Regulation, for instance, provides fundamental protection regarding the processing of personal data, but can also have an impact on the development of AI-systems.

- **Robustness and accuracy**: requirements on robustness and accuracy should not lead to source code disclosure. This might stifle innovation as firms will be concerned about the protection of trade secrets (and it might prevent them from capitalizing on their IP). Contextual algorithmic audits and training data review might perform better in addressing risk.

- **Transparency and explainability**: the section on transparency is too vague and guidance on the required level and scope of transparency is unclear, this will hamper implementation efforts.

  Explainability can be a challenge for many AI-systems, both in terms of feasibility and practicality. It is will be very difficult to come up with detailed explanations on how the outputs of AI systems are provided, especially for many deep learning systems. Furthermore, it will hamper the development of advanced algorithms that offer high levels of accuracy but cannot easily be explained. Some have argued to work more towards algorithmic accountability, which implies that the AI-system includes a variety of controls to make sure the operator can verify the algorithms work in accordance with its intentions and can identify and rectify potentially harmful outcomes.

  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. The EC could play a role in providing guidelines on how to properly integrate human rights into AI-systems.

- **Human oversight**: 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). The requirement of human oversight should, however, be ensured for all AI-systems, regardless of the risks they entail. The White Paper may 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).

- **Addressees**: the White Paper stipulates that "It is the Commission's view that, in a future regulatory framework, each obligation should be addressed to the actor(s) who is (are) best placed to address any potential risks" (p. 22). This approach seems to imply that everyone is able to understand and assess the working of AI-systems, which remains wishful thinking. It is not realistic to expect that all users understand the working of AI-systems.

- **Compliance and enforcement:** Each level of governance slows down development and raises the question who will be responsible for independent assessments. There should be room for experiments in a lab setting before an entire risk assessment is performed.

- **Prior conformity assessment**: The EC 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). The revised version of the White Paper may stress once more the importance that all AI-systems should comply with the applicable regulatory requirements.

- **Governance**: the White Paper rightly stresses the importance of a proper European governance structure (p. 24). At the same time, however, it should be determined which authorities will eventually supervise compliance of AI-systems with the applicable regulatory requirements. It should be clear for users to which authorities users can turn when encountering problems with AI-driven systems, services or products. A question in this regard is whether this should be organised at the local, national or supranational level. It may also be organised at all levels, but it should be clear for users/consumers who they should go to upon encountering problems.
  It is important to ensure that protection against AI-related threats to fundamental rights is not diminished or obstructed by having multiple and potentially conflicting decision-making bodies on AI-driven discrimination. The institutional division of labor in regulating algorithmic

biases should be functional, based on the specific expertise and suitability of a particular regulatory actor for the task at hand.

We also note that the division of labor within a decision-making body on AI and fundamental rights requires attention as human rights organizations will have to work with engineers and data scientists to come to a conclusion based on research of the latter.

5. CONCLUDING REMARKS

In this report, the Knowledge Centre Data & Society 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.