CONSULTATION ON THE WHITE PAPER ON ARTIFICIAL INTELLIGENCE – A EUROPEAN APPROACH

Submission of the

Chair Legal and Regulatory Implications of Artificial Intelligence

MIAI (Université Grenoble Alpes)

JUNE, 2020

THEODORE CHRISTAKIS
KARINE BANNELIER
MATHIAS BECUYWE
STEPHANIE BELTRAN GAUTRON
MAÉVA EL BOUCHIKHI
KATIA BOUSLIMANI

AI-REGULATION.COM

MIAI Grenoble Alpes

Consultation on the White Paper on Artificial Intelligence – A European Approach

Submission of the

Chair Legal and Regulatory Implications of Artificial Intelligence

MIAI (Grenoble Alpes)

June , 2020

Introduction

The Chair on the Legal and Regulatory Implications of Artificial Intelligence (Chair AI-Regulation) was set up in September 2019. The Chair has been chosen by an international panel of experts to be part of the Multidisciplinary Institute in Artificial Intelligence (MIAI) created atthe Université Grenoble Alpes.

MIAI Grenoble Alpes Institute aims to conduct research in artificial intelligence at the highest level, to offer attractive courses for students and professionals of all levels, to support innovation in large companies, SMEs and startups and to inform and interact with citizens on all aspects of AI.

AI has the potential to make breakthrough advances in several areas, but its growing applications raise complex questions and provoke broad concerns throughout society. The main mission of the Chair AI-Regulation is to research how regulation can support sustainable and ethical innovation.

Chair's members are experts in law, economics, computer and data science, all actively working in the field of data protection, privacy, cybersecurity and AI. The Chair's objective is to become a valuable point of reference regarding the legal and regulatory questions raised by artificial intelligence and to contribute to national, European and international debates on these issues.

The present submission has been drafted by Professors Theodore Christakis and Karine Bannelier with the contribution of the four research fellows of the Chair: Mathias Becuywe, Stephanie Beltran Gautron, Maéva El Bouchikhi and Katia Bouslimani. The authors would like to thank all the members of the MIAI (Grenoble Alpes) for their input and comments during the preparation of this contribution.

The Chair on the Legal and Regulatory Implications of Artificial Intelligence is currently conducting several research projects in relation with the topics covered by this consultation and would be very glad to contribute to future activities of the European Commission, the European Parliament and, more generally, European and national Institutions on these issues.

For more information about us visit our website: https://ai-regulation.com

Our website aims to become a forum to provide some answers to the hard questions related to the legal and regulatory implications of AI and share the results of our research as well as insights on these issues from external collaborators and contributors. We publish substantive articles and reports as well as brief notes and news updates on worldwide developments in AI regulation.

Section 1 - An ecosystem of excellence

To build an ecosystem of excellence that can support the development and uptake of AI across the EU economy, the White Paper proposes a series of actions.

In your opinion, how important are the six actions proposed in section 4 of the White Paper on AI? (1-5: 1 is not important at all, 5 is very important)

Working with Member states	**5**
Focusing the efforts of the research and innovation community	**5**
Skills	**4**
Focus on SMEs	**4**
Partnership with the private sector	**4**
Promoting the adoption of AI by the public sector	**4**

We welcome the European Commission's awareness on the need to establish a better cooperation with and between EU Member States on AI and its regulatory and legislative frameworks. Some Member States are currently working on national legislation to frame the development and deployment of AI systems. A lack of common knowledge and harmonization of the different rules applying to these technologies could lead to a fragmentation of the market and legal uncertainty. This, in turn, would be counterproductive for the development of AI in Europe. Some Member States are already ahead in terms of technological advances or AI expertise. A lack of global European cooperation or common focus on AI innovative research could leave behind part of the European family. The willingness of the Commission to focus on the societal and environmental well-being as a key principle for AI is also welcome.

There is a real need to focus, help and support SMEs specialized in AI and ensure that all SMEs can have access and benefit from AI.

Partnership with the private sector could be an important action but should be done under win-win conditions. For instance, this should not lead to a unilateral transfer of public data (such as health data) towards the private sector without sufficient protections and real counterparts.

Promoting the adoption of AI in the public sector is also an important action as this could help improve public services for the benefit of all. However, promoting the adoption of AI by the public sector should not be a standalone goal. Careful consideration must be given to the need for an AI system in a specific field and to the eventual associated risks.

Are there other actions that should be considered?

The EU has an important role to play at the international level in this field. It should cooperate with other regional and international organisations in order to promote international

cooperation, environmental-friendly innovation and AI initiatives that put the protection of fundamental rights and European rules of values at their very core. Europe could act as a global leader in this field, as it did with data protection.

Revising the Coordinated Plan on AI (Action 1)

The Commission, taking into account the results of the public consultation on the White Paper, will propose to Member States a revision of the Coordinated Plan to be adopted by end 2020.

In your opinion, how important is it in each of these areas to align policies and strengthen coordination as described in section 4.A of the White Paper?

(1-5: 1 is not important at all, 5 is very important)

Strengthen excellence in research	**5**
Establish world-reference testing facilities for AI	**5**
Promote the uptake of AI by business and the public sector	**4**
Increase the financing for start-ups innovating in AI	**5**
Develop skills for AI and adapt existing training programmes	**4**
Build up the European data space	**5**

In your opinion how important are the three actions proposed in sections 4.B, 4.C and 4.E of the White Paper on AI? (1-5: 1 is not important at all, 5 is very important)

Support the establishment of a lighthouse research centre that is world class and able to attract the best minds	**3**
Network of existing AI research excellence centres	**5**
Set up a public-private partnership for industrial research	**4**

Supporting the establishment of a lighthouse center of research, innovation and expertise that will coordinate networks and centres dedicated to AI might be interesting. However, issues might arise about where this centre will be located and how it would interact with national centres in Europe already very active in the AI field. Helping to ensure a functional and performing network of national excellence centres should be the priority. As members of one of these centres, the Multidisciplinary Institute in Artificial Intelligence (MIAI, Grenoble Alpes, France) we will greatly welcome any European projects permitting to enhance pan-European cooperation on these issues and develop common transdisciplinary projects. Europe should be very careful not to neglect the importance of social sciences in relation with future AI-related research projects.

Focusing on Small and Medium Enterprises (SMEs)

The Commission will work with Member States to ensure that at least one digital innovation hub per Member State has a high degree of specialisation on AI

In your opinion, how important are each of these tasks of the specialised Digital Innovation Hubs mentioned in section 4.D of the White Paper in relation to SMEs? (1-5: 1 is not important at all, 5 is very important)

Help to raise SME's awareness about potential benefits of AI	**4**
Provide access to testing and reference facilities	**4**
Promote knowledge transfer and support the development of AI expertise for SMEs	**5**
Support partnerships between SMEs, larger enterprises and academia around AI projects	**5**
Provide information about equity financing for AI startups	**5**

These tasks are all important as promoting the uptake of artificial intelligence for SMEs could boost innovation in Europe. However, at the same time, it is important to ensure that SMEs are aware of the potential risks of AI systems and the ethical tensions surrounding them. Also European SMEs and start-ups need access to financing. From this point of view they often find themselves in a competitive disadvantage in relation with their counterparts in the US or Asia. Supporting partnerships between SMEs, larger companies and academia could be a particularly welcomed action as experience has shown the richness and usefulness of the resulting cross-fertilisation.

Section 2 - An ecosystem of trust

Chapter 5 of the White Paper sets out options for a regulatory framework for AI.

In your opinion, how important are the following concerns about AI (1-5: 1 is not important at all, 5 is very important)?

AI may endanger safety	**4**
AI may breach fundamental rights (such as human dignity, privacy, data protection, freedom of expression, workers' rights etc.)	**5**
The use of AI may lead to discriminatory outcomes	**5**
AI may take actions for which the rationale cannot be explained	**4**
AI may make it more difficult for persons having suffered harm to obtain compensation	**5**
AI is not always accurate	**3**

A. ***AI may endanger safety (4)*** : Some AI systems could become dangerous if their safety is not clearly established or if they are hacked or otherwise misused. This is evident, for instance, for embedded systems in autonomous cars which could be either “hacked” or lead to accidents due to flaws in their object recognition technology. Another example is the hypothesis of “home AI” controlling entrance doors and windows hacked by a hostile person. AI systems must be robust and based on the principle of safety by design. Safety can be ensured by the establishment and respect of clear cybersecurity rules and standards as well as precautions against accidental or deliberate misuse of AI systems. On the other hand, it must also be emphasized that AI can enhance safety and security. For instance, well-built digital identity systems could help improve security in relation with identification of persons, online connexions and economic or other transactions.

B. ***AI may breach fundamental rights (5)*** : The misuse of AI can endanger several human rights. Facial recognition technologies, for example, could be used for bulk surveillance greatly endangering the right to privacy. Freedom of expression could fall victim of AI-driven content moderation and “filtering machines” unable to understand context. Systems taking eventually important decisions on the basis of “emotion recognition” could be based on flawed scientific premises and lead to discrimination and flawed decisions. The breach of fundamental rights can happen both at the individual and collective level. For instance, the misuse of data and AI can seriously hamper the democratic electoral process, as has been demonstrated by the Cambridge Analytica case occurring during the presidential American elections in 2016, or as shown in the Commision’s report on

disinformation practices. It could lead to the "manipulation" of humans and influence operations.

The objectives and construction of AI systems should thus be based on existing human rights, ethical principles and solid scientific standards. Measures should also been taken against unauthorized and abusive function creep.

C. ***The use of AI may lead to discriminatory outcomes (5)*** *:* Databases used by AI systems might be biased or based on pre-existing incomplete or biased data. Furthermore, the way algorithms are made could also be biased or lead to discriminatory outcomes. There is a generalized awareness about the risk of bias into databases and AI systems. Nevertheless, ensuring fairness and having unbiased databases could sometimes be difficult as AI systems might use pre-existing databases with pre-existing biases based on ethnic characteristics, gender, etc.

Furthermore, we should focus not only on discriminatory AI but also on discriminatory use of AI. For instance, facial recognition can be discriminatory in itself (because the training sets or relevant databases might be biased). But *the use of facial recognition* could also be discriminatory if, for instance, it becomes a tool for the surveillance of a specific ethnic group. Interestingly, already-discriminated populations could become the focus of enhanced surveillance in the very name of reducing discrimination in the datasets (ex: over-monitoring black people in order to have more data to eliminate bias toward black people).

In a more positive note, well-done IA systems could help reduce or eliminate human bias in some contexts. They could also help improve the condition of disabled persons.

D. ***AI may take actions for which the rationale cannot be explained (4)*** *:* Explicability is very important in some contexts (for instance when the use of AI systems leads to decisions affecting persons and their rights). However, we should not ignore the complexity of some AI systems that may lead to actions for which the rationale cannot always be easily explained. In their study for the European Parliament, Claude Castelluccia and Daniel Le Metayer, affiliates to the AI Regulation Chair, underlined that "ADS are often complex systems that are difficult to understand". They added that "ADS that are based on machine learning are even more challenging to understand, and therefore to explain, since their models are generated automatically from training data".

Technical issues could be attenuated through several mechanisms mentioned in the study, such as the collaboration with the scientific community, the authorisation of reverse engineering or the implementation of explainability-by-design. Yet, those solutions might imply some legal adjustments. For instance, "a key condition is the possibility to provide the research community with access, under specific conditions and the strictest confidentiality, to datasets held not only by public entities but also by private companies". Similarly, the two researchers argue that "reverse engineering should not be limited by intellectual property and trade secret issues. As Articles 3 and 4(3)c of the Directive 2016/943 could be in opposition in the case of a contractual prohibition of reverse engineering, the EU should clarify that reverse engineering is not possible to contractually prohibit".

Concerning this last issue, it is important to note that the interplay between transparency principles and intellectual property rights is a serious issue in the digital environment. For instance, recital 63 of the GDPR states that the right to access "*should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software"*. This consideration also applies to the necessity of explainable AI because, as the OECD highlighted, "*without access to the code, there are only limited ways to examine the validity and reliability of the tools"*.

Another interesting source on the issue of explicability is the Guidance provided by the United Kingdom's Data Protection Authority, the ICO, with the Alan Turing Institute on

Explaining decisions made with AI. This guidance provides practical advice to help explain the processes, services and decisions delivered or assisted by AI, to the individuals affected by them.

E. ***AI may make it more difficult for persons having suffered harm to obtain compensation (5)*** *:* As discussed in Section 3 *infra*, the specific characteristics of AI systems, including their complexity, connectivity, opacity, autonomy and vulnerability to cyberattacks, could make it more difficult for victims to present a claim of compensation or establish the causal link between the victim's harm and the defendant's action. Moreover, it could be interesting to think about the concept of "invisible" harm and modification of human attitudes caused by the use of very intrusive AI technologies such as facial recognition.

As noted by MEP A. Voss in his 2020 Draft Report on Civil liability regime for artificial intelligence to the European Parliament: "Using AI-systems in our daily life will lead to situations in which their opacity ("black box" element) makes it extremely expensive or even impossible to identify who was in control of the risk of using the AI-system in question or which code or input has caused the harmful operation. This difficulty is even compounded by the connectivity between an AI-system and other AI-systems and non-AI-systems, by its dependency on external data, by its vulnerability to cybersecurity as well as by the increasing autonomy of AI-systems triggered by machine-learning and deep-learning capabilities".

F. ***AI is not always accurate (4)*** : Accuracy is important, but requiring from AI systems to be "completely accurate" would be illusory and could also pose a very high and costly standard that European SMEs could sometimes be unable to meet. Do we really impose such high standards to humans when they perform (often in an imperfect way) similar tasks? It is necessary to tailor these requirements in relation with the AI sector concerned, the expected outcomes and the risks for human rights. It is especially important to think about requirements ensuring that AI systems can adequately deal with errors or inconsistencies during all life cycle phases and also focus on the reproducibility of outcomes.

Do you think that the concerns expressed above can be addressed by applicable EU legislation? If not, do you think that there should be specific new rules for AI systems? (Highlight the chosen answer)

- Current legislation is fully sufficient
- **Current legislation may have some gaps**
- There is a need for a new legislation
- Other
- No opinion

As shown in page 13 of the White Paper, several EU legal instruments are applicable in the field of AI - not to mention international and national rules and standards. The current legal and regulatory framework in Europe thus already includes an important number of technology-neutral rules and provides satisfactory solutions to several problems.

However, this framework might need to be adapted sometimes in order to address new challenges, created by AI and rapid technological developments. Improvements to the current regulatory framework are welcomed in order to prevent/mitigate/exclude a series of AI-related

risks. This adaptation should take place in a careful way, following a risk-based assessment and a sector by sector approach. There are, indeed, several ways to improve the current regulatory framework.

i. **Fragmentation among European countries in the interpretation and application of these rules should be avoided**. Our research within the AI Regulation Chair shows that the existing rules are not always interpreted in a harmonized way. For instance, while the concern about an eventual fragmentation of the GDPR is a more generalized problem, it appears particularly important in the field of the relation between data protection and AI. The European Data Protection Board could play a major role in issuing guidelines about how exactly the GDPR applies in relation with some specific sectors (for instance health), applications (for instance facial recognition or digital assistants), or rules and principles (for instance necessity and proportionality).

ii. **European authorities could provide guidance on "technology-neutral" rules**. While technologically neutral rules present the interest to apply to all technologies, their specific operation and application in relation with new digital technologies often needs to be determined with precision. If national and European authorities don't offer guidance, standardization organizations (sometimes dominated by industrial lobbies with very limited or no participation of NGOs and civil society) will take the driver's seat.

iii. **New rules could be necessary in fields where regulatory gaps exist**. This could be the case, for instance, in relation with civil liability regimes in order to avoid a situation where individuals who suffered damage because of AI systems are unable to request compensation (see *infra* Section 3).

iv. **Assuring adequate oversight**. Another important objective should be to ensure that AI applications presenting important risks for human rights are subject to adequate oversight and control.

If you think that new rules are necessary for AI system, do you agree that the introduction of new compulsory requirements should be limited to high-risk applications (where the possible harm caused by the AI system is particularly high)? (Highlight the chosen answer)

- YES
- NO
- **Other**
- No opinion

Other, please specify :

A risk-based approach is welcome. However, one could think of possible improvements to the Commission's approach.

First, it is necessary to better define high-risk and the methodology to assess it (see *infra* our response to the next question).

Second, one should take into consideration that "high risk" applications might not need new rules if they are adequately addressed by existing rules. We thus need a sector by sector approach.

Third, and inversely, “high risks” might exist in some low-risk sectors. For instance, digital assistants (that do not seem to fall under the Commission’s “high-risk” sectors) might raise very important risks for privacy.

Fourth, the Commission’s proposal seems to be based in an “all or nothing” dichotomy between “high-risk” applications (that would be subject to a new regulatory framework) and “non-high-risk” applications that would remain under the current regulatory framework. It might be better to adopt a more nuanced approach which could include, for instance, different rules for different degrees of risk. This could permit to have a more refined regulatory spectrum than a “black or white” approach.

Do you agree with the approach to determine “high-risk” AI applications proposed in Section 5.B of the White Paper?

(Highlight the chosen answer)

- YES
- NO
- **Other**
- No opinion

Other, please specify :

The Commission’s approach to determine “high-risk” might need more work and refinement. We have identified three potential problems:

The limits of the first of the two cumulative criteria: The idea of providing a list with sectors where “given the characteristics of the activities typically undertaken, significant risks can be expected to occur” is interesting and could provide legal certainty for business and lawmakers. However, as mentioned earlier, there might be high-risk applications in “low-risk” sectors.

The limits of the second of the two cumulative criteria: **The definition of this second criterion seems a little bit cyclical**: *“high-risk” applications are the ones that are used in such a manner that “significant risks” are likely to arise”*! It might be more interesting to focus on the potential *severity of the harm* that might occur together with the *likelihood* of its occurrence. Depending on the *severity/likelihood* ratio of specific applications different regulatory solutions might apply. This could lead to nuances (extreme/high/moderate/low risk) and appropriate regulatory responses, instead of what seems to be a rather binary (“all” or “nothing” – see *supra*) approach by the Commission.

The limits of the “exceptional instances” clause: While the Commission is right to acknowledge that “there may also be exceptional instances where, due to the risks at stake, the use of AI applications for certain purposes is to be considered as high-risk as such – that is, irrespective of the sector concerned”, this clause introduces a lot of uncertainties. Who will define which are these “exceptional instances”? Is it going to be the EU by new introductions to the applications listed under the first criterion? Or could this assessment become part of the margin of appreciation of developers themselves? As the 2020 Juri Committee Draft Report on a framework of ethical aspects of artificial intelligence, robotics and related technologies noted “the determination of whether artificial intelligence, robotics and related technologies are to be considered high-risk as regards compliance with ethical principles should always follow from an impartial, regulated and external assessment”.

If you wish, please indicate the AI application or use that is most concerning ("high-risk") from your perspective:

Lethal autonomous weapon systems are the most famous and notable example. The use of facial recognition technologies in order to undertake mass surveillance is also scary. The same applies to several AI tools that might be used in order to "manipulate" humans or "hack the human mind". The use of data and AI tools for disinformation and influence operations is also a major concern as it could lead to the destabilization of the very democratic foundations of our societies.

In your opinion, how important are the following mandatory requirements of a possible future regulatory framework for AI (as section 5.D of the White Paper)

(1-5: 1 is not important at all, 5 is very important)?

The quality of training data sets	**5**
The keeping of records and data	**4**
Information on the purpose and the nature of AI systems	**5**
Robustness and accuracy of AI systems	**5**
Human oversight	**5**
Clear liability and safety rules	**5**

A. ***The quality of training data sets (5)*** *:* The quality of AI outputs is based on the quality of the training data sets. Instances have already demonstrated the garbage-in garbage-out principle. Having "perfect" and unbiased training data might of course be challenging in some situations. However, training data sets should be as unbiased and representative as possible in order to avoid discriminatory outcomes in cases where the decision of an AI has a significant impact on an individual (i.e. legal decision, breach of human rights etc.).

   The quality of training data sets implies the existence *of* and access to quality tools in order to assess them. The EU should thus encourage a multi-stakeholder and multidisciplinary approach in order to provide such technology and methodology tools to companies, in particular SMEs.

B. ***The keeping of records and data (4)*** *:* The keeping of records in relation to the programming of an algorithm and the kind of datasets used to train AI systems is very important in cases when the use of these systems results in decision making and/or affects the life of persons and their rights. This permits to ensure the transparency and traceability of the process and to explain AI outcomes to the persons concerned or the relevant oversight authorities. The lack of records could enhance the fear that AI systems are complex and opaque (the popular perception of AI systems as "black-

boxes") and could make it difficult to identify and prove possible breaches of laws, including of legal provisions that protect fundamental rights, attribute liability and enable affected persons to claim compensation.

When it comes to the White Paper's suggestion to "keep the data themselves" this could be more problematic.

First, the conservation of personal data would often be incompatible with the limits posed by existing EU law on storage limitation (for instance Art. 5(1)(e) GDPR or consent-related time constraints) or data retention (including the *Tele2 Sverige and Watson* CJEU judgment).

Second, the indefinite conservation of data would only increase the risks of cyberattacks and other data breaches.

Third, the systematic keeping of mass volumes of data sets could have an important energetic and environmental impact.

C. ***Information on the purpose and the nature of AI systems (5)*** *:* It is important to ensure the transparency principle, in order to allow individuals to exercise an informed consent and their rights. People have to be warned and aware that they are interacting with an AI system and be in a position to object to a decision totally or partially adopted by an AI-system.

D. ***Robustness and accuracy of AI systems (5)*** *:* As the Commission repeatedly said, trustworthy AI must be reliable and robust enough to manage any mistake or incoherence that can happen from internal and external sources during the life of the AI system. Reliability and resilience of the system are essential. The determination of the level of accuracy of such systems mentioned in the White Paper has to be done with care in order for them to be functional, reliable, secure and trustworthy. Such approach is already implemented in companies in the domain of cybersecurity where the robustness principle allows for an anticipation of risks. Indeed, in the field of AI safety by design and cybersecurity are key concerns for robustness.

   For instance, autonomous car is a sector where resilience and robustness are particularly important. Security and safety issues inherent to autonomous cars provoked fatal accidents, and tests proved that the technology is vulnerable to cybersecurity and data protection issues. The robustness of the system is key as the flaws that may occur within the security systems could lead to very serious damages and harm to consumers and third-parties. They might also hamper the functionable capacities of the car themselves.

E. ***Human oversight (5)*** : Human oversight is of particular importance when AI can have implications in terms of safety, human rights or lead to discriminatory outcomes. The compliance with several of the requirements mentioned above can only be ensured validation and supervision of humans during the AI system's development and posterior human monitoring during its operation.

   As the White Paper rightly emphasizes, the appropriate degree of human oversight might vary from one case to another. The High-Level Expert Group on AI established by the Commission noted in its 2019 Ethics Guidelines for Trustworthy Artificial Intelligence that the idea of human intervention in every decision cycle of the system is neither possible nor desirable in several situations. A human-on-the-loop approach, referring to the capability for human intervention during the design cycle of the system and monitoring the system's operation, should be more appropriate and should be combined with a human-in-command approach, referring to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation.

The ultimate objective of human oversight is that we should never lose control of machines more powerful than we are. The Chair reads with interest in this respect Stuart Russell's approach (in Human Compatible, 2019) to the AI control problem. Russel argues that a major objective should be to prevent catastrophic misunderstandings of human preferences and encourage cooperation and communication with humans. He suggests to develop provably beneficial AI machines that focus on deference to humans. Unlike in the standard model of AI, where the objective is rigid and certain, this approach would have the AI's true objective remain uncertain, with the AI only approaching certainty about it as it gains more information about humans and the world. "Uncertainty about objectives implies that machines will necessarily defer to humans: they will ask permission, they will accept correction, and they will allow themselves to be switched off", concludes Russell.

F. ***Clear liability and safety rules (5)*** *:* Clear liability and safety rules are essential in order to provide legal certainty to business and to protect the users.

In addition to the existing EU legislation, in particular the data protection framework, including the General Data Protection Regulation and the Law Enforcement Directive, or, where relevant, the new possibly mandatory requirements foreseen above (see question above), do you think that the use of remote biometric identification systems (e.g. face recognition) and other technologies which may be used in public spaces need to be subject to further EU-level guidelines or regulation: **(Highlight the chosen answer)**

- No further guidelines or regulations are needed
- Biometric identification systems should be allowed in publicly accessible spaces only in certain cases or if certain conditions are fulfilled (please specify)
- Other special requirements in addition to those mentioned in the question above should be imposed (please specify)
- Use of Biometric identification systems in publicly accessible spaces, by way of exception to the current general prohibition, should not take place until a specific guideline or legislation at EU level is in place.
- Biometric identification systems should never be allowed in publicly accessible spaces
- No opinion

Please specify your answer:

The use of Facial Recognition Technologies (FRTs) creates important risks and problems that go well beyond the subject matter of the question (their "use in public spaces"). Biometric data are particularly sensitive data and thus FRTs are almost never a completely "harmless" type of processing. The risks for privacy, the cybersecurity risks, the issue of how consent should operate, the risks of errors, the risks of bias and discrimination (such as the fact that error rates of facial recognition algorithms can vary with gender or skin colour) and other concerns should always be taken into consideration when using these technologies. The **limited scope of the question** does not provide the opportunity to the Commission to receive feedback on some major issues such as whether the legal framework for the use of FRTs for identification purposes by law enforcement agents in European countries, within the context of criminal investigations, is adequate or not – and what could be the eventual improvements. **The AI**

Regulation Chair believes than a wider consultation on FRTs could be useful and would be delighted to participate to such a development.

Coming to the specific question asked by the Commission, we should note from the outset that there is today in Europe an important number of rules, starting with the GDPR and the Law Enforcement Directive, permitting to regulate the use of Facial Recognition Technologies (FRTs).

As noted by the White Paper, the GDPR **prohibits, in principle**, the processing of biometric data. However, some **exceptions** are announced permitting to use FRTs in some cases such as when there is an explicit consent of the persons concerned; to protect their vital interests; or on the basis of a substantial public interest.

The Law Enforcement Directive follows the same logic, only allowing such data to be processed in cases of "strict necessity" and only when specifically "authorised by Union or Member State law".

In both cases the exceptional use of FRTs is subject to a strict necessity and proportionality test and to appropriate safeguards for the rights and freedoms of the data subject.

It is fortunate that we have, in Europe, these and other rules applicable in relation with the use of FRTs in public spaces. Be that as it may be, a careful analysis is needed in order to assess whether the "technologically neutral" rules of European instruments and the *specific* provisions concerning biometric data are sufficient and permit to address all the risks related to this particularly sensitive sector.

Our research projects undertaken within the MIAI Chair on the Legal and Regulatory Implications of Artificial Intelligence shows that several improvements could be considered:

- First, there is a **risk of fragmentation in the interpretation and the implementation** of the GDPR and the Law Enforcement Directive in this field by EU Member States. In some States Data Protection Authorities are systematically consulted before the deployment of FRTs in public spaces – while in other States this does not seem to be the case. When Data Protection Authorities intervene, it is hard to assess it they proceed to a harmonized interpretation of the GDPR as their opinions are not always published or easily accessible, and are only available in the national languages of each member State. The European Data Protection Board could play a very useful role here in providing guidelines for a harmonized interpretation of existing rules. Guidelines on how strict necessity and proportionality should be interpreted as well as how exactly consent should operate in this field could be particularly useful.
- Second, we believe that there is a **lack of adequate information and transparency concerning the existing and intended uses of FRTs in public spaces** in the different Member States. This renders the assessment of intended uses of FRTs in public spaces cumbersome.
- Third, there is no doubt about the fact that evolutions in FRTs present an **unprecedented surveillance potential**, capable of undermining societal choices and raising several important ethical issues. A **public and democratic debate** should precede the use of FRTs and lead to important political decisions on these matters.
- Fourth, this democratic debate will permit to **fix from the outset some "red lines"** beyond which no use, not even "experimental", of FRTs should be allowed.
- Fifth, if and when the use of FRTs is considered as legitimate and proportional, a common understanding of what exactly should be the **"appropriate safeguards"** would be useful. The question of **oversight and control** by democratic bodies, regulators and judicial bodies is also crucial.
- Finally, a **specific attention should be given to the relevant databases**. Rules governing the entry of biometric data into a database as well as the exit of biometric

data could be considered, as well as the monitoring process of biometric identification systems. In addition, there is a need to ensure that comparison databases are relevant to the intended purpose.

Do you believe that a voluntary labelling system (Section 5.G of the White Paper) would be useful for AI systems that are not considered high-risk in addition to existing legislation? (Highlight the chosen answer)

- Very much
- **Much**
- Rather not
- Not at all
- No opinion

Voluntary labelling systems could be useful but the developments in the White Paper do not provide sufficient information about how exactly they should work and how some risks related to their use might be addressed.

The usefulness of labelling and certification systems has been proven in some other sectors. In the field of cybersecurity certifications granted by National Cybersecurity Agencies (ANSSI in France) are very useful in order to assess whether some products meet the necessary safety standards. The European Parliament's Draft Report on Civil liability regime for artificial intelligence also suggests that a labelling system in the AI field might be beneficial for companies and society as whole. The Voss Report suggests that in the case of a fault-based liability regime for no-high-risks AI systems, the deployer will have the obligation to establish he/she acted with due diligence for not being found liable of the damage/harm caused by the AI system (art.8). One of the ways to show that he/she acted with due diligence, is to having selected "a suitable AI-system for the right task and skills, putting the AI-system duly into operation, monitoring the activities and maintaining the operational reliability by regularly installing all available updates''. If this scenario materializes, labelling could thus facilitate SMEs' ability to prove their due diligence in the framework of liability claim.

Voluntary labelling systems rise several questions and difficulties: how are they going to work? Who will grant the quality labels? How to avoid them becoming *de facto* mandatory due to the fact that people would mainly use labeled applications putting aside non-labeled ones? How to define with precision the standards and constraints for the label in such a young and quickly evolving field as AI? This might take years to develop. Putting in place a labelling system would also bring new actors offering to help companies to make their AI get the labelling. This could add significant costs to SMEs and favor large groups. All these and other concerns should be addressed thoroughly.

What is the best way to ensure that AI is trustworthy, secure and in respect of European values and rules? (Highlight chosen answer(s))

- **Compliance of high-risk applications with the identified requirements should be self-assessed ex-ante (prior to putting the system on the market)**
- Compliance of high-risk applications should be assessed ex-ante by means of an

external conformity assessment procedure

- Ex-post market surveillance after the AI-enabled high-risk product or service has been put on the market and, where needed, enforcement by relevant competent authorities
- **A combination of ex-ante compliance and ex-post enforcement mechanisms**
- Other enforcement system
- No opinion

Do you have any further suggestion on the assessment of compliance?

Compliance of high-risk application with the legal and ethical requirements should first be self-assessed by the developers themselves. We could draw here a comparison with the Data Protection Impact Assessment (DPIA) existing under the GDPR under the “privacy by design” principle. According to Article 35 of the GDPR: “Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data”. Self-assessment should not be viewed as an administrative burden but as an important step to realise “safety by design”. It will also provide protection against liability claims, by showing, for instance, that the developer exercised due diligence.

More generally, high-risk applications should be tested and certified before they reach the single market and also be submitted to posterior conformity controls. We could make here an analogy with what happens with conformity assessments and posterior technical inspections of cars. However, these requirements and controls should be subject to several conditions and caveats. For instance, conformity assessment requirements already exist in EU law in relation with a large number of products and it is necessary to avoid duplication. Also, one should take into consideration the reservations expressed above in relation with identifying “high-risk” applications.

Section 3 – Safety and liability implications of AI, IoT and robotics

The overall objective of the safety and liability legal frameworks is to ensure that all products and services, including those integrating emerging digital technologies, operate safely, reliably and consistently and that damage having occurred is remedied efficiently.

The current product safety legislation already supports an extended concept of safety protecting against all kind of risks arising from the product according to its use. However, which particular risks stemming from the use of artificial intelligence do you think should be further spelled out to provide more legal certainty? (Highlight chosen answer(s))

- **Cyber risks**
- **Personal security risks**
- **Risks related to the loss of connectivity**
- **Mental health risks**

In your opinion, are there any further risks to be expanded on to provide more legal certainty?

There is a public awareness about the need to take into consideration cybersecurity and personal security risks. Nonetheless, other risks, such as the risks for mental health explained in the White Paper are issues which are not adequately addressed by the current regulatory framework or do not reflect a general legal consensus at the national, European or international level.

Do you think that the safety legislative framework should consider new risk assessment procedures for products subject to important changes during their lifetime? (Highlight the chosen answer)

- **YES**
- NO
- No opinion

Do you have any further considerations regarding risk assessment procedures?

As mentioned in the report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, new risk assessment procedures should be considered in the case of important products change during the lifetime cycle of the AI system. However, the concept of “important change” needs further clarification.

Do you think that the current EU legislative framework for liability (Product Liability Directive) should be amended to better cover the risks engendered by certain AI applications? (Highlight the chosen answer)

- **YES**
- NO
- No opinion

Do you have any further considerations regarding the question above?

The Product Liability Directive, has served since 1985 as an effective tool and contributed to enhancing consumer protection, innovation and product safety.

Some of its technologically neutral rules might nonetheless need adjustment and update. As emphasized in the 2019 Report Liability for Artificial Intelligence of the Expert Group on Liability and New Technologies set up by the Commission: "some key concepts underpinning the EU regime, as adopted in 1985, are today an inadequate match for the potential risks of emerging digital technologies".

The Product Liability Directive is based on the principle that the producer is liable for damage caused by the "defect" in a "product" – very much perceived as a material object. However, in complex systems using AI (such as autonomous cars) there might be a constant interaction between traditional products (cars) and associated *services* making a sharp separation between them unfeasible. It is thus necessary to carefully think whether and how the Product Liability Directive might dial with services where there is a significant risk of damage to persons or property. An issue that needs particular attention is whether software is covered by the legal concept of product or product component and how an updated Directive could deal with this issue without imposing an impossible burden to software developers.

The concept of "defectiveness" also needs careful thinking in relation with sophisticated autonomous systems based on machine-learning. If we go back to the software debate mentioned in the previous paragraph, it might be problematic to consider cyber vulnerabilities as a "defect" and impose a strict liability regime to software developers – especially taking into consideration that patches are regularly released to deal with such vulnerabilities. An obligation of due diligence in order to regularly examine vulnerabilities and provide patches over the lifecycle of a software used in such high-risk systems could be more appropriate to deal with this issue.

A better definition of the repartition of liability between the various stakeholders involved, especially between the producer, the operator or what the Voss Report calls the "deployer" (= the person who decides on the use of the AI-system, exercises control over the associated risk and benefits from its operation), should also be considered.

Do you think that the current national liability rules should be adapted for the operation of AI to better ensure proper compensation for damage and a fair allocation of liability?

(Highlight the chosen answer)

- Yes, for all AI applications
- **Yes, for specific AI applications**
- No
- No opinion

Please specify the AI applications:

As already discussed in relation with regulation in general (see Section 2 above) or in relation with the Product Liability Directive (see the answer to the question above), **some adjustments** to the existing safety and liability framework might be useful and even necessary in order to build trust on AI and to provide users of AI systems and applications with adequate protection.

National liability legislations offer an **important and technology neutral framework** in order to cover the challenges arising by the use of new technologies and to enable persons who suffered damage by the operation of such technologies to request reparation.

However, EU Member States laws of tort are **far from being harmonized**. Also, it is questionable whether all of them include clear liability rules specifically applicable to damage resulting from the use or misuse of AI systems. Indeed, the specific characteristics of AI systems, including their complexity, connectivity, opacity, autonomy and vulnerability to cyberattacks, **could make it more difficult for victims to present a claim of compensation or establish the causal link** between the victim's harm and the defendant's action. The EU has noted several times that it considers the liability risk to be one of the key factors that defines the success of new technologies, products and services and that a satisfactory regime could help the **public trust** AI technology.

Current national liability rules could thus be updated and harmonized in order to better address some issues:

Systems of **alleviation (or even reversal) of the burden of proof** could be useful in order to avoid a situation in which persons who suffer harm or whose property is damaged end up without compensation.

New regimes of **strict liability** should be introduced in situations where the operation or the (mis)use of AI-systems involves a significant potential to cause important harm to persons. This should be particularly the case when such systems are used in public spaces with, potentially, an important number of persons exposed to a specific high risk. As a matter of fact, one could expect that in "private" situations, such as in smart home appliances, the existence of a contractual relation with the operator should in principle enable victims to seek compensation. In any case, what is important is that persons who suffered damage because of AI systems should be able to seek compensation.

We should recall that the Axel Voss' draft report to the European Parliament on the creation of a civil liability regime for artificial intelligence provides for a **distinction** between a strict liability regime for "high-risk AI systems" (art.4); and a presumption of fault-based liability for "other AI-systems" (art.8).

An important issue to consider nevertheless is **whether the same definition of "high risk" should be adopted** in relation with the standard regulatory framework analysed above

(Section 2) and in relation with the specific question of liability rules. In any case the concept of “high-risk” needs to be clearly defined. The Voss proposal to **list** as an Annex to an eventual future instrument all AI-systems with a high risk potential (and to review and amend if necessary the Annex every six months) could provide legal certainty. The creation of a 'Technical Committee – high-risk AI-systems' (TCRAI) which would support the Commission in its review is an interesting proposal.

Axel Voss’ proposal to require from deployers of high-risk AI-systems to hold an **adequate liability insurance** is also important. How exactly such a proposal could be implemented should be studied thoroughly with the participation of the insurance industry.

We also note the opposition of the Voss Report to publicly funded compensation mechanisms, which seems justified if new liability regimes with mandatory insurance are introduced.

Yet, the important issue of **hacking** should also be addressed. Hacking is a serious threat to users of software-based technologies and traditional (existing) tort law rules may often prove insufficient because of the victim’s inability to identify the tortfeasor.

In such situations there can be two solutions.

First, the legal regime of strict liability for high risks applications could exclude the exoneration of responsibility of the deployer. This is what the Voss Report suggests in Article 8(3): “Where the harm or damage was caused by a third party that interfered with the AI-system by modifying its functioning, the deployer shall nonetheless be liable for the payment of compensation if such third party is untraceable or impecunious”.

If such a solution is not adopted, the only alternative solution could be, as proposed in the Expert Group’s 2019 Report Liability for Artificial Intelligence, to introduce a non-fault compensation scheme, equivalent to that applicable to victims of violent crimes, to ensure that persons who suffered damage from the hacking of high risk AI systems will be able to seek compensation.

Thank you for your contribution to this questionnaire.