FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

COMMENTS TO THE EUROPEAN COMMISSION'S WHITE PAPER ON ARTIFICIAL INTELLIGENCE – A EUROPEAN APPROACH TO EXCELLENCE AND TRUST

Prof. Dr. Franziska Boehm
With Diana Dimitrova, Stephanie von Maltzan, Fabian Rack, Francesca Pichierri

11 June 2020

Content

1. INTRODUCTION

The European Commission recently released the White Paper on Artificial Intelligence (AI) in which the European Commission, amongst others, makes proposals with regards to and seeks consultation on how to regulate AI technologies throughout their lifecycle and across the entire supply chain in full respect of the values and rights of the concerned individuals. The AI strategy, which is quintessential European, is designed to put people first and promote a trustworthy AI. This strategy tries to make an ethical approach to AI technologies in order to have a competitive advantage over China and the US.

The White Paper on AI defines AI as "a collection of technologies that combine data, algorithms and computing power."[1] Data and algorithmic combination clearly includes algorithmic decision-making systems such as profiling techniques and biometric applications.[2] Such tools could be used to support operations such as a Passenger Name Record (PNR) profiling operation or knowledge extraction from Big Data, which can have an effect on the rights of individuals.

The below paragraphs will highlight the issues in relation to AI which are particularly challenging from a legal, but also from more practical, political and technical perspectives. The legal analysis will focus on the following topics: the Commission's proposed high-risk approach to the regulation of AI, challenges related to data protection and effective remedies in the framework of profiling in the law enforcement field, such as PNR, biometric technologies, oversight and explainability of AI and liability. The topic of oversight and explainability will include also technical aspects related to the explanation of the reasoning of AI and the results it produces. When highlighting the issues, suggestions for regulatory measures will be made. The present consultation will also highlight certain legal and ethical problems related to mental manipulation, which have not been mentioned by the Commission but which need more attention. Our response to the consultation will then go on to comment on the

[1] European Commission, "White Paper on Artificial Intelligence – a European Approach to Excellence and Trust," COM (2020) 65 final, 19 February 2020, p.2. (Hereinafter "White Paper on AI").
[2] See examples of AI such as references to remote biometric identification and other examples of algorithmic applications/technologies throughout the White Paper on AI, e.g. p. 16 and 18.

FAIR principles which the White Paper mentions, and which legal, practical and political issues should be taken into account when applying these principles to AI and AI development projects. Then, from a more practical perspective, the analysis will make suggestions about what should be included during the development of AI, especially as concerns research projects on AI. Recommendations on the focus of training in relation to AI will also be made.

2. High-risk approach

The White Paper sets out a two-level test to define a high-risk AI application, which, unlike the low-risk applications, would be subject to forthcoming regulation, as the Commission suggest.[3] The forthcoming regulation will introduce a list of high-risk sectors as suggested by the White Paper, such as health care, transport, energy and parts of the public sectors.

The question arises whether the separation between high and low risk applications and the criteria for risk-applications are adequate. Further legal questions are raised by the requirements that will be imposed on high- risk applications and on the proposed testing centres.

With regards to the first question, the Commission proposes two cumulative criteria for high-risk applications, namely that high-risk applications are those (1) used in a sector where "significant risks can be expected" and (2) additionally "used in such a manner that significant risks are likely to arise."[4] Besides this two-level test there might be exceptional instances where an AI application will be seen as a high-risk application – regardless of the sector.[5] Such high-risk technologies would be, for example, biometric technologies for remote identification and applications for recruitment processes.[6]

[3] White Paper on AI, p.17.
[4] White Paper on AI, p.17.
[5] Ibid.
[6] White Paper on AI, p. 18.

This approach is problematic for three reasons:

- First, it assumes that it is not only possible to calculate risks but also differentiate AI applications in only two levels of high or low-risks.
- Second, this high-risk approach leaves open questions about the scope of the notion "high-risk" in relation to AI applications and thus opens lots of opportunities for interpretation.
- Third, there is also a risk that following this approach many AI applications would fall outside the scope of high-risk applications.

For example, the high-risk approach might not apply to advertising technology or scoring systems in general. However, existing scoring systems can create substantial risks by the use of statistical analysis. The problem is that these also mostly do not fall under the scope of Article 22 GDPR. This is because Article 22 GDPR addresses solely automated processes in which an algorithmic decision implies a direct action and which has a legal effect concerning the data subject or similarly significantly affects him or her. Thus, Article 22 GDPR only covers a small amount of algorithmic decision processes. Some scoring activities, for example, might also not be covered by Article 22 GDPR, because the automated evaluation only prepares a decision, which is then later taken by a human. Scoring and statistical analysis, however, could have big impact on individuals and might thus remain unregulated, either by Article 22 GDPR or the forthcoming regulation on high-risk applications. Sometimes other provisions of the GDPR might also not apply, e.g. in the framework of statistical analysis when it is performed with anonymous data only.

The difficulty with applications not passing the threshold of high-risk applications is that low-risk applications will only be part of a voluntary labelling scheme, which could also create the illusion of responsible behaviour. However, these applications will need more than voluntary standards to address the inherent risks. A more defined and scaled risk approach should be considered.

The best way forward should be **to apply the requirements to all risky automated decision-making systems** and not exclude so called low-risk applications, even if they do not use machine learning techniques. It is thus recommended to develop a more-level based risk assessment which can nuance different automated decision-making systems more effectively. For example, the

proposal of Germany's Datenethikkommission[7] divided automated systems into five categories of risk. On that background, to range the different AI applications and their risk regarding the European values and rules, the differentiation in a two-level risk approach is barely enough. Admittedly, not each AI application requires an in-depth inspection. The two-level approach, however, assumes that people are already protected from low-risk applications, because developers have to comply with the GDPR. It ignores the fact that there are applications and uses of AI systems that do not fall under the scope of Article 22 GDPR as mentioned above. On the other side, a large number of moderately risky applications, such as Smart Mobility applications, would fall under onerous and disproportionate requirements. This could stifle innovation, especially with regard to SMEs. A more-level based risk approach should therefore be considered without slowing down the innovation.

It is not clear why the European Commission decided against a more-level risk-based approach. The different scale of applications and the spectrum of potential automated decision-making systems along with their benefits and risks should therefore be captured and dynamically balanced. A classification regarding the risk-level approach should be determined according to the relevance and scope of the AI applications and the impact on the data subjects. Thus, it is recommended that the Commission reconsiders it proposal on regulating only high-risk applications and adopts a more nuanced approach.

3. Gaps in the Regulation of AI in the Law Enforcement Field

The Commission White Paper acknowledges that citizens' rights might be "most directly affected" by AI technologies in the law enforcement field and in the judiciary.[8] However, in the White Paper relatively little attention is paid to AI applications in the law enforcement field. Example of what regulatory

[7]Gutachten der Datenethik Kommission, available at: https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.html (accessed 11 June 2020).
[8] White Paper on AI, p. 10.

measures might be lacking, e.g. in terms of safeguards for the affected individuals, is the EU PNR Directive.[9] It is presumed that AI in the law enforcement field would most likely fall within the scope of high-risk AI because of the sensitivity of the sector and the high risks for individuals. Thus, it is presumed that the forthcoming regulation would apply to such AI applications. The following paragraphs will demonstrate where there are gaps in the regulation of AI in the law enforcement field, taking PNR as an example.

Pursuant to the EU PNR Directive, air carriers are obliged to transfer passenger reservation data 24 to 48 hours before the flight to specially established passenger information units (PIUs) in the EU Member State of departure/arrival.[10] The PIUs, after receiving the data, shall perform passenger risk-assessment by comparing the submitted data against pre-established abstract criteria and background databases to discover patterns through this automated analysis of the PNR data.[11] The purpose is to identify individuals who might pose a serious security threat, but who are not known to the law-enforcement authorities yet.[12] However, no final decision as to whether to subject a passenger to a further checks may be made by purely automated means, i.e. each hit should be individually reviewed.[13] Nevertheless, it is still the profiling algorithm which automatically assesses the risks an individual could pose and "selects" the individuals which it considers to be risky.

When the PIUs and the Member State competent authorities process personal data in the framework of PNR, then Directive 2016/680 on data protection in the law enforcement field is applicable, in parallel to the data protection provisions in the PNR Directive itself.[14]

---

[9] Directive (EU) 2016/681 of the European Parliament and of the Council of 27 April 2016 on the use of passenger name record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime, OJ L 119, 4.5.2016, p. 132–149 (Hereinafter "PNR Directive").
[10] Article 8 PNR Directive. For the complete list, see Annex I, PNR Directive.
[11] Article 6 (3) PNR Directive and Recital 7 PNR Directive.
[12] Recital 7 PNR Directive.
[13] Article 6 (5) and Article 7 (6) PNR Directive.
[14] Recital 27 and Articles 6, 7 and 13 PNR Directive; Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the pre-

As a background note, Article 11 Directive 2016/680, similarly to Article 22 GDPR,[15] prohibits taking decisions "based solely on automated processing, including profiling, which produces an adverse legal effect concerning the data subject or significantly affects him or her."[16] However, it allows automated decisions to be taken and profiling to be performed if: (1) the respective measures are based in Union or Member State law to which the controller, e.g. a police and investigatory authority, is subject, and (2) if this law "provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller."[17] However, safeguards such as expressing one's point of view and contesting the decision are missing from Article 11 Directive 2016/680.[18] There are further essential safeguards which are missing, as will be demonstrated below. Some of these safeguards concern issues which have been pointed out by the Commission in the White Paper as important requirements in the regulation of AI technologies, e.g. transparency, traceability and human oversight. Others, e.g. **sunset clauses and judicial review were not explicitly listed in the White Paper and should be included in future**.

a. Transparency and Directive 2016/680

The importance of **transparency** via the provision of **information** is highlighted by the Commission in the White Paper.[19] It is noted that an obligation to inform individuals that they are potentially subject to such measures, e.g.

vention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, O.J. L 119/89-131 (Hereinafter "Directive 2016/680").

[15] Article 22 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and re-pealing Directive 95/46/EC (General Data Protection Regulation), OJ L119/1, (Hereinafter "GDPR").

[16] Art. 11 (1) Directive 2016/680.

[17] Art. 11 (1) Directive 2016/680.

[18] Such safeguards can be found in Article 22 (3) GDPR, although it is acknowledged that these two additional safeguards do not apply to automated decisions which are taken on the basis of a measure based in law. Article 11 Directive 2016/680 also prohibits taking decisions solely based on sensitive data and profiling measures which result in discrimination and which are based on special categories of data.

[19] White Paper on AI, p. 20.

automated decision-making, is missing from Directive 2016/680.[20] Thus, there is no legal guarantee that data subjects would be made aware of the fact that they are subject to such a measure, e.g. in the framework of PNR. This lack of awareness might make it more difficult for the data subject to assert their additional rights, e.g. to rectification and erasure. Thus, the information obligations in Directive 2016/680 with regards to automated decision-making should be strengthened or the rights should be alternatively boosted in the new regulation on AI. This will ensure that the controller provides such information across similar AI applications in the law enforcement sector.

Such a harmonisation is all the more pertinent because information obligations are also not systematically enshrined in sectoral legislation, e.g. in the EU PNR Directive.[21] Therefore, it is recommended that both the information obligations in Directive 2016/680 and in sectoral legislation such as PNR are strengthened via **legislative amendments.**

b. INDEPENDENT CONTROL

Furthermore, one needs to establish mechanism for **independent control** over the **abstract profiling criteria (abstract scenarios) and the quality of training data sets.**

It is recommended that **regular reviews by an independent supervisory authority with powers to request amendment of the criteria and the training data, are carried out**. This might be in the framework of prior consultations of the data protection supervisory authority, but such reviews should be not restricted to prior consultations only.[22] Independent review, e.g. by data protection supervisory authorities, should not be prejudiced by the fact that an AI

[20] Such an obligation was not found in Article 11 on prohibition of ADM, Art. 13 on the right to information or Art. 14 on the right of access. By contrast, such an obligation exists in Articles 13 (2) (f) and 14 (2) (g) GDPR on the right to information.
[21] The latter only provides in Recitals 29 and 37 that the data subject should be informed about the collection of PNR data, its transfer to the Passenger Information Units (PIU) in the respective Member States and their rights as data subjects as enshrined in Directive 2016/680. From these Recitals it is not explicitly clear whether the controller is obliged to inform the data subject about the existence of profiling and also because such an obligation is not explicit in Directive 2016/680.
[22] Article 28 Directive 2016/680.

application, e.g. PNR, is prescribed in EU law. Such control is necessary because, as the CJEU held in its Opinion on the EU-Canada PNR Agreement, "the extent of the interference" of the automated processing of PNR data with Articles 7 and 8 CFREU "essentially depends on the pre-established models and criteria and on the databases on which that type of data processing is based."[23] For this reason, the principles and explicit rules concerning the scenarios, the pre-determined assessment criteria and the background databases to be consulted should be set out in the applicable legal provisions.[24] This is to ensure that the PNR profiling identifies only targets against whom there is a "reasonable suspicion" of being involved in serious crime, i.e. ensure accurate results of the profiling.[25]

Thus, even if the abstract criteria of algorithmic applications – whether in the framework of other law-enforcement applications and programmes, or in other fields such as the health and commercial ones - cannot be disclosed (in their entirety) to the public for various reasons, then at least an independent supervisory authority should be able to regularly monitor the adequacy of the scenarios/criteria when such applications could infringe the users' fundamental rights. In addition, these should be able to examine also the legality of both the abstract criteria and the individual results, i.e. the application of the abstract criteria to individuals, because, as the CJEU noted, the abstract criteria could interfere with the rights to privacy and data protection. *A fortiori,* one could conclude that their application to individuals should, too, be seen as an interference. It has been pointed out that often trade secrets, e.g. related to the software which is designed to perform an AI task, prevent the disclosure of the algorithms or the logic of individual decisions, which in turn prevents full inspections from taking place.[26] This conflict of laws should be regulated in law. For example, an explicit obligation could be added that in such situations supervisory authorities, e.g. data protection authorities, should, subject

[23] CJEU, Opinion 1/15 of the Court (Hereinafter "Canada PNR Opinion"), 26.07.2017, par. 172.
[24] Opinion of Advocate General Mengozzi, Opinion 1/15, (2016), ECLI:EU:C:2016:656, par. 255. See also Diana Dimitrova, "The Right to Explanation under the Right of Access to Personal Data: Legal Foundations in and beyond the GDPR," (Forthcoming 2020).
[25] CJEU, Opinion 1/15 of the Court (Hereinafter "Canada PNR Opinion"), 26.07.2017, par. 172. See concerns on algorithmic decision-making quality also in Dennis Broeders et al, "Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data," Computer Law & Security Review, 33 (2017), pp. 317.
[26] Dennis Broeders et al, (n. 25 above), p. 317-318.

to their professional secrecy obligations, be given full access to all parts of the AI, including the trade-secret protected software. At the very least, as some have recommended "research results, profiles and correlations must be open to oversight: the data-processing party must be able to show clearly how they arrived at particular results."[27] Further remarks on oversight and explainability, especially from a more technical perspective, are made in Sections 5 and 6 below.

c. Record-keeping obligations

It is further noted that existing **record-keeping** obligations in existing legislation, e.g. in Directive 2016/680 on data protection in the law enforcement field and in sectoral legislation, e.g. the PNR Directive, are not always adjusted to the traceability requirements in the context of AI. For example, Article 24 (1) (e) Directive 2016/680 requires the keeping of records about the use of profiling. However, profiling is not the only possible process, which might rely on AI technologies. In addition, it is not clear whether pursuant to this obligation the controller should simply note that a profiling measure has been performed or also record the criteria/scenarios, etc, as well as how a decision or profile were made and applied to an individual. For the sake of legal certainty, such recording obligations should be clearly added. This is necessary, because sectoral legislation does not always anchor such obligations and fill the gap in Directive 2016/680. An example of the deficiencies in sectoral legislation is the PNR Directive, in which there are no explicit provisions about recording the justifications behind the decision to treat a certain passenger as potentially posing a risk, e.g. for national and public security. The PNR Directive provides a *non-exhaustive* list of information to be recorded, which does not mention explicitly the justifications for why a certain conclusion was reached in a certain case. It enshrines explicit documentation obligation only with regards to the collection, consultation, disclosure and erasure of the data.[28] Thus, again, due to the fact that neither Directive 2016/680 nor each piece of sectoral legislation enshrine detailed tracing obligations, explicit traceability obligations in law with regards to abstract criteria and their application to individuals are

[27] Dennis Broeders et al, (n. 25 above), p. 318.
[28] Article 13 (6) PNR Directive.

needed. Such a traceability requirement would enable the **accountability** of controller (Art. 4 (4) Directive 2016/680).[29]

d. PROFILING ACCURACY

Challenging the accuracy of AI applications, especially the accuracy of the produced results, could pose a problem to the concerned individuals. This is because the right to rectification in EU data protection law[30] has not been adapted to AI scenarios and under the Directive 2016/680.[31] In addition, as demonstrated above, in Directive 2016/680 data subjects do not have the opportunity to contest the automated decision/profiling result and express their point of view. This is a gap in the system of safeguards which needs to be corrected in Directive 2016/680 and the rights of the data subjects should be specified in relation to each application, e.g. to be extended in scope so that they can be effectively exercised. To enable individuals to challenge the accuracy of AI and thus the legality of the produced results individuals and their lawyers should have the opportunity to understand these decisions. Thus, the data protection principle that the controller, e.g. a police or judicial authority, remains responsible for the accuracy of AI processes should be enshrined in legislation. This responsibility means that the controller should be able to demonstrate „what a decision is based on and what factors and considerations were taken into account."[32] Otherwise there is a risk that the burden of proof will not shift to the data subjects.[33] Sections 5 and 6 will discuss the technical issues related to understanding such AI decisions. The opportunity to contest automated decisions and profiling results would also respect the right to effective remedies, a fundamental right in the EU.[34] The above suggested traceability requirements could help achieve this purpose. In addition, trade secrets should not prevent individuals from exercising their rights, i.e. they should not

[29] See on the importance of accountability Dennis Broeders et al, (n. 25 above), p. 320.
[30] Article 16 GDPR and Article 16 Directive 2016/680.
[31] Diana Dimitrova, "Personal Data Quality: Scope of the Legal Principle and Its Impact on the Right to Rectification" (Forthcoming 2020).
[32] Dennis Broeders et al, (n.25 above), p. 319.
[33] Dennis Broeders et al, (n.25 above), p. 319.
[34] Article 47 Charter of Fundamental Rights of the European Union (2012) OJ C326/391 (CFREU).

be absolute, and controllers should be able to disclose adequate information about how they reach results without disclosing commercial secrets.

e. JUDICIAL REVIEW

As some scientists have pointed out, AI technologies such as Big Data applications are difficult to challenge in courts and thus subject such applications to **judicial review**. They attribute this largely to the fact that it is difficult to prove individual harm. Thus, they recommend, e.g. providing for the opportunity for launching collective proceedings in relation to such technologies.[35] For example, in the PNR context it might be difficult for an individual to challenge the abstract scenarios and the algorithm as such. Under data protection law one could potentially challenge the application of these results to oneself, e.g. where they have been mistakenly flagged as a potential terrorist and was subject to additional checks. However, it looks unlikely that a challenge of the abstract criteria/scenarios and the logic of the algorithm could be challenged by an individual who might not have suffered individual harm. Looking at the CJEU EU-Canada PNR Opinion, one could argue that hypothetically the logic and criteria used for profiling could be challenged because they could interfere with the fundamental rights to privacy and data protection in Articles 7 and 8 CFREU. However, whether individuals have access to information on the logic and criteria in order to argue that these violate Articles 7 and 8 CFREU seems uncertain, as well as of course their standing in court.

f. SUNSET CLAUSES

Finally, AI applications are novel and challenging and their success outside the testing environment is not guaranteed. Therefore, such applications should have the opportunity to be revised and potentially dismantled, i.e. they should contain **sunset clauses**.[36] For example, the PNR Directive provides that once, by 25 May 2020, the Commission should prepare a review of the elements of the PNR Directive, paying special attention, amongst others, to "the necessity

[35] Dennis Broeders et al, (n.25 above), p. 320.
[36] Dennis Broeders et al, (n.25 above), p. 318.

and proportionality of collecting and processing PNR data for each of the purposes set out in this Directive."[37] However, this does not seem to be equal to re-considering the whole PNR programme as such or at least individual parts of it, based on the overall functioning and effectiveness of the programme or its individual components, as well as the necessity and proportionality of having such an application. Thus, an explicit sunset clause, at least with reference to elements of the PNR profiling technology, should be added. Such sunset should be considered also for other AI applications which could have an impact on people's fundamental rights and whose effectiveness in practice needs to be proven.

4. BIOMETRICS

Biometric applications are considered to be particularly sensitive due to the unique identification of individuals they enable.[38] The issues discussed above, especially on transparency, on sunset clauses, algorithmic matching accuracy, administrative and judicial control, are equally valid for biometric applications.

However, because of the sensitivity of the processed data, especially when DNA is included, and the unique identification they make possible, additional suggestions for improvement are proposed below.

First, some have noted that in the GDPR biometric data are classified as **sensitive data** in Article 9 (1) GDPR only if they are processed to uniquely identify an individual. Thus, it remains unclear whether the processing of biometric data for other purposes or in the framework of applications which might currently not allow unique identification but which might allow such identification in the future, are included. Thus, the Commission should pay more attention to this problem and strengthen the protection in relation to biometric data to close the existing gaps.[39]

[37] Article 19 (2) (b) PNR Directive.
[38] Article 29 Working Party, "Opinion 3/2012 on developments in biometric technologies," 27th April 2012, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2012/wp193_en.pdf, p.9 (accessed 11 June 2020).
[39] E.J. Kindt, "Having Yes, Using No? About the new legal regime for biometric data," CLSR, 2017, available at: https://daneshyari.com/article/preview/6890475.pdf, p.5.

Then, it is noted here that biometric applications *per se* cannot always produce 100% **accurate** matches.[40] Thus, for each biometric application adequate accuracy standards, adapted to the purpose of the processing, e.g. whether 1:1 verification or 1:n identification, should be developed and complied with. Reusage of the biometric data for new purposes and new applications should be strictly assessed. This is not only necessary in order to comply with the purpose limitation principle in EU law.[41] It would also prevent that biometric quality standards adapted for one purpose result in false acceptances and rejections when used for another purpose/application.

Similarly to the discussion above on judicial review, it is noted that **individuals should have the opportunity to challenge a biometric match or mismatch**. The consequences of the challenge should be regulated in law. For example, the right to rectification should be expanded to allow the individual to have the right to a new enrolment where it has been proven that the biometric data as enrolled by the controller are not of sufficient accuracy,[42] especially when the biometric application is used for identification purposes. For the concerned individual to be able to make his case that a certain (mis)match is not correct, **technical advice** should be made available to him/her. In addition, one should note that because biometric technologies are not perfect, the results of biometric matching should not be assumed to be automatically correct. The risk is that if too much trust is put in a technology which produces errors, negative consequences for the people on whom the technology is used will follow.[43] As has been stated, "when, for example, a non-match is routinely assumed to be a false identity claim by an imposter, this may lead to automatically putting the burden-of-proof on this person and, hence, to a violation of the presumption of innocence".[44]

Most importantly, before the deployment of each application the **necessity and proportionality** should be assessed. These are two of the principles which

[40] Ibid, p. 6.
[41] Article 5 (1) (b) GDPR and Article 4 (1) (b) Directive 2016/680.
[42] Council of Europe, "Progress report on the application of the principles of Convention 108 to the collection and processing of biometric data (2005)," T-PD, February 2005, par. 93.
[43] Sanneke Kloppenburg & Irma van der Ploeg (2020) Securing Identities: Biometric Technologies and the Enactment of Human Bodily Differences, Science as Culture, 29:1, 57-76, p. 73.
[44] Sanneke Kloppenburg & Irma van der Ploeg (2020) Securing Identities: Biometric Technologies and the Enactment of Human Bodily Differences, Science as Culture, 29:1, 57-76, p. 73.

need to be satisfied when the fundamental rights to privacy and data protection are interfered with.[45] There have been applications whose necessity has not been convincingly proven, e.g. taking students' fingerprints when access school facilities such as canteens.[46] It is not clear why access control could not have been designed differently, i.e. by processing less intrusive and less sensitive data such as alphanumeric data on school cards. Thus, biometric applications should not be universally allowed. This is especially the case where the **processing of data of children** is concerned, e.g. cameras is school rooms. This poses not only data protection risks, but also prevents children from being sensitivised to surveillance technologies.

A final point that deserves attention is that, usually, public authorities rely on private companies to acquire the technology and deploy it. As a consequence, as pointed out by a recent report by the European Union Agency for Fundamental Rights, it is also necessary that considerations about fundamental rights are directly "**built into technical specifications and contracts to ensure that the industry pays due attention to them**".[47] It has been suggested that technical specifications could, for example, refer to high quality standards in order to reduce wrong identification rates.[48]

5. Types of requirements for high-risk AI

According to the White Paper, if AI has been identified as posing a risk under the premises of the two-level test, the producer should examine the safety before deployment. The proposed requirements are: 1) that AI systems should be trained using data which "respects European values and rules" and a record

[45] Article 52 (1) CFREU; Case C-293/12 Digital Rights Ireland and C-594/12 Seitlinger and Others [2014] ECLI:EU:C:2014:238, par. 38.
[46] Stephen Mayhew, "UK school to use fingerprint for payment in cafeteria" Biometric Update.com (2012), https://www.biometricupdate.com/201210/uk-school-to-use-fingerprint-for-payment-in-cafeteria (accessed 11 June 2020).
[47] European Union Agency for Fundamental Rights, Facial recognition technology: fundamental rights considerations in the context of law enforcement: FRA focus, November 2019, p. 34, available at: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf (accessed 11 June 2020). Emphasis added.
[48] Ibid.

of such data should be kept; 2) that an AI system should provide "clear information to users about its purpose, its capabilities but also its limits" and it should be clear to users when they are interacting with an AI instead of a human; 3) AI systems must be "technically robust and accurate in order to be trustworthy"; and 4) they have always to ensure "an appropriate level of human involvement and oversight".

This raises several questions: How can these requirements be measured and enforced in practice? Are SMEs able to afford the cost of compliance and the cost of retraining algorithms? Furthermore, where the requirements sit in the broader regulatory context – as part of the regulation? How would they interact with other AI frameworks, e.g. the EU's High Level Expert Group's (HLEG) Ethic Guidelines[49]? In the debate regarding the forthcoming regulation these questions should be addressed. The focus within this analysis will nevertheless lie on the main issues such as "providing information" and "human oversight".

The White Paper proposes the need to provide clear information[50] about the capabilities as well as the limitations of an AI application and the need for human oversight.[51] The paper recognises that "the appropriate type and degree of human oversight may vary from one case to another,"[52] without proposing any differentiation.

We fully agree with the principle of transparency and explainability[53] as well as human oversight and acknowledge its necessity, as seen above in the section on AI in the law enforcement field. **At the same time, we note that AI applications are implemented in an increasingly complex way and that there are technical challenges to ensuring transparency and effective oversight.** Human oversight clearly depends on the ability to review and validate AI applications. This implies a good understanding of the internal processes. The complexity of these systems and their internal structure make their inner mechanism difficult to understand. Numerous hidden layers between input

[49] European Commission, High-Level Expert Group on Artificial Intelligence, available at: https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence (accessed 11 June 2020).
[50] White Paper on AI, p. 20.
[51] White Paper on AI, p. 21.
[52] Ibid.
[53] See for instance Amina Adadi and Mohammed Berrada, Peeking inside the Black-Box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access,* vol. 6, (2018) pp. 52138-52160.

and output layers and millions of parameters turn in particular neural networks into opaque models. Conveying this logic to technical laymen in a way that allows them to reason about the behaviour of algorithmic decision-making is extremely challenging. Explaining the functionality of complex algorithmic decision-making systems and their rationale is furthermore a technically challenging problem.[54] Therefore, designing processes in a way that ensures more explainable decisions is of social interest. Explainability is nevertheless a complex topic, in which several open questions exist.

This raises the question how these inner processes can be represented in such a way that people can understand and interpret the inner structure and processes. From deductive and rule-based systems (decision trees) to statistical probabilistic models (Bayesian networks) to artificial neural networks (multilayer perceptrons), a variety of different model types can be used. The ability to explain is strongly influenced by the model being used.

The technical representation of the model also raises the question which information about the inner process should be disclosed and to whom. Explainability or human oversight does not mean full access to all the sensible information. Approaches that require solely the disclosure of the training data as well as the algorithm are not effective due to the limited technical understanding of the average user. Many models process high-dimensional data which cannot be understood by technical laymen due to the lack of an explicit declarative knowledge representation. Depending on the receiver (user or testing centre) a different disclosure of information should be considered. Lipton[55] and Kahnemann[56] propose concrete implications for good explanations regarding the average user. For example, explanations of the decision-making process are contrastive and easier to understand for a technical layman in

[54] Joshua Kroll et al., Accountable algorithms, University of Pennsylvania Law Review, (2017) 633.

[55] Zachary C. Lipton, The mythos of model interpretability (2016), arXiv:1606.03490.

[56] Daniel Kahneman and Amos Tversky, The simulation heuristic (1981) available at: https://apps.dtic.mil/dtic/tr/fulltext/u2/a099504.pdf (accessed June 11, 2012); as well as Andrej Dobnikar, Uroš Lotrič and Branko Šter, Adaptive and natural computing algorithms: 10th international conference, ICANNGA 2011, Ljubljana, Slovenia (Lecture notes in computer science, vol. 6594, Springer 2011).

counterfactual,[57] short cases.[58] Thus, the users do not necessarily have to understand the internal structures. The most important aspect is the knowledge that an AI application decides (by the mandatory use of labels) and under which conditions the decision was made.

To sum up, the current business model is mainly based on the exploitation of people's data. A truly forward-thinking technology policy must ensure that those who are most vulnerable are protected. And at the same time paving the way to real alternatives without compromising individual rights. Increasing technical research in explainable AI promises deeper insight into the workings of AI solutions and increased acceptance both by lawmakers and by those potentially affected by automated decisions. Furthermore, an explanation of the rationale behind could be helpful to identify potential grounds for validations, such as inaccuracies in the input data, problematic inferences, or other flaws in the algorithmic reasoning.[59] This might increase the probability of successfully testing AI applications and their risks.

6. Testing centres for high-risk AI

The White Paper suggests a process of objective prior conformity assessment to verify and ensure high-risk application will meet the mandatory requirements. Procedures like testing, inspecting or certifying are specifically mentioned.[60] These *ex ante* reviews focus not only on the output but also the algorithms or training datasets and could be repeated in case of learning algorithms. The application could also be required to be retrained.

[57] An example of contrastive and counterfactual explanation is if a loan was denied because the annual income of the applicant for a loan was only 40.000€. In the explanation it is stated that if the income had been 45.000€, the loan would have been offered.

[58] See Sandra Wachter et al., Explanations Without Opening the Black Box: Automated Decisions and the GDPR, Harvard Journal of Law & Technology (2018) 31 (2), who let the AI explain what would have been necessary for another decision. In contrast to an attempt to convey the internal logic, counterfactuals describe a dependency on the external facts that leads to a decision.

[59] Brent D. Mittelstadt et al., The ethics of algorithms: Mapping the debate, Big Data & Society (2016).

[60] White Paper on AI, p. 23.

This raises the question which information is needed to test and certify the applications. Because of trade secrets and IT-Security requirements in relation to accuracy and robustness, the full disclosure of algorithms might not always be possible. Algorithm auditing is nevertheless a very limited way of inspecting the inner structure and decision-making in general. The focus is therefore shifted to an input-output analysis since most models are too complex and dynamic for a structural analysis. This input-output analysis is primarily for detecting biases and does not intend to break down any opacity or lack of transparency. The algorithm reveals the machine learning method and not the data-driven decision rule. In addition, companies have an interest in not sharing details of their algorithms to avoid disclosing trade secrets and violating the rights and freedoms of others. Such disclosure may, for example, be accompanied by a manipulation of the system. Revealing information such as the training datasets as well as the algorithms might be essential for testing the accuracy and robustness of AI applications by the use of audits and explainable AI. In order to protect the company's interests regarding intellectual property and data protection as well as to prevent manipulations the information should solely revealed to the mentioned testing centres in for example in-camera reviews and subject to professional secrecy obligations. The testing centres should be non-profit and preferably assigned with research as well. Since also the EU data protection supervisory authorities might be competent to sometimes examine algorithms when they fall within the competence of the data protection authorities, it should be clarified how the testing centres should cooperate with the said supervisory authorities.

The White Paper proposes the combination of *ex ante* and *ex post* enforcement mechanisms.[61] Unquestionably it is the only appropriate solution to deal with different AI applications throughout their lifecycle. **It is nevertheless important to create a regulatory body with a cadre of experts who are able to test and judge the design, development and deployment of AI applications as legally and ethical applicable.** Such an independent control was also recommended by the CJEU, as examined in the PNR case above. This body or testing centre should be able to provide recommendations and enforce fines and moratoria if an AI application does not comply. This should be combined

[61] White Paper on AI, p. 24.

– depending on the risk classification of the applications – with several other requirements, such as algorithm auditing and standardisation as well as labels.

7. LIABILITY

The discussion in the White Paper on the outstanding liability issues and the range of proposed solutions are controversial. To achieve a uniform solution, the White Paper proposes a liability regime according to the actor best placed to address the risk of harm. This consideration is interesting and thoughtful, but **in case of unforeseeable damages, how do they determine the cause of harm and the responsible party**? The applicable law does not always lead to appropriate solutions in view of the particular constellations of causation and responsibility.[62] From a legal point of view, it is of crucial importance whether the occurrence of the damaging event could have been foreseeable by complying with the duties of care. In AI, a uniform description of action, causality, and consequences is hardly achievable due to the increasing complexity, non-linear dependencies, and interdependencies. Further with regards to AI technologies, the knowledge learned on the basis of classification models and its effects on decision-making in a concrete context are generally unpredictable and consequently uncontrollable. Thus, in the framework of AI, in retrospect, it will be difficult to determine whether the damage-causing misconduct can be traced back to the original programming, later training or other environmental factors and who ultimately set the relevant cause. The "Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics" should aim to clarify this question.

[62] See Peter Asaro, The liability problem for autonomous artificial agents, AAAI Spring Symposia (2016).

8. FAIR PRINCIPLES FOR DATA: LICENSING AND INFRASTRUCTURE QUESTIONS

It is to be welcomed that the White Paper mentions the FAIR principles in dealing with the data that arise in the AI environment: "Promoting responsible data management practices and compliance of data with the FAIR principles will contribute to build trust and ensure re-usability of data."[63] The FAIR principles set certain requirements for Findability, Accessibility, Interoperability and Reusability of research data.[64] The principles should also be applied in a broader sense beyond science context, as the White Paper suggests.

This contribution takes the White Paper's mention of the FAIR principles as an opportunity to address licensing issues and related infrastructure issues, such as the certification of data repositories.

a. SHARING OBLIGATIONS IN THE PRIVATE AND PUBLIC SECTOR

In regard of the FAIR criterion Accessibility, a European Approach should address the issue of data access obligation for the private sector. This issue is primarily discussed at the levels of data protection law, antitrust law and competition law. Our comments do not intend to take a position on the complex question of data sharing obligations, but they do **suggest that the discourse be conducted also within the framework of the European AI approach.** There can be good reasons for such obligations, in particular when a market failure in a specific sector is detected or can be foreseen.[65] For the public sector, an open data framework was reformed with the directive on open data and the re-use of public sector information[66] in 2019.

At the same time, the FAIR principles do not interpret the "Accessible" criterion in such a way that it is mandatory for Open Access. Rather, data can be

[63] White Paper on AI, p.8.
[64] https://www.go-fair.org/fair-principles/ (accessed 11 June 2020).
[65] European Commission, "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - A European strategy for data," COM (2020) 66 final, 19 February 2020, p. 13, in particular footnote 39 (Hereinafter "A European Strategy for Data").
[66] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, OJ L 172, 26.6.2019, p. 56–83.

shared under restrictions and still be FAIR.[67] In this context it is coherent that the European data strategy announces to foster the business-to-government/business-to-business data sharing.[68] **However, the guiding principle should be "as open as possible, as closed as necessary".**[69] Paths that have already been started should continue in future funding programs, according to which certain open access obligations apply (see Horizon 2020). Therefore, it should be clear when training data contain confidential commercial information or personal information – which could be valid reasons for restrictions, or when further action can be taken as the anonymisation of personal information within training data. In this context, it should be supported that the Commission is aiming for a single European market for data in a Data Act. This is what the Commission demands in its European strategy for data.[70]

b. IP AWARENESS

In terms of interoperability and reusability, the FAIR principles encourage license terms to be clear and unambiguous.[71] The first requirement for this is that data producers are aware of their own IP rights that arise in their work. Possible rights chains may have to be traceable so that data falling under database protection, training data containing copyright protected works or other property rights can be made accessible and copyright designation requirements can be fulfilled.

It is also important not to create false expectations about IP rights that actually do not exist, for example, if training data are not protected under IP rights. Hence, as an example, data should not be incorrectly declared as protected or – with actual IPR protection – falsely provided with a public domain mark. Also, from the re-user's perspective, legal certainty must be sought.

[67] Final Report and Action Plan from the European Commission Expert Group on FAIR Data, 2018, p. 21.
[68] A European strategy for data, p. 13.
[69] Open Research Data Pilot in Horizon 2020 follows this principle, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm (accessed 11 June 2020). Emphasis added.
[70] A European strategy for data.
[71] R1.1 of the FAIR principles states: „(Meta)data are released with a clear and accessible data usage license".

Practical guidelines and dialog-based storage and upload mechanisms should be fostered, particularly for SMEs and researchers and their institutions.

c. New IPR, reform existing IPR?

The World Intellectual Property Organisation (WIPO) has already taken up the discussion about AI regulation with regard to intellectual property.[72] The key issue is whether works and inventions that are not created directly by humans can be protected by copyright or related rights and patent law. This issue concerns mainly the "output" of AI technologies.

The European Approach to AI and the Commission's statements on IP regulation regarding AI should run in sync. The introduction of new property rights must be carefully justified as they can have a major impact on information law regulation. It would therefore have to be evaluated precisely whether and to what extent new industrial property rights are incentivizing or rather undermining the intended pioneering role in the field of artificial intelligence, and where such new rights unduly restrict the public interest. A revision of the existing framework should be taken into account, in particular the Database directive.

d. Role of repositories

Repositories play an important role for making data available.[73] As the White Paper states: "Equally important is investment in key computing technologies and infrastructures."[74] Infrastructures, including repositories for both storage and making available of data, must hold licenses in accordance with the status of IP rights. Standardized licenses such as the CCPL[75] and open data licenses should continue to be used here, as the Re-usability criterion R.1.1 requires

[72] https://www.wipo.int/about-ip/en/artificial_intelligence/faq.html; https://www.wipo.int/about-ip/en/artificial_intelligence/policy.html; WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI), Draft issues paper on Intellectual Property and Artificial Intelligence, WIPO/IP/AI/2/GE/20/1 (accessed 11 June 2020).
[73] See for the Research Data Repository RADAR, FIZ Karlsruhe: https://www.radar-service.eu/sites/default/files/publications/RADAR_FAIR_Principles.pdf (accessed 11 June 2020).
[74] White Paper on AI, p. 8.
[75] https://creativecommons.org (accessed 11 June 2020).

"clear and accessible data usage licenses". In addition, license machine readability must be guaranteed. Here it should be evaluated whether existing certification procedures for repositories such as the Core Trust Seal[76] are sufficient for the AI context and the European approach. **If not, the European Approach might consider encouraging the establishment of certification mechanisms at Union level, equivalent to Article 42 GDPR in data protection law.**

9. MENTAL MANIPULATION

Mental manipulation has been recently defined as a form of influence, hidden and intentional, which subverts another person's capacity for conscious decision-making by exploiting in particular his/her cognitive, emotional, or other decision-making vulnerabilities.[77] There are many opportunities to manipulate individuals and some increasingly easier to exploit with new developments in the field of Artificial Intelligence. For example, online behaviour analysis and psychological assessment may become easier and more accurate. Personalized tactics designed to silently exploit people vulnerabilities, e.g. via the use of text synthesis, image synthesis or video manipulation techniques, may become more sophisticated and difficult to detect.[78]

[76] https://www.coretrustseal.org (accessed 11 June 2020).

[77] See Daniel Susser et al. Online Manipulation: Hidden Influences in a Digital World (December 23, 2018). 4 Georgetown Law Technology Review 1 (2019), pp. 14-23.

[78] See Vesselin Popov, Lecture: How to Wield the Data-Driven, Double-Edged Sword: Navigating the Ethics of Psychological Profiling and Targeting with Big Data, in The Political Economy of Data: On Psychometrics, Deep Learning and Net Populism, edited by Daniel Irrgang, Peter Weibel and Siegfried Zielinski, ZKM, Center for Art and Media, Karlsruhe (2018) pp. 4-11; Panel Discussion with Vesselin Popov, Florian Cramer, Matteo Pasquinelli, Peter Weibel, Siegfried Zielinski and Daniel Irrgang, in The Political Economy of Data: On Psychometrics, Deep Learning and Net Populism, edited by Daniel Irrgang, Peter Weibel and Siegfried Zielinski, ZKM, Center for Art and Media, Karlsruhe (2018) pp. 11-25; see also Ref. Ares(2019)2266862, EU project SHERPA report, Security Issues, Dangers and Implications of Smart Information Systems, March 2019, pp. 1-68, p. 59, available at https://pdfs.semanticscholar.org/4d32/adabdb782382c1fe1f0237a0585771f2b700.pdf; Philip N. Howard et al., "Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration". Journal of Information Technology & Politics (2018) 15 (2): 81-93; Philip N. Howard, and Bence Kollanyi, Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum (2016)

A personal robot may trick and mislead its owner into purchasing products if designed to do so.[79]

Such forms of mental manipulation are characterized by hidden interferences with another person's mental integrity which undermine mental control capacities and exploit mental weaknesses.[80] The subversion of a person mental capacities is a mental/psychological intervention which poses mental/psychological risks for the individual. Manipulative interventions have the potential to impair mental capacities or alter decisions, mood, preferences and will formation. **Despite the risks which such mental manipulation through AI technologies may pose**, **it is unfortunate that it is not mentioned in the White Paper**. Whereas the "Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics" mentions that "explicit obligations for producers could be considered also in respect of *mental safety risks* of users when appropriate."[81], no definition of "mental safety risks" has been provided. Does the European Commission recognize psychological risks of AI, such as the risk of being manipulated, as a mental safety risk? **We recommend that the risk of manipulation should be explicitly covered not only within the concept of product safety but also as one of the major risks regarding the use of AI in our society**. New incentives for innovation that make

available at SSRN: https://ssrn.com/abstract=2798311; see also Ben Nimmo, "Measuring Traffic Manipulation on Twitter." Working Paper 2019.1. Oxford, UK: Project on Computational Propaganda, available at https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/01/Manipulating-Twitter-Traffic.pdf; Fenwick McKelvey and Elizabeth Dubois, Computational Propaganda in Canada: The Use of Political Bots, Working Paper No. 2017.6, Oxford, UK: Project on Computational Propaganda, available at http://blogs.oii.ox.ac.uk/politicalbots/wp-content/uploads/sites/89/2017/06/Comprop-Canada.pdf; The Guardian , Maeve Shearlaw, From Britain to Beijing: how governments manipulate the internet (2015) available at https://www.theguardian.com/world/2015/apr/02/russia-troll-factory-kremlin-cyber-army-comparisons; Samuel C. Woolley and Philip N. Howard, Computational Propaganda, Oxford University Press (2018) p. 8.

[79] See European Parliamentary Research, Panel for the Future of Science and Technology, The ethics of artificial intelligence: Issues and initiatives Service, March 2020, p. 18, available at https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf#page=100&zoom=90,39,749 (accessed 11 June 2020).

[80] See Jan C. Bublitz and Reinhard Merkel, Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination, in Criminal Law and Philosophy (2014) pp. 51-77, p. 60.

[81] White Paper on AI, p. 15.

the mental integrity or mental health of EU citizens a key ethical requirement should also be created. Technology companies and political institutions which legislate on AI in Europe need credible and independently trained ethicists since often technology is actually designed (and desired) to mislead, trick, manipulate users (e.g. dark patterns). That is why designers need to take ethical considerations (as well as legal considerations) seriously.[82]

With regard to the discussion on the regulatory framework for AI, in the White Paper (Section 5) requirements are called for regarding transparency and the keeping of records in connection with the "programming of the algorithm, the data used to train high-risk AI systems, and, in certain cases, the keeping of the data themselves."[83] As stated in the White Paper, "these requirements essentially allow potentially problematic actions or decisions by AI systems to be traced back and verified."[84] Therefore, investigating and promoting solutions that by design aim at ensuring that the use of AI systems does not lead to outcomes entailing mental manipulation, e.g. by allowing the traceability of the workings of the algorithm to enable the corrections of mistakes and undesirable outcomes, is of great importance.

In addition, paramount importance should also be given to investigating how to legally protect the mental sphere of EU citizens. Article 3 of the *European Union Charter of Fundamental Rights and Freedoms* (CFREU), for example, guarantees that "everyone has the right to respect for his or her physical and *mental integrity*."[85] So far the European Court of Justice has limited its interpretation of the article to the fields of medicine and biology.[86] Thus, no guidance from it exists as to whether mental manipulation could fall within the scope of "mental integrity." However, Court could interpret Article 3 CFREU in such a way that mental manipulation by AI is provided under this

[82] See Mark Coeckelbergh and Thomas Metzinger, Tagesspiegel, Europe needs more guts when it comes to AI ethics, April 2020, available at https://background.tagesspiegel.de/digitalisierung/europe-needs-more-guts-when-it-comes-to-ai-ethics.
[83] White Paper on AI, p. 19 and 20.
[84] White Paper on AI, p. 19.
[85] Article 3 CFREU, emphasis added.
[86] CJEU – Joined Cases C 148/13, C 149/13 and C 150/13 / *Opinion A, B and C*, 2014; Netherlands/ Council of State /ECLI:NL:RVS:2018:1802, 2018; see also the analysis of Jan C. Bublitz and Reinhard Merkel (no 80 above); Andorno Roberto and Ienca Marcello, Towards new human rights in the age of neuroscience and neurotechnology, Life Sciences, Society and Policy (2017) 13:5.

right. The normative framework should keep up with the new technological advances of the last 21 years (the Charter was adopted in 2000) and extend the protection of people's mental integrity to the era of AI.[87] The Commission should encourage protection from any use of AI that breaches the fundamental right to integrity under Article 3 CFREU.[88]

10. TRAINING AND SKILLS

Section 4 (C) of the White Paper on Skills focuses largely on upskilling the workforce to become fit for the digital age. Upskilling can have different foci. We recommend that upskilling through adequate training schemes should focus especially on the following aspects:

- **The human in the loop should have adequate technical skills** in order to understand how a result, based on an AI application, was reached. This understanding should help him especially to be able to challenge the result and to potentially reverse it. This would both make human control effective and real, and would help control the legality (accuracy) of results.[89]
- Special attention should be paid to **technical skills in relation to biometric technologies**. The controller of a certain biometric application should make sure that biometric matches are verified by a competent human being before decisions in relation to individuals are taken and that thus the controller can live up to his legal obligation to ensure the accuracy of the personal data he processes.[90]
- Next to ethical training,[91] **data protection training should be compulsory.**

[87] Andorno Roberto and Ienca Marcello (no 86 above).
[88] For a better understanding of how to do so under Article 3, see Francesca Pichierri and Mark Leiser, "Post-Panoptic Surveillance: State-Sponsored Manipulation" (forthcoming).
[89] See on human agency Strasbourg, 23 January 2017 T-PD(2017)01 Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (T-PD), "Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data, point. 7.
[90] Article 5 (1) (d) GDPR and Article 4 (1) (d) Directive 2016/680.
[91] White Paper on AI, p. 6.

- Training opportunities about the functioning of complex AI technologies should be offered not only to the controller. They should be offered especially to the **members of the independent supervisory authorities**, who can then inspect the different applications and suggest improvements, but also understand better how individual results were reached and thus help data subjects exercise their rights, e.g. where they claim that a certain (biometric) match is a false positive. In addition to training, more technical experts should be employed by such supervisory authorities.
- Individuals subject to AI applications should be informed not only about the existence of a certain AI application, but also **about how it operates and what the consequences are.**

11. Research and Development Projects on AI

A significant number of AI initiatives and applications are developed in the framework of research projects, including projects which receive EU financing. The research results, including the scientific methods relied on and the training data, are not always made publicly available. **This does not render the development of AI applications transparent, e.g. about the scientific methods used, the success of the technology, etc.** This is especially the case where the technology, if operational, would pose serious risks to peoples' rights and freedoms and where the underlying scientific methods are debatable.[92] This prevents an informed public debate on the technologies under development and about their ethical and legal desirability in a certain society.

Furthermore, technologies might easily and quickly reach the market after their development stage, as very often a clear exploitation plan is required by certain EU funded projects. It is therefore essential, for the sake of legal and ethical compliance, **that already at the development stage there is sufficient independent oversight from legal and ethical experts who are not members of the research project and whose opinions should be publicly available.** For

[92] Daniel Boffey, "EU border 'lie detector' system criticised as pseudoscience," The Guardian, 2 November 2018, https://www.theguardian.com/world/2018/nov/02/eu-border-lie-detection-system-criticised-as-pseudoscience (accessed 11 June 2020).

example, EU funded projects on AI, which are reviewed annually by external reviewers, should always have on their reviewers board independent ethical and legal experts. Also, the technological readiness for a certain technology to become operational should be carefully assessed and the exploitation potentially delayed, e.g. until the technology is improved or safeguards are put in place. The weaknesses of the technologies should be disclosed to avoid overreliance on their accuracy.

In that respect there should be clear guidance about how transparency should be balanced against trade secrets and **IPR issues**. A situation should be avoided in which IPR and trade secrets are absolute and are used to prevent any disclosure, especially with regards to controversial applications, and to prevent the public and ethical and legal experts from participating in the development of such technologies.

Finally, in the development of any new AI application, whether in the framework of research projects or not, there should be **adequate end – user involvement**. This is especially the case for AI projects which are supposed to become operational soon after their development and are not theoretical and purely exploratory. End – users should be understood to mean not only the potential controllers and processors and their specific needs. In this way the risk that a technology is developed only according to the wishes of the developers and is imposed on the end-users where either the technology would be irrelevant or does not have the wished functionalities or is not user-friendly, would be avoided.

End-users should also be understood to mean the individuals who would be potentially subject to the technology. This would allow citizens to become aware of the risks of the technology in order to build in adequate safeguards and would spark the public debate about whether a certain society wishes to use/be subject to a certain application.

12. Conclusion

The present response to the consultation on the Commission's White Paper on AI focuses on six main legal issues:

- The legal problems in relation to the Commission's proposed high-risk approach for the future regulation of AI technologies. It is suggested that this approach might not be effective because it is difficult to classify AI applications as either high- or low-risk. It would also allow certain technologies to "escape" regulation. It is thus recommended to develop a more-level based risk assessment which can nuance different automated decision-making systems more effectively.
- We point out what safeguards in relation to AI in the law enforcement field are missing and how the legal framework (whether the future regulation on AI or existing instruments such as Directive 2016/680) should be improved in order to provide a more robust protection in relation to AI technologies in this sensitive field. Thus, our response recommended strengthening the information obligations about the existence of automated decision-making, ensuring independent oversight of the algorithms, enshrining in law the obligation for the controller to keep detailed records about the algorithmic decision-making operations, ensuring the accuracy of profiling results, opening up the opportunity for more judicial review of AI applications and anchoring sunset clauses in the respective regulations (legal bases) on AI which are particularly privacy-invasive.
- We focus on the problems arising from the proposed types of requirements for high-risk applications and the proposed testing centres. It is suggested that the proposed requirements should be specifically set in accordance to the different levels of automated decision-making systems. Furthermore, it should be discussed how these requirements can be measured and enforced in practice. To ensure these requirements it is important to create an independent regulatory body with a cadre of experts who are able to test and judge the design, development and deployment of AI applications as legally and ethical applicable.
- We discuss aspects concerning the processing of biometric data, such as the importance of assessing the necessity and proportionality of biometric

applications prior to their deployment, as well as paying attention to the unique accuracy problems they pose. Thus, a system should be designed to allow individuals to inspect the accuracy of the biometric matching and to challenge it effectively.

- We point out legal problems concerning the Commission's proposed liability in relation to AI products, especially in relation to the attribution of responsibility.
- We draw attention to the importance of considering mental manipulation as one of the major risks regarding the use of AI in our society and that it should be therefore covered by the concept of product safety. Furthermore, our contribution emphasized the necessity of investigating how to legally protect the mental sphere of EU citizens, which can be particularly undermined by manipulative practices through the use AI technologies.

We also discuss more practical topics such as the FAIR principles which the Commission mentions in the White Paper and focused specially on the licensing issues related to them. The present response further made suggestions in relation to the future R&D work on AI, e.g. to ensure more regulatory and public oversight of the research methods and results, the ethical and legal compliance of research, as well as independent and balanced review of publicly funded projects. Our response also looked at what should be one of the foci of the training programmes in relation to AI, so that training is not restricted only to AI developers, who need training also in ethics and data protection. Also controllers who work with and are responsible for the results of AI applications, data protection authorities and ordinary citizens need adequate technical skills to safely interact with and be able to control AI technologies. Finally, we note that the White Paper also did not discuss topics such as AI in lethal weapons, which is an important aspect of AI.