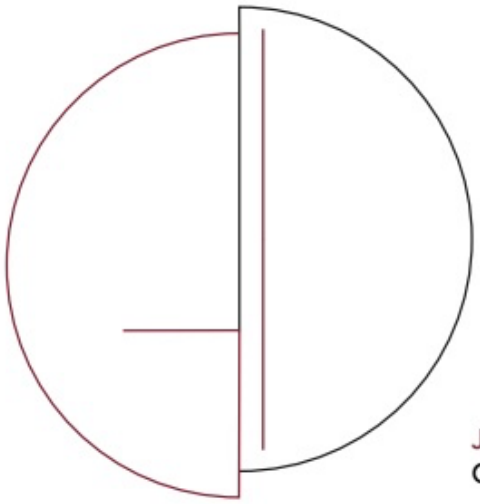
Ref. Ares(2020)3355667 - 26/06/2020



Juridical Observatory on Digital Innovation
Osservatorio Giuridico sulla Innovazione Digitale

CONSULTATION ON THE WHITE PAPER COM(2020) 65 FINAL

"ON ARTIFICIAL INTELLIGENCE - A EUROPEAN APPROACH TO EXCELLENCE AND TRUST"

Rome, 19 May 2020

INTRODUCTION: SCOPE OF THIS DOCUMENT

This document has been prepared and authored by Professor Salvatore Orlando ([1]) in the context of the submission by the Juridical Observatory on Digital Innovation ("**JODI**") ([2]) of its responses to the online questionnaire relevant to the open public consultation procedure on the EU Commission's White Paper - COM(2020) 65 final - On Artificial Intelligence - A European Approach to Excellence and trust (the "**Questionnaire**" and the "**White Paper**", respectively).

This document addresses exclusively certain issues relevant to **Section 2 of the Questionnaire "An ecosystem of trust"** referring to a series of options for a regulatory framework for AI, as set up in **Chapter 5 of the White Paper.**

More specifically, this document contains 3 sections aimed at explaining more in detail the following:

- JODI's comments included in the box of Section 2 of the Questionnaire relevant to 'any other concerns about AI':
  «With reference to the seven "key requirements" mentioned in the White Paper, and in connection with the concern that AI's systems may be employed to exploit human behaviour, it is suggested to identify a specific notion of the "fairness" principle consisting in the duty of care of AI's systems for human behavioural vulnerability including but not limitedly to AI's applications affecting consumers».

([1]) Full Professor of Private Law at the University of Rome "Sapienza" - https://web.uniroma1.it/deap/informazioni-su-salvatore-orlando
([2]) Transparency Register number 164451038125-33 - https://web.uniroma1.it/deap/ogid

- JODI's comments contained in the box of Section 2 of the Questionnaire relevant to the response 'other' in connection with the option to limit the introduction of new compulsory requirements to 'high-risk applications' of AI:
  «With reference and in addition to the purposes of AI's applications to be considered "high risk as such", as indicated and listed on a non-exclusive basis in the White Paper, it is suggested to expressly mention also the marketing purpose so as to cover all AI's marketing applications, including those directed at influencing non-economic decisions».

- JODI's comments contained in the box of Section 2 of the Questionnaire relevant to the question 'Do you have any further suggestion on the assessment of compliance?'
  «(1) Ex-ante compliance should include "fairness by design"; (2) Ex-post enforcement should be focussed on collective remedies and public enforcement for their higher ability to act as deterrents against AI's unfair marketing applications compared to individual ex-post remedies».

A. ***JODI'S COMMENTS INCLUDED IN THE BOX OF SECTION 2 OF THE QUESTIONNAIRE RELEVANT TO 'ANY OTHER CONCERNS ABOUT AI'***

AI's applications not only can unintentionally make discriminatory decisions, as remarked in the White Paper, but also have the technological potential for tracking, processing and utilizing personal or statistical data and other data concerning the daily habits of people with the malicious purpose of ***(i)*** clustering people with respect to average conditions of people's vulnerability (such as average conditions of ignorance, lack of circumspection or attention, lack of psychological or economic ability to resist or to react, impressionability, hypochondriasis etc.) on the basis of observed, provided, derived or inferred data (such as data relevant to income and wealth, health, age, family of origin, instruction, online behaviour etc.) and ***(ii)*** targeting real people on the basis of an algorithmic association of the targeted people to such categorizations/clusters of vulnerability, with the aim of taking advantage of certain average conditions of subjective vulnerability which are presumed to be borne by the targeted persons, and, thus, with the aim – and, in any case, with the result - of creating the conditions for distorting (as opposed to influencing) the typical decision-making processes of the targeted people in connection with certain specific decisions (e.g. a certain purchase decision, a certain complaint decision, a certain electoral decision, a certain healthcare decision etc.) or in connection with decisions relevant to long-term affiliations or status (e.g. decisions relevant to the affiliation or the continued affiliation to communities or organizations, such as online communities, political parties, churches, foundations, associations, unions etc).

When electoral or political decisions, or decisions relevant to the affiliation or continued affiliation to communities or organizations are at stake, these would typically be distorted by online targeting practices maliciously disseminating fake news including fake news aimed at perpetuating or exacerbating biases or at radicalizing ideologies or beliefs.

In EU law, the purpose of protecting human decisional vulnerability is traditionally relevant to consumer-protection legislation, as it is made evident by the very important "framework" directive of 2005 intended for combating unfair business-to-consumer commercial practices: Directive 2005/29/EC.

As known, the focus of the discipline of Directive 2005/29/EC has been set in the concept of the consumer's transactional decision and, more specifically, around certain average consumer's conditions of vulnerability associated with the consumer's decision-making process. Acting in a manner contrary to the requirements of professional diligence (Article 2(h) of Directive 2005/29/EC) means acting in a way that allows the trader to take advantage of certain average consumer's conditions of vulnerability associated with the consumer's decision-making process. In essence, it is forbidden to influence consumers' transactional decisions in such a way as to take advantage of certain average consumer's conditions of vulnerability. Whenever the trader puts in place commercial practices that present this feature he is said to act against the professional diligence and to 'distort' (as opposite to 'influence') the consumers' economic behaviour, and, in brief, to act unfairly (Article 5, para. 2 (a) and (b) of Directive 2005/29/EC).

This is expressly recognized in the White Paper, at **page 13 and foot note 42**, where expressly quoting Directive 2005/29/EC among the rules to be taken into account as regards the protection of fundamental rights and consumer rights; at **page 14** [«*economic actors remain fully responsible for the compliance of AI to existing rules that protects consumers, any algorithmic exploitation of consumer behaviour in violation of existing rules shall be not permitted and violations shall be accordingly punished*»]; at **page 16** [«*The EU has a strict legal framework in place to ensure inter alia consumer protection, to address unfair commercial practices and to protect personal data and privacy. In addition, the acquis contains specific rules for certain sectors (e.g. healthcare, transport). These existing provisions of EU law will continue to apply in relation to AI, although certain updates to that framework may be necessary to reflect the digital transformation and the use of AI*»].

However, since the realm of AI's applications is not limited to business-to-consumer commercial practices, it appears sensible to remark that the issue of fairness/unfairness in connection with AI's systems should be aimed at protecting human decisional and behavioural vulnerability and at combating practices exploiting human behaviour and taking advantage of human decisional and behavioural vulnerability in ANY possible field of application of AI's systems, e.g. online targeting for electoral purposes, health care decisions, affiliation to online communities, political parties etc.

In addition to the White Paper, the following documents have been issued during the last year, which also support the idea that a specific notion of fairness should be highlighted with respect to AI's systems, consisting in the need to protect human decisional vulnerability against AI's applications that may be exactly designed for distorting the human ability to make free and aware decisions:

- The *Ethics Guidelines for Trustworthy AI* published on 8 April 2019 by the Independent High Level Expert Group on Artificial Intelligence set up by the European Commission

- The *Review of online targeting: Final report and recommendations*, dated February 2020, released by the UK's Centre for Ethics and Innovation

- The *AI Rome Call*, signed in Rome on 28 February 2020 by representatives of the Pontifical Academy for Life, Microsoft, IBM, FAO and the Italia Government.

- *Shaping Europe's digital future*, Communication from the EU's Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, dated 19 February 2020

- *A European strategy for data*, Communication from the EU's Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, dated 19 February 2020

More in particular:

The *Ethics Guidelines for Trustworthy AI* published on 8 April 2019 by the Independent High Level Expert Group on Artificial Intelligence set up by the European Commission: **page 11** ["This also requires adequate respect for potentially vulnerable persons and groups, such as workers, women, persons with disabilities, ethnic minorities, children, consumers or others at risk of exclusion] and definition of Vulnerable Persons and Groups at **page 38** ["No commonly accepted or widely agreed legal definition of vulnerable persons exists, due to their heterogeneity. What constitutes a vulnerable person or group is often context-specific. Temporary life events (such as childhood or illness), market factors (such as information asymmetry or market power), economic factors (such as poverty), factors linked to one's identity (such as gender, religion or culture) or other factors can play a role. The Charter of Fundamental Rights of the EU encompasses under Article 21 on non-discrimination the following grounds, which can be a reference point amongst others: namely sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age and sexual orientation. Other articles of law address the rights of specific groups, in addition to those listed above. Any such list is not exhaustive, and may change over time. A vulnerable group is a group of persons who share one or several characteristics of vulnerability."]; **page 12** ["Moreover, the use of AI systems should never lead to people being deceived or unjustifiably impaired in their freedom of choice"]; **page 16** ["AI systems can sometimes be deployed to shape and influence human behaviour through mechanisms that may be difficult to detect, since they may harness sub-conscious processes, including various forms of unfair manipulation, deception, herding and conditioning, all of which may threaten individual autonomy."]; **page 34** ["Human beings should always know if they are directly interacting with another human being or a machine, and it is the responsibility of AI practitioners that this is reliably achieved. AI practitioners should therefore ensure that humans are made aware of – or able to request and validate the fact that – they interact with an AI system (for instance, by issuing clear and transparent disclaimers). Note that borderline cases exist and complicate the matter (e.g. an AI-filtered voice spoken by a human). It should be borne in mind that the confusion between humans and machines could have multiple consequences such as attachment, influence, or reduction of the value of being human."].

The *Review of online targeting: Final report and recommendations*, dated February 2020, released by the UK's Centre for Ethics and Innovation: see especially the many paragraphs on "Autonomy and Vulnerability" at **page 36 ff. especially pp. 38-39 f.** (paragraphs 108, 109, 110, 111 and 112) **, page 53** (paragraphs 153 and 154)**, page 58** (paragraphs 170, 171 and 172)**,** Glossary **page 130 ff.**

The *AI Rome Call*, signed in Rome on 28 February 2020 by representatives of the Pontifical Academy for Life, Microsoft, IBM, FAO and the Italia Government: **first page** ["The transformations currently underway are not just quantitative. Above all, they are qualitative, because they affect the way these tasks are carried out and the way in which we perceive reality and human nature itself, so much so that they can influence our mental and interpersonal habits. New technology must be researched and produced in accordance with criteria that ensure it truly serves the entire "human family" (Preamble, Univ. Dec. Human Rights), respecting the inherent dignity of each of its members and all natural environments, and taking into account the needs of those who are most vulnerable. The aim is not only to ensure that no one is excluded, but also to expand those areas of freedom that could be threatened by algorithmic conditioning."]; second **page** ["AI-based technology must never be used to exploit people in any way, especially those who are most vulnerable."]

***Shaping Europe's digital future*, Communication from the EU's Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, dated 19 February 2020 :** A fair and competitive economy: A frictionless single market, where companies of all sizes and in any

sector can compete on equal terms, and can develop, market and use digital technologies, products and services at a scale that boosts their productivity and global competitiveness, and consumers can be confident that their rights are respected **(page 2)**; The more interconnected we are, the more we are vulnerable to malicious cyber activity. **(page 5)**; In the digital age, ensuring a level playing field for businesses, big and small, is more important than ever. This suggests that rules applying offline – from competition and single market rules, consumer protection, to intellectual property, taxation and workers' rights – should also apply online. Consumers need to be able to trust digital products and service just as much as they would any other. There is a need to pay attention to the most vulnerable consumers and to ensure the enforcement of safety laws, also in relation to goods originating from third countries. Some platforms have acquired significant scale, which effectively allows them to act as private gatekeepers to markets, customers and information. We must ensure that the systemic role of certain online platforms and the market power they acquire will not put in danger the fairness and openness of our markets. (page 8); Delivering a new Consumer Agenda, which will empower consumers to make informed choices and play an active role in the digital transformation (Q4 2020). (page 10)

***A European strategy for data*, Communication from the EU's Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, dated 19 February 2020:**

The European data space will give businesses in the EU the possibility to build on the scale of the Single market. Common European rules and efficient enforcement mechanisms should ensure that:

- data can flow within the EU and across sectors; - European rules and values, in particular personal data protection, consumer protection legislation and competition law, are fully respected; … **(page 5)**

Since increasingly large amounts of data are generated by consumers when they use IoT devices and digital services, consumers may be faced with risks of discrimination, unfair practices and 'lock-in' effects. Considerations of consumer and innovation empowerment underlie the provisions on data access and reuse of the Payment Services Directive **(page 10).**

The fairness principle has been spelt out at **page 9 of the White Paper** as forming, together with "diversity" and "non discrimination", one of the "seven key requirements" identified in the Guidelines of the High-Level Expert Group, and welcomed by the Commission in the Communication COM(2019) 168.

However, little or no discussion is contained in the White Paper as to how to differentiate "fairness" from "diversity" and "non discrimination", nor a specific notion of fairness is provided therein. This brings about the risk of dilution or amalgamation of the concept of "fairness" with those of "diversity" and "non discrimination".

We also note – and this is a further element of concern – that the recent **European Parliament's Draft Report with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL))** of 21 April 2020 does not even mention the fairness principle, while only addresses the non-discrimination principle.

In conclusion, and based on the foregoing, JODI's proposal is to address the above concerns by way of:

a) identifying a specific notion of the fairness principle with respect to AI consisting in the duty of care (of AI's systems) for human decisional and behavioural vulnerability;

b) stating that the scope of application of the above specific notion of the fairness principle should be as broad as to cover the extent of application of any AI's systems, and shall not be limited to AI's applications affecting consumers.

B. ***JODI'S COMMENTS CONTAINED IN THE BOX OF SECTION 2 OF THE QUESTIONNAIRE RELEVANT TO THE RESPONSE 'OTHER' IN CONNECTION WITH THE OPTION TO LIMIT THE INTRODUCTION OF NEW COMPULSORY REQUIREMENTS TO 'HIGH-RISK APPLICATIONS' OF AI***

The comments under Section **A.** above suggest that the general approach to limit the introduction of new compulsory requirements to those AI's applications identified as high-risk in accordance with the two cumulative criteria indicated at **page 17 of the White Paper** (i.e. sector and use) should be combined with measures and rules that recognize the potential for AI to harm people's fundamental rights in any field of application of AI, and in particular the potential for AI's systems to exploit human behavioural and decisional vulnerability, in the sense indicated under Section **A.** above, in any possible field or sector of application of AI.

As already noted under Section **A.** above, the first lines at **page 14 of the White Paper** read as follows: «*… economic actors remain fully responsible for the compliance of AI to existing rules that protects consumers, any algorithmic exploitation of consumer behaviour in violation of existing rules shall be not permitted and violations shall be accordingly punished*».

Based on a literal interpretation of the above passage, one might conclude that any algorithmic exploitation of human behaviour would be possible except only in case the human behaviour in question is the behaviour of a "consumer", pursuant to EU existing rules protecting consumers.

While it is fair to say that the above conclusion would hardly sound sensible, it is also fair to remark that the White Paper does not request to draw this (unreasonable) conclusion.

In fact, the "high-risk" approach based on the two cumulative criteria is neither absolute nor exclusive under the White Paper.

The "high-risk" approach based on the two cumulative criteria is not intended to be absolute in the White Paper, while it is clearly intended to be an "in principle – approach", as stated in the last lines at **page 17 of the White Paper** («*The mandatory requirement contained in the new regulatory framework on AI (see section D below) would in principle apply only to those applications identified as high-risk in accordance with these two cumulative criteria*.») and confirmed at **page 18 of the White Paper**, where, in fact, some cases are indicated in which the use of AI applications for certain purposes should be considered as being "*high-risk **as such***" (i.e. without the need to verify whether or not the two cumulative general criteria are met), and the White Paper is clear in stating that these cases are indicated "*as an illustration*", i.e. on a non-exclusive basis («*… As an illustration, one could think in particular of the following …*»).

The driving concept for assessing the "high-risk as such" feature, is that of the intended **purpose** of the AI's application (**page 18 of the White Paper**: «*Notwithstanding the foregoing, 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 …*»).

In order to identify what is the purpose of the AI's applications that shall be deemed to be relevant in the above normative approach for combating AI's applications that are exactly designed to distort human behaviour by taking advantage of people's average conditions of decisional vulnerability, the concepts underlying the discipline of Directive 2005/29/EC are certainly of help.

The subject matter of of Directive 2005/29/EC consists in business to consumer commercial practices that can influence the consumer's transactional decisions. Influencing consumers' transactional decisions is not per se prohibited (to give a pair of indisputable examples: advertising and marketing in general are not prohibited as such). The discipline is precisely intended at fixing general criteria for establishing upon which conditions influencing is legitimate ('fair') and upon which conditions it becomes illegitimate ('unfair').

As already mentioned, under the norms of Directive 2005/29/EC, it is forbidden to influence consumers' transactional decisions in such a way as to take advantage of certain average consumer's conditions of vulnerability. Whenever the trader puts in place commercial practices that present this feature he is said to act against the professional diligence and to 'distort' (as opposite to 'influence') the consumers' economic behaviour, and, in brief, to act unfairly (Article 5, para. 2 (a) and (b) of Directive 2005/29/EC).

On the basis of the above comments, the "purpose" of the AI's applications that shall be deemed relevant for affirming the fairness principle is that of influencing people's decisions, which is, by definition, the intended purpose of any AI's marketing applications.

For the reasons already illustrated above, there is no reason to limit the fairness principle to AI's marketing applications affecting consumers; on the contrary, the fairness principle should apply to any AI's marketing applications, including those directed at influencing non-economic decisions.

It is therefore proposed that, in addition to the cases indicated as an illustration of the purposes to be considered as high-risk as such at **page 18 of the White Paper**, the following should also be included:

- **The use of AI applications for purposes of marketing would always be considered "high risk", irrespective of the sector of activities to which the marketing practices are addressed, and including in case of AI's marketing applications directed at influencing non-economic decisions.**

Finally, it is very clear, based on the above, that careful consideration would have to be paid to the issue of conflict between the many EU norms governing marketing practices in the many different sectors where AI's applications may be employed for marketing purposes, and to the various criteria that should be observed for combining those norms also in view of the expected new compulsory requirements under the future regulatory framework for AI (see Section **C.** below). This is especially true considering that, as mentioned, AI' s marketing applications are not limited to business to consumer marketing. As a consequence, while the Directive 2005/29/EC on business to consumer unfair commercial practices could and should of course be taken into consideration as a source for general principles and guidance for all AI's marketing applications, it is nevertheless true and undeniable that the rules of this directive do not directly apply to all AI's marketing applications.

C. ***JODI'S COMMENTS CONTAINED IN THE BOX OF SECTION 2 OF THE QUESTIONNAIRE RELEVANT TO THE QUESTION 'DO YOU HAVE ANY FURTHER SUGGESTION ON THE ASSESSMENT OF COMPLIANCE?'***

It is herein suggested that the design of AI's marketing applications should be made subject to ex-ante requirements aimed, in essence, at prohibiting and avoiding ***(a)*** the processing of data with the purpose of creating clusters of vulnerable people with respect to stated categories of behavioural and decisional vulnerability, and ***(b)*** the triggering of functionalities aimed at targeting people on that basis for obtaining a certain decision or behavioural response.

Based on the experience of the implementation of Directive 2005/29 EC on unfair commercial practices in the various Member States, it is recommended to give a central role to collective remedies and public enforcement for their higher ability to act as deterrents against AI's unfair marketing applications compared to individual ex-post remedies.

Respectfully submitted

Rome, 19 May 2020

Prof. Salvatore Orlando