Consultation on the European Commission’s Proposal for Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act)

Submission of Dessislava Fessenko

8-6-2021

1. Introduction

The technical complexity, pervasive penetration and multifaceted social implications of the use of artificial intelligence (“**AI**”) in every-day life of EU citizens unequivocally necessitate the introduction of clear and sensible rules for development, deployment and use of AI. The endeavors in this respect of the European Union institutions, and of the European Commission in particular, are laudable. The strive to shape those rules in a way that promotes innovation and the uptake of AI in the European Union (the “**EU**”) is also praiseworthy.

However, any future regulation of AI should also be introspective, smart and fair. Introspective in the sense that it should learn from and avoid deficiencies and past mistakes in other relevant areas of business regulation (e.g. data protection, antitrust). Smart in the sense that it should have built in already now a conceptual design and a toolkit that would allow addressing issues arising from the advance in AI (e.g. the emergence of strong and general AI). Fair in the sense that a future regulation of AI should – while creating a favourable business environment – also account and cater for the interests of society at large, including end users and citizens more generally.

This position paper focuses on analysis of the proposal for a regulation of the European Parliament and the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) (the “**Draft AI Regulation**” or the “**Draft**”) and recommendations on the conceptual design of and certain fundamental rules under the Draft with view to the need of an introspective, smart and fair legal framework for design, development, deployment and use of AI in the EU.

2. Risk-based approach and ethical/human rights backbone

The Draft AI Regulation retains the risk-based approach and the ethical/human rights backbone initially contemplated in the European Commission’s White Paper on AI[1] and the recommendations of the European Parliament of October 2020,[2] while building upon regulation in other areas (e.g. data protection, consumer protection, standardization) to address the more specific technical aspects of AI design, development, deployment and use. This approach generally makes sense given how contextually dependent the deployment and use of AI are and how quickly the technology evolves. Too legalistic requirements might result in under- or overregulation.

However, clearer emphasis needs to be put in the final text of the Artificial Intelligence Act on outcomes, rather than on formal course of action. Art. 13 to Art. 15 of the Draft already require attainment of certain overall outcomes (transparency, accuracy, etc.). Yet, other obligations under Chapters 2 and 3 of Title III presuppose mere process-like actions (e.g. documenting, record-keeping, risk management based on step-by-step actions) in order to demonstrate compliance. If the emphasis in this framework does not firmly and eloquently lie with ultimate outcomes – and that various processes are one of the means to that end – compliance would morph into a “box-ticking”- and “window-dressing”-type of adherence to the formal requirements under the Artificial Intelligence Act, as is the case with similar rather “technical” requirements under the European Union Genera Data Protection Regulation.[3]

3. Prohibited use of AI

The explicit bans under Art. 5 of the Draft AI Regulation are a welcome development in an attempt to limit the use of AI to manipulative and/or privacy-invasive ends.

However, the proposed language and approach have some deficiencies, as set out below.

3.1 AI employing subliminal techniques or exploiting vulnerabilities

The language of the bans under Art. 5(1)(a) and (b) is vague and the legal standard introduced is too general. As a result, it is unclear:

(i) Whether certain forms of AI – such as algorithms for behavioural advertising, including in political campaigns – fall into that category of "subliminal techniques", given that they do not generally cause physical and psychological harm but may have larger societal impact.

(ii) To what extent a specific AI system may fall under these bans due to typical context, in which it is used, and/or users' typical perception of the interaction with the AI system. Context and perception are always part of the wider equation when assessing distortion. Possible distortive effects may not be dismissed merely because context/perception are outside of the control of the provider of the AI system, as recital 16 of the Draft suggests. For example, a seven-year-old child conversing with a personal assistant with a human-like voice can be reasonably expected to perceive the statements of the personal assistant (e.g. recommendations about "Harry Potter" books) as trustworthy (and even compelling) because of the similarity of this communication with such with a well-intended adult. The provider of the AI system may still model typical context and perception and try to curb the distortive effects of its system in such circumstances (e.g. by deactivating advertising functionalities of the algorithm on which the personal assistant is based).

Clarifications to that effect in recital 16 of the Draft and fine-tuning of the text of Art. 5(1)(a) and (b) are recommendable.

3.2 Real-time biometric identification versus facial recognition

The ban under Art. 5(1)(d) on real-time biometric identification for law enforcement purposes, as defined, captures identification techniques only. Other forms of facial recognitions (e.g. emotion recognition, which is widely used in public and private domains)[4] and such used for other purposes are permitted, though considered high-risk. Some of those other forms are subject under the Draft only to self-assessment and not to a conformity assessment by a notified body. Which would still allow for a largely unvetted wider spread, including by governments and private actors for extensive surveillance.

Extension of the ban under Art. 5(1)(d) of the Draft to include more generally facial recognition (including emotion recognition) intended for mass surveillance and/or social controlling is necessary.

3.3 Real-time biometric identification for law enforcement purposes

The legal standard for exempted use of "real-time" remote biometric identification for law enforcement purposes – when "*strictly necessary*" – is not waterproof enough in my view. As worded in the Draft AI Regulation,[5] this standard implies necessity, and not mere efficacy. However, this standard does not imply admissibility only as a last resort. In practice, a fair amount of cases may still merit such exempted use. Same goes for the exempted use prior to obtaining court authorization. Concerns around state surveillance (examples of which unfortunately persist also in EU member states[6]) would not be effectively curbed in this way.

Amendments to Art. 5(1)(d) are needed. Suggested language (amendments underlined) reads:

"(d) the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary <u>as a last resort, respects the essence of the fundamental rights and freedoms and constitutes a necessary and proportionate measure in a democratic society</u> for one of the following objectives:"

The proposed approach reflects an established standard for human rights protection adopted by the Council of Europe in international treaties concerning data protection,[7] as well as the standard for exempted biometric identification under Art. 9(2) of the General Data Protection Regulation.

4. Principle-based obligations, non-discrimination and fairness

The Draft AI Regulation introduces[8] explicit obligations around transparency, accuracy, robustness, cybersecurity, data quality and governance. They are welcome – in my view – given the fundamental concerns around those aspects of AI operations.

Yet, the Draft somehow elegantly misses out on explicit obligations for non-discrimination and fairness. The Draft introduces requirements for data quality/governance, which is much needed given that training, testing and input data predetermine to a certain extent predictions and inferences that AI systems come up with. However, those requirements are not sufficient to allay possible discriminatory or unfair outcomes from the use of AI.[9] The requirements for data quality essentially cater only for the quality of the input into the AI system. They do not alone warrant bias-free conceptual design, technical development, practical deployment and use of the system.[10]

Explicit obligations must be introduced in the final text of the Artificial Intelligence Act to the effect that AI systems must be designed, developed, deployed and used in a way that ensures non-discrimination and fair treatment of the individuals or categories of individuals concerned by the use of those AI systems.

5. Compliance requirements

The Draft AI Regulation introduces[11] a complex web of compliance requirements that hinge on risk assessment, quality management and data management. They would require internal controls and dedicated resources by AI developers/providers in order to comply.

Whether this would spur or rather stifle innovation would largely depend on the interpretation and application of those requirements by the authorities in charge (a potentially large amount of them at national level, given the institutional framework under the Draft). There are sufficient grounds for skepticisms in this respect given how similar institutional frameworks have worked out in other areas (e.g. data protection,[12] consumer protection, anti-money laundering).

If formalistic interpretations/application prevail, this would likely result in:

(i) The AI regulation turning into a *de facto* barrier to entry/expansion for smaller AI providers (e.g. start-ups);

(ii) Partnerships between start-ups and bigger, more resourceful organizations (most of which are US-based corporations) in development of AI;

(iii) Industry consolidation (similar trends have been observed in similarly regulated industries[13]).

With this hindsight, consistent, sensible and pragmatic interpretation of and guidance on the requirements under the Artificial Intelligence Act would be vital for promoting the development of AI in the EU. The European Artificial Intelligence Board envisaged under Art. 56 of the Draft must be solely responsible for providing guidance on the interpretation and implementation of the Artificial Intelligence Act. The national authorities should play a limited role (as part of enforcement) in those areas as this would inevitably result in divergence and legal uncertainty.

6. Self-assessment

Self-assessment emerges as the main form of control mechanism under the Draft. While this allows for a light-touch, less onerous regulation for the industry, similar attempts in other areas (e.g. antitrust, data protection, anti-money laundering) have shown that this approach may result in suboptimal outcomes for businesses (ultimately, more legal uncertainty, especially when coupled with high fines, as is the case under the Draft) and society (general ethical and compliance race to the bottom).

Those side effects might be resolved by envisaging additional mechanisms in the final text of the Artificial Intelligence Act for regulatory guidance (such as more structured soft guidance, comfort letters to be issued by national supervisory authorities following consultations among themselves and with the European Artificial Intelligence Board, etc.). Amendments to that effect in the governance framework under the Draft are worthwhile.

7. Checks and balances

AI may impact the standing of various societal groups (consumers, patients, employees, actual/potential competitors, etc.) who are typically not part of the self-assessment process, and of the deliberations between providers of AI systems, on the one hand, and notified bodies or regulators, on the other. In order for regulatory processes to yield better informed and fairer outcomes overall (including for affected parties and society at large), more checks and balances appear necessary, in particular:

(i) A possibility must be envisaged in Chapters 2 and 3 of Title III and Chapter 1 of Title VIII for external independent oversight (e.g. by academia and civic organizations) of the self-assessment and post-marketing monitoring of high-risk AI systems that are widely employed (e.g. to a number of end users above a certain threshold).

(ii) A possibility must be envisaged in Art. 43 of the Draft for third interested parties to express views in the course of the conformity assessments of AI systems as to the possible effects of their use. This could be done either in the form of open calls for expression of views or of market testing as part of the respective assessment procedure.

(iii) Regulators and notified bodies must be required to seek feedback from actually or potentially affected groups as part of proceedings and the conformity assessments, respectively, concerning high-risk AI systems.

(iv) A possibility under Chapter 3 of Title VIII must be provided for third affected parties to file complaints with the supervisory authority in cases of alleged breaches by developers/deployers/users of AI systems of ethical principles or legal obligations specific to the Artificial Intelligence Act (e.g. regarding robustness, accuracy, transparency), which would typically not be covered by redress mechanisms under data protection, non-discrimination, consumer protection, etc., legislation.

*

*　　*

[1] White Paper On Artificial Intelligence - A European approach to excellence and trust, Brussels, 19.2.2020 COM(2020) 65 final, available at: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en .

[2] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)), available at: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html .

[3] E.g. the principles of transparency and fairness under Art. 5 of the General Data Protection Regulation, OJ L 119, 4.5.2016, p. 1–88, (the "GDPR") and respective implementing requirements to that effect have been understood and practiced by way of production of lengthy compliance documentations, which may not always correspond to the spirit of the GDPR.

[4] On the same topic also Crawford, K. (2021). *Time to regulate AI that interprets human emotions.* Nature, available at: https://www.nature.com/articles/d41586-021-00868-5; Madhumita, M. (2021). *Emotion recognition: can AI detect human feelings from a face?* The Financial Times, available at: https://www.ft.com/content/c0b03d1d-f72f-48a8-b342-b4a926109452; ARTICLE 19 (2020). *Emotional Entanglement: China's emotion recognition market and its implications for human rights*. Available at: https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf

[5] See recital 16 et seq. of the Draft AI Regulation.

[6] Panyi, S., P., Pethő, A. (2021), *Hungarian journalists and critics of Orbán were targeted with Pegasus, a powerful Israeli cyberweapon*. Direct36, available at: https://www.direkt36.hu/en/leleplezodott-egy-durva-izraeli-kemfegyver-az-orban-kormany-

kritikusait-es-magyar-ujsagirokat-is-celba-vettek-vele/ . Nikolov, K (2021), *Bulgarian secret services suspected of eavesdropping opposition politicians*. Euractiv, available at: https://www.euractiv.com/section/politics/short_news/bulgarian-secret-services-suspected-of-eavesdropping-opposition-politicians/

[7] Such as those adopted by the Council of Europe in Convention for the protection of individuals with regard to processing of personal data (Convention 108+).

[8] In Chapters 2 and 3 of Title III.

[9] To that effect, for example, Wachter, S., Mittelstadt, B. (2019) *A Right to Reasonable Inferences:Re-thinking Data Protection Law in the Age of Big Data and AI*. Columbia Business Law Review, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3248829

[10] Wachter, S., Mittelstadt, B. (2019), supra. O'Neil, C. (2016). *Weapons of Math Destruction*. Crown

[11] In Chapters 2, 3 and 5 of Title III.

[12] Empirical evidence to that effects available in Jia, J. et all (2019). *GDPR and the Localness of Venture Investment*. American Economic Association, available at: https://www.aeaweb.org/conference/2020/preliminary/paper/7dfztbb4 .

[13] E.g. pharma, with respect to more complex solutions (e.g. innovative treatments).