KU LEUVEN CiTiP

CENTRE FOR IT & IP LAW

KU Leuven Centre for IT and IP Law's Comments to the proposed Artificial Intelligence Act

Authors: Koen Vranckaert, Lydia Belkadi, Noémie Krack, Emine Özge Yildirim, Jan Czarnocki, Katherine Quezada-Tavarez, Jenny Bergholm.

Introduction

First, we offer our congratulations to the Commission for this great piece of legislation, consisting of a solid combination of general principles and very concrete answers to the many questions raised in recent years regarding AI. It is a very rich document, offering the long-awaited standards and clarity about AI applications and uses, efforts and results that are welcomed enormously. Nevertheless, there are some reservations and areas of improvement that need to be highlighted, as listed below.

On the scope of 'artificial intelligence system' under the proposed AI Act (Article 3(1)) - Koen Vranckaert

We note that the Commission has departed from prior, open conceptions of 'artificial intelligence', as had been proposed by its Communication for AI, as well as the reports of the AI High-Level Expert Group. Rather, the Commission has opted for a 'closed list' of techniques that constitute 'artificial intelligence' for the purpose of regulation. The reference to techniques match the requirement of legal certainty and ease of enforceability. However, this approach may also invite AI developers to attempt 'designing around' the proposed AI Act (insofar as that is possible), developing new techniques to not have to comply with the proposed AI Act. We understand that the Commission grants itself the power to add new AI techniques to the closed list as listed under Article 4. However, we believe that the option should be considered to keep the list under Annex I, but as an exemplative rather than an exhaustive listing of techniques covered by the notion of 'artificial intelligence', in order to ensure resilience of the proposed AI Act's provisions in the face of rapid technological innovation.

On the regulation of biometric categorisation, emotion recognition and real-time remote biometric identification systems - Lydia Belkadi

The establishment of new sets of safeguards applicable to AI systems processing human features is welcomed. In particular, the definitions of *emotion recognition system, biometric categorisation system,* and *remote biometric identification system* are promising (article 3.34 to 38). This new set of definitions implement innovative methods of concept building. For instance, the definition of *real-time remote biometric identification system* introduces an interesting mechanism to avoid circumvention. The wording of these definitions also expresses the collective impact of these systems ('natural persons'; 'on the basis of their biometric data').

However, some points of attention will require further clarifications.

Article 3

The wording of the definition of *biometric categorisation system* (article 3.35) seems confusing. Indeed, it seems to mix prohibited grounds of discrimination (e.g. ethnic origins, sexual or political orientation) and types of soft biometric characteristics (e.g. hair color, eye color, tattoos) under the qualification of *specific categories*. Soft biometric characteristics may indeed be *used as proxies* for discriminatory processing, such as ethnic origins, sexual or political orientation. Nonetheless, the formulation of the definition should precise the wording and clarify the intention of the legislator further.

Furthermore, the definition of *remote biometric identification system* introduces some problematic components. In particular, it is unclear whether time ('without a significant delay') and space ('at a distance') are relevant emphases to characterize these systems. Indeed, the risks associated with biometric identification systems usually stem from their technical architecture. In that context, two elements should be emphasized: the constitution of centralized biometric databases and the necessity to process the data of every person present in a given space. This collective impact of biometric identification systems is not apparent in the current definition, which remains largely individualistic.

The interactions between the definitional frameworks of the proposed AI Act and existing legal instruments may cause unforeseen difficulties. As detailed in Recital 7 of the proposed AI Act, the new definitional framework applicable to biometric systems builds upon the established notion of *biometric data* as defined in Regulation 2016/679, Regulation 2018/1725, and Directive 2016/680. If this interpretative consistency is in principle necessary, the current definition of biometric data under European laws still enshrines unresolved interpretative challenges.

Article 5

As underlined by the EDPB and the EDPS in their Joint Opinion 5/2021, the proposal fails to adequately mitigate the societal and group risks posed by these AI systems. To this extent, it is important to further consider and integrate the risk analysis proposed under this Joint Opinion.

Real-time and *post remote identification* should be regulated consistently whether in *physical* or *online* spaces. Indeed, the proposal does not provide any element justifying how these distinctions mitigate the intrusiveness of the processing or comply with the necessity and proportionality requirements. Similarly, emotion recognition and categorisation systems should be regulated more strictly.

Additionally, it remains unclear whether the sole focus on *identification* systems, as opposed to so-called '*authentication'* or *identity verification* systems, is desirable. Recent evolutions in practice have seen the emergence of remote (contactless) authentication systems operating in real-time (e.g. remote iris, face, or palm authentication). If the identity verification function does not present the same level of risk as identification systems, it remains unclear why authentication systems have been fully excluded from the definitional framework and the new transparency requirements established by the proposed Act.

Finally, additional interpretative guidance on the qualifications of subliminal techniques beyond a person's consciousness and the exploitation of individuals' vulnerabilities will be crucial. Both of these criteria are conceptualized in very abstract terms. For instance, would covert biometric processing, including emotion recognition or biometric categorisation, qualify as a subliminal technique beyond a person's consciousness? Would the covert use of biometric categorisation systems qualify as exploiting individuals' vulnerabilities? Other unresolved issues arise from the specification that such systems should cause a material distortion of a person's behavior. For instance, how would such a distortion be proven or documented? And how would the causality link between the AI system and the distortion be established?

On the prohibition of AI systems intended for social scoring - Emine Özge Yildirim

In the AI Act's current state, private social scoring is not enumerated as a prohibited practice under Article 5(c). While we welcome that public social scoring is finally considered a practice that curtails human rights with the threat of mass surveillance, private social scoring should also not be considered as posing a lesser degree of danger to fundamental rights or autonomy by EU policymakers. For instance, *the High Level Expert Group on A*I found that any form citizen scoring – whether conducted by public authorities or private actors – can endanger principles of human autonomy and non-discrimination. Thus, the HLEG suggests that if their practices impact human rights, private social scoring should also be banned. Additionally, even though the provision seems to be *prima facie* prohibiting practices interfering with cognitive liberty, the proposed regulation did not seem to take into account the effect of any sort of social scoring on self-determination, mental autonomy, and autonomous decision-making. Many times, individuals, with the fear of private or public surveillance, try to fit in the mold of the rules of the majority. In other words, they conform to the not-so-novel paths of the society, dominated by majoritarian ideologies. Unfortunately, the proposed regulation did not seem to think far ahead regarding how individuals could silence themselves and their actions, caused by the chilling effect of such social scoring. Consequently, I urge the policymakers to consider that mental autonomy does not only have to be bypassed by subliminal techniques to deem the practice an illegitimate interference. In the era of digital power asymmetries between the watchers and the watched, the 'ordinary' chilling effect of living under the threat of being constantly watched by

powerful private or public actors could also lead to harm to or interference with mental autonomy and cognitive liberty, resulting in the creation of conformist societies without any substantial self-determination, meaningful democratic participation, and lively debate. Finally, policymakers should think twice before allowing practices that would silence or chill the already marginalized communities and minorities further, by not providing adequate or perhaps equal protection concerning their cognitive liberty and mental autonomy, which are deemed to be the precursors of almost every individual liberty.

On the risks of overregulation and harm to innovation - Jan Czarnocki

The categorisation of AI systems in several risk categories, as well as the requirements imposed upon the developers of AI systems, are likely to create a high administrative burden. This high administrative burden may become an excessively high compliance burden which will hamper start-ups and individual researchers and developers. The innovation sandboxes mentioned under Articles 53 *et seq.* will likely not compensate for the high compliance burden of (over)regulation.

A few examples can be listed here:

- The definitions of 'emotion recognition' (Article 3(34)) and 'biometric categorisation systems' (Article 3(35)) are dependent on narrowly understood biometric data processing. As such systems are not truly biometric data processing in and of themselves, it would make sense to place these out of the scope of the AI Act Proposal. Nonetheless, whether they fall under the AI Act's scope or not is currently unclear.
- From the perspective of an AI developer, it will be difficult to develop a heuristic or benchmark to assess whether or not their AI systems will be high-risk or not. The current definition also relates mostly to the use of AI in specific sectors, which risks being over-inclusive. Either standardisation must clarify this gap, or the provision must be abolished. The requirement for the developer to nonetheless assess the risk of its AI system prior to deployment is likely to create a chilling effect.

Moreover, the establishment of the Artificial Intelligence Board again invites Member States to re-establish specific supervisory bodies to oversee the compliance with the rules regarding Artificial Intelligence. In order to avoid an infinite multiplication of institutes, the competences of the Artificial Intelligence Board should be bestowed on already existing institutions as much as possible.

On the list of prohibited practices, the definition of high-risk AI systems, the data governance requirements, the database for high-risk systems as well as several unclarities, with a focus on law enforcement - Katherine Quezada-Tavarez

<u>Art. 5: Prohibited practices</u>

Title II provides a welcome list of prohibited practices. However, it might be missing some important points. First, significant gaps remain in the Art. 5 prohibitions regarding menacing government uses (other than law enforcement) of different forms of remote biometric surveillance, such as the use of real-time biometric applications in immigration and refugee decision-making, which raise equivalent concerns to those in law enforcement. Moreover, the wealth of concerns relating to real-time biometric systems are not limited to the use by the public sector, but also to private uses of the technology, currently not banned.

Second, the range of exceptions to the rule in Art. 5 risks creating loopholes that authorities could exploit. For instance, the exception for the "prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack" makes the prohibition too narrow and could leave room for discretion as it is for national authorities to decide whether a certain use case is permitted or not. Law enforcement authorities already have unparalleled discretion to use violence and intrude on rights and liberties, therefore the importance of setting clear and strict rules to worrisome uses of AI technology in this sector.

Third, some of the hoped-for bans are absent. In particular, the expected prohibition on the use of AI to automatically recognize gender and sexual orientation is lacking. Using digital clues in an attempt to predict someone's gender or sexuality is fundamentally flawed because the technology tends to limit gender to a simple binary and, as a result, usually causes harms to persons who do not fit into these limited categories, such as trans and non-binary people. Automatic gender recognition technology is incompatible with the EU's commitment to human rights, and should thus be included in the list of prohibitions in the final iteration legislative text.

<u>High-risk AI systems</u>

While the heading of Annex III.1 reads "Biometric identification and categorization of natural persons", the actual text only covers identification. Thus, it seems to be that biometric categorization systems, such as gender classification AI, are not even considered high-risk, but in principle only subject to the transparency obligations in Art. 52, while in fact they should be put on the ban list (as argued above).

Also, according to Art. 43, only biometric identification systems will be subject to third party conformity assessment. In other words, other AI used in law enforcement (such as predictive policing, often found to be affected by biases), while classified as high-risk, will not have strong oversight as they will only be subject to self-assessment. Given the grave risks to fundamental rights, it is advisable to revise and strengthen the requirements.

<u>Art. 10: Data and data governance</u>

The "by-design" approach promoted by the Commission is also a welcome aspect of the proposal. However, some requirements for high-risk AI applications may need to be revised according to their feasibility as they may be difficult to comply with. Regarding the data quality requirements, while the datasets described in the regulation ("relevant, representative and free of errors") are the dream goal for developers, such data might not even exist yet, and, if it does exist, might not be available for development purposes. For instance, data generated in sectors

that tend to be characterized by systemic biases (e.g. law enforcement) will almost inevitably reflect those biases, thus making the data quality requirements unachievable for AI systems that need to use such data.

Article 60: Database for high-risk systems

An unexpected and welcome addition to the proposal is the creation of a new EU-wide database of high-risk AI systems. This is a good measure that will foster more transparency about AI uses across the Union and facilitate the public scrutiny of those.

Unclear issues

There is unclarity about certain aspects. For example, what does it mean "to put in the market" an AI system? Would AI systems developed in-house by law enforcement agencies fall under the scope of the regulation (as this would not be marketed, but limited to in-house use)?

Another issue to clarify is whether a different regime will apply for research and prototype building, or whether all rules would be the same in all situations (unless within regulatory sandboxes).

AI & media in the AI act - Noémie Krack

Prohibited practices by Noémie Krack and Koen Vranckaert

It is not clear whether the use of subliminal techniques could cover some AI systems already used in practice such as the recommender system or system used for targeted advertising. The requirements imposed to manipulative AI, such as the use of subliminal techniques or the use of a specific vulnerability of a specific group of persons, as well as the requirement of intent, can result in these provisions having a very limited scope, to the point of lacking any *effet utile* in practice. More incidental manipulative systems (such as targeted advertising) is therefore not likely to be covered. In order for the protection of manipulative AI to have an *effet utile* beyond what is present in existing law, the scope of this system should be included to allow protection against more subtle forms of manipulation.

High-risks AI systems

AI applications in media are not listed in the high risk list in the annex. However, more and more studies are being conducted and show the risks associated to it : discrimination, bias, threats to fundamental freedom such as freedom of expression when AI systems are used for content moderation. Therefore, the Commission should consider adding AI recommender systems in media to be added to the list.

Transparency obligations for certain AI systems

As explained in the impact assessment report for the AI act, the Digital Services Act is indeed already containing some transparency provisions on recommender systems (article 29 DSA) but

only for very large online platforms and targeted advertising for online platforms (article 24 DSA). However, there is room for improvement and for even better taking into account the considerable impact that AI has on media.

The scope of article 52 is not clear which will lead to legal uncertainty. It is not clear why §1 is targeting providers and §2&3 users.

Could article 52 §1 be applicable to recommender systems or robot journalism?

Article 52§3 covering deep fakes provides the following exception: this obligation does not apply the exception where “necessary for the exercise of the right to freedom of expression and the right to freedom of arts and sciences guaranteed by the Charter and subject to appropriate safeguards for the rights and freedoms of third parties”. The scope of the exemption is not clear, no further information can be found in the recital or in the impact assessment documentation. In addition, there is also an exemption for ‘personal non-professional’ use which asks question in terms of impact for this provision. Article 52§3 should also contain an obligation to make the deep fake identification information undeletable in case of transfer or further modification of the material in order not to lose track of the deep fake original information.

The transparency requirements should include more precisions on what should be communicated (the type of information), when (at which stage this should be revealed) and how. Providing more information such as some transparency provisions present in the Digital Services Act (for instance article 13) will help to gain a better understanding of the obligations for the users and providers and ensure teeth to the provision.

Regarding the relationship between the AI Act and the GDPR - Jenny Bergholm

It is important that the EDPB/EDPS Joint Opinion 5/21 is taken into account in its entirety. As is stressed in the Opinion, it is of utmost importance that the AI Regulation is aligned with the scheme of the GDPR. This is detriment in order to to ensure a coherent application of both the GDPR and the AI Regulation, and an effective protection of EU fundamental rights. Rights and obligations stemming from data protection law are based on the aim to both protect the fundamental right to privacy and data protection, as well as other related fundamental rights. These rights and obligations must hence also be efficiently enforceable when personal data is processed by AI systems of different levels of risk. Consequently, a clarification with regards to the relation between the two regulations is important.

It is clear, that the GDPR applies when personal data is processed with the help of artificial intelligence. Even if that fact is not doubted, the relation between the GDPR and the AI Regulation should be clarified. The (long term) quality of the AI Regulation would benefit from clear and strong references to the principles of the GDPR, and so would the innovation community, the internal market and the data subjects.

One point of concern regarding the alignment between the data protection regime and the proposal concerns the approach to risk management. Whereas certain artificial intelligence practices are directly prohibited, most emphasis is put on high-risk AI systems. Those are subject to internal market conformity procedures, with notified bodies and CE marking of conformity. Other AI systems, those that are not considered high-risk, are subject to transparency obligations (Article 52). This risk approach based system is not only detrimentally different from an obligation perspective in relation to data protection law, but also concerning the processing of personal data of AI systems. As the EDPB/EDPS also point out in their opinion, it should be further stressed, that this risk management perspective is not clearly aligned with that of the GDPR. Also, should the regulator found it unreasonable to align the two, the discrepancies should be clearly addressed in a recital. Further, the impact of the risk assessment on the compliance with the GDPR/AI regulation need to be addressed.

It is understood that the AI Regulation is a risk assessment based approach. The scale of risk is graded from unacceptable to minimal risks. Risk-based approaches are also common in cybersecurity legislation, and hints thereof can also be found in the 2020 Cybersecurity Strategy. From a data protection perspective, it can be noted that some risk management elements also exists in the GDPR, such as the data protection impact assessments, as discussed by Gellert and the Article 29 Working Party. Also data protection law is developing in a risk-based direction, as can be seen in the recently published Standard Contractual Clauses for international data transfers, which allow a risk-based approach.

Nevertheless, the rights and obligations of the GDPR stretch further than only high-risk operations. .[A1] The issue is, that even though the AI Regulation is intended to complement the GDPR, it provides very little clarity of processing of personal data by any other AI systems than high-risk AI systems. The Proposal includes a declaration that the proposed AI Regulation does not affect the application of the GDPR, but more guidance in the legal text on how the Regulation should be applied with regards to processing of personal data is missing. The lack of clear reference to the GDPR and other data protection legislation has been raised by the EDPB and the EDPS in a joint opinion.[A2] This is especially the case for data processing of AI systems which are not considered as high-risk systems. Even for high-risk AI systems, it mainly considers [A3] special categories of personal data [A4] . A clear example is Article 10 of the draft regulation, concerning data and governance. The article is set to regulate the training, validation and testing of data sets of high-risk AI systems. It refers to the data collection, but with no reference to rights and obligations of the data subjects. Only in Article 10(5), the GDPR is mentioned, as the providers of systems may process personal data of special categories if strictly necessary for the purpose of ensuring bias monitoring, detection and corrections. In such cases, the provider of the system need to adopt appropriate safeguards, such as technical limitations on the re-use of data and privacy-preserving measures. However, read in the light of the GDPR, many of these safeguards are already *per-se* obligatory due to the GDPR and the principle of data minimisation and privacy by-design and by-default.

Further, it is not clear how the recommendations of the High-Level Expert Group on AI highlighted the need for AI systems to guarantee privacy and data protection, especially for information provided by the user of a system but also data generated about the data subject

while using AI-based tools. The European Commission White Paper on AI already identified risks to data protection and privacy rights by *e.g.* using AI to retrace or de-anonymise data, even for data sets which do not, as such, include personal data.

Other elements of attention - Noémie Krack

The place of research in AI

It is crucial to clarify the situation of research when it comes to the AI act. There is a strong call for more provisions, information concerning the role of research and an extension of the research exemption.

European Data sets

The development of an EU sovereignty in data and ethical use of the data is crucial. There is the need to have European data sets fulfilling ethical and legal principles and rules in order to base ethical and legally compliant progress in research and commercial application of AI systems.

Transparency

Transparency has different faces and multiple notions can be understood as transparency, there is a need to clarify the legal vocabulary between transparency, explainability, interpretability, record keeping, communication to the end-user. Some elements present in previous policy text on AI such as the HLEG guidelines are not taken in the proposal and more consistency on this aspect will be beneficial to transparency in general.

In addition, making use of signs, symbols could be a way forward and to develop an easy understanding for the consumers. The text is complex and hard to read for non-lawyers and might create problems of enforcement and transparency towards the population already suffering from information asymmetry.

Consumers and society: what role do they have?

There is a strong presence/mention of fundamental rights in this internal market proposal (Article 114 TFEU as legal basis) : 80 mentions through the all document and a clear will to establish a text with positive impact for citizens. However, in the text there is no concrete redress and action mechanisms for citizens, civil society or consumer rights associations to act against non-compliant providers or users. Specific provisions should be established especially on the burden of proof, as there is already an asymmetry of information context. It is important to have provisions helping victims for lodging complaints to the competent authorities and asking for redress.