Position Paper and Suggested Changes for the European Artificial Intelligence Act

Kaspar Rosager Ludvigsen*

June 2021

1 Key Messages

- The choice to not let AI be its own distinct field, and let it be part of the product legislation in the European Union is admirable and pragmatic.
- The Act in its current form is well shaped and structured, and because of its emphasis on both rights as well as innovation of the "product" (AI) itself, is bound to have a similar effect on EU-law as GDPR did, despite not creating a whole new legal area[1].
- Despite this, there are certain issues with some preambles as well as articles, with the biggest one being Art. 5. It is proposed that the prohibitions there become concrete and direct. Further suggestions for clearer wording in both preambles and other articles are included in this paper as well.

2 Comments on structure and choice of the New Legislative Framework

The proposal for an Artificial Act is one the most interesting, well written and thoroughly thought-out pieces of legislation that has come out of the European Union. Attempting to regulate an innovative, immersive and hastily growing sector is important, but doing this while deliberately wanting to apply existing rules on products to it is equally so. Wanting artificial intelligence (AI) to be seen in the same manner as other product legislation is a very wise move, as they

*PhD Student at the University of Strathclyde, Scotland, Danish Lawyer, and member of the (Cyber)Security group at the Department of Computer and Information Sciences, as well as member of the Strathclyde Centre for Internet Law and Policy in the Law School at the same university. Kaspar is furthermore a former civil servant with independent decision-making powers in appealed cases within Public Law in Denmark.

[1]For GDPR, being an act/regulation instead of a directive, since data protection was already regulated before it entered into force.

are already sold as services and products, and because this is a strong measure of fairness[2]. Taking inspiration and being part of the New Legislative Framework, we see the usual suspects of Notified Bodies, Conformity Assessments etc., which is very encouraging and a tried and true method in the Union. Unlike a lot of these other products however, AI works with data separately and a more intensive way, which puts them thoroughly apart. But this may be more or less solved in the Act, by putting a very intense focus on the requirements for staff of the regulators, which is unique but otherwise very fitting to the area, as well as re-acknowledging the increasing value of the GDPR outside of a few exceptions. Furthermore, the Act also has several well written articles that touch upon the (cyber)security aspects which are extremely important regarding AI, so these are not purely left out for authorities and guidance.

3 Suggestions for specific changes and comments of the Regulation to better reach the Goals of the Act

Regardless of the comments above, there are certain small changes which can heavily improve the draft of the AI Act. Some changes are proposed here, while others that could be mentioned are left to different providers of feedback.

3.1 Preambles

Suggested changes and comments on preambles follow below. It has to be noted, that some preambles need to be changed if certain articles are as well, but we will not list those here.

3.1.1 Preamble 6.

The draft does not work with clearly defined AI, it does the exact opposite with the use of Annex I. Legal certainty can therefore not be attained if AI is not properly defined, according to this preamble, which makes no sense because of the goals in the memorandum. Preamble 6 should not contradict the rest of the Act, and a change of the wording is in place.

A change to preamble would be as follows:

AI systems are defined in a technology neutral manner, which provides flexibility to accommodate future technological developments. The definitions are based on the current key functional characteristics of the software, in particular the ability, for a given set of human-defined objectives, to generate outputs..

The Commission choose a technology neutral approach with a few technology specific choices, but they clash with public statements regarding the Act, as well as with some parts of the Act itself. By leaving the rest of preamble 6 as

[2]Not attempting to create extremely specialised *lex specialis* is justified in the memorandum of the draft, which is equally well written and worth reading.

it is, it allows for the Regulation to focus on catching all AI possible (without hampering innovation), and leave it up for the Annex I to define further. While not all outputs for the AI may be relevant for anyone else but the AI or its user/provider, this does not mean they should not potentially be able to be regulated long term.

3.1.2 Preamble 15.

This preamble stands out[3] for pointing out that AI can be used in novel ways to be *tools for manipulative, exploitative and social control practices.* But pointing out that AI can do this, without preventing existing tools and human means to manipulate, exploit and socially control populations, is contradictive and therefore goes against all existing legislation. What is meant here, is that if the AI Regulation wants to prevent these practices when it comes to AI, the EU should prevent these practices with all types of technologies at every instance (including Member States that make use of these practices themselves). This is especially important for the New Legislative Framework, as this novel approach to preventing these means should be extended to all product legislation in the EU.

There are therefore two ways to improve this preamble, the first being to remove it entirely, or change preamble 15 into: *Like any technology or means to manipulate, exploit and socially control humans, AI must be prevented from being used for such goals as well, regardless of whether it is misused or not*[4]*. Such practices are particularly harmful and should be prohibited because they contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law and Union fundamental rights, including the right to non-discrimination, data protection and privacy and the rights of the child.*

3.1.3 Preamble 47.

Any AI, regardless of high-risk or not, should not be incomprehensible to natural persons. This does not mean that no AI should be useable by anyone but experts, but assuming the opposite seems to go against the spirit of the Act[5]. The following suggested change therefore is: *To address the opacity that may make AI systems incomprehensible to or too complex for natural persons, a certain degree of transparency should be required for all AI systems, with higher amounts of transparency needed for high-risk AI systems. Users should be able to interpret the system output and use it appropriately. High-risk AI systems should therefore be accompanied by documentation and instructions of use and include concise and clear information, including in relation to possible risks to fundamental rights and discrimination, where appropriate.*

[3] Also for the out of context use of "that" in the Act.

[4] This way, the preamble does not sound like these practices are prevented elsewhere, as it does now.

[5] This is written knowing the intention is to regulate non-high-risk as little as possible.

3.1.4 Preamble 68.

Which situations? This preamble stands out, and has little purpose other than predicting technology that does not yet exist, or "AI" being sold as more than it is. Nonetheless, allowing such exceptions for technology that does not exist, seems out of place compared to the rules otherwise, and it does not work as an exception when there is no justification for it.

Two changes can therefore be suggested; either remove the preamble entirely, or change it to this:

Rapid availability of innovative technologies may be crucial for health and safety of persons and for society as a whole. Under extremely exceptional reasons of public security or protection of life and health of natural persons and the protection of industrial and commercial property, Member States could authorise the placing on the market or putting into service of AI systems which have not undergone a conformity assessment. This way, the exception is limited somewhat.

3.2 Articles

3.2.1 Art. 5.

Prohibition of AI is a great idea. But the exceptions in each prohibition makes them work in the exact opposite way. These definitions are not written in a way where AI is prohibited, but instead set up exceptions as to when the AI cannot be used - and open up for their use the rest of the time. Considering the memorandum, this may not have been the purpose of the Commission, and so a list of changes is suggested that will effectively enable prohibition, but includes an exception clause in the end that is far narrower than what has been proposed.

This would follow what the Commission verbally said during their presentation, and it would follow the purpose of the Act in the first place. Several of these prohibited practices can never be used because they will emphalways be considered a threat to the individual and communal rights of all European Citizens, and there is no scientific basis for their usefulness, only commercial and state interests, which cannot suffice when it comes to the potential misuse and power they can wield.

Before we show these changes, the following comments for each prohibited use are listed:

- (A). The prohibited technique seems to be "sloshing", which is very reasonable, with the goal of harming the individual or someone else. But what about sloshing that does not harm anyone, but almost does? What about sloshing to abuse consumers? Why is this not prohibited, or at least discussed in the Act? As said earlier, prohibiting or regulating techniques that humans employ on each other (which is not normally regulated), such as business practices in apps and during civil purchases, is surprising to say the least, but sloshing is always negative (hence it being the opposite of nudging). Why not ban it entirely?

- (B). This is very well written and needs to no changes. It plays well into existing anti discrimination legislation on a EU level.
- (C). Banning the concept of "trustworthiness" or social scoring is necessary. This is not the case further into the paragraph of the article, which is why it must be changed. The distinction between being mistreated because of data taken out of a different situation, versus when it is disproportional or unjust, is interesting and not followed up on elsewhere.
- (D). This is not a prohibition. This allows all EU member states to use real-time remote identification biometric systems[6] within many exceptions, and can theoretically lead to certain member states making use of them as the rule instead of the exception.

The following changes may be worth considering.

It is suggested that (c) is changed to: *the placing on the market, putting into service or use of AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons regardless of the method and whether or not the end result is social scoring;* The current version has means of circumvention based on how the wording of (i) or (ii) is read. They are therefore not worth keeping, as it prevents an actual prohibition. Details about how these scores etc. are generated were removed, as these are bound to change to follow any current trends, this makes the prohibition more technology neutral and lets it effectively ban all means to do so[7]. This follows the oral presentation of the Commission as well (banning all social scoring). The problem with that statement is yet again, that it does not ban private (including in education) social scoring, but this can be changed later.

It is suggested that (d) is changed to: *the use of 'real-time' remote biometric identification systems in publicly accessible spaces* The 3 exceptions are far too wide to allow to stay. And merely focusing on law enforcement seems to miss the purpose of the Act, as private organisations can cause the same damage or may be employed by the State to do the same work. Otherwise, allowing them in general (outside of law enforcement) is not suitable. This is both down to the chance of misuse, the severity of the personal data and profiling that is created, and the proportionality - it is known from surveillance research in general, that there will never be a 100 percent success on anything, and because this kind of technology is employed en masse, the 10 percent or more that go through will lead to hundreds of thousands of false positives/negatives and errors. Regardless of this, it is acknowledged that the member states have interest (and private entities) in using it, which is why narrower exceptions are suggested.

Furthermore, *in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm;* should be removed from Art. 5(1)(a) and (b), because they

[6] These may not even fulfill the criteria to be AI, why include them here then?

[7] Not for private entities though.

allow abuse that does not meet the threshold of the excluded part of each sentence. By making the prohibitions clearer, prevention any sort of abuse is more likely.

Art. 5(2) should be removed, as it is included it in the new version of Art. 5(2) below, same goes for Art. 5(3). Art. 5(4) includes rules for exceptions that are so wide and circumventable, that it should not be included or be relevant at all.

The new Art. 5(2) would then be as follows: *Under very exceptional circumstances, 'real-time' remote biometric identification systems may be individually deployed in publicly accessible spaces. Each individual use of a 'real-time' remote biometric identification system in publicly accessible spaces shall be subject to a prior authorisation granted by a judicial authority of the Member State in which the use is to take place, issued upon a reasoned request and in accordance this Regulation. The authorisation can never be the basis for continuous use of systems.*

The severity of private entities making use of the same technology in public spaces can be great as well, so they should be included in this authorisation.

It is urged that the Commission considers taking as many of any arbitrary requirements that Art. 5(4) could have created, out of the hands of the member states, and therefore see good reason to have them base their authorisation on the Act, in lieu of Art. 5(2) which is changed to Art. 5(3) with the following wording:

The use of 'real-time' remote biometric identification systems in publicly accessible spaces shall take into account the following elements, before and during its authorisation: (a) the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm caused in the absence of the use of the system; (b) the consequences of the use of the system for the rights and freedoms of all persons concerned, in particular the seriousness, probability and scale of those consequences.

Creating an analogy to the kind of authorisation that authorities must have to enter private property, should be used for deploying real-time remote biometric identification systems in public spaces[8] in member states. This guarantees (in theory) the highest possible safeguards for the citizens, and limits the use adequately. By not allowing the technology to be used in emergencies, a technology which like many others employed hastily, has not shown increased likelihood of anything[9], a safer public space is created for the population at large in the EU.

[8] It may necessary to still include the original Art. 5(2), last sentence, but it may be redundant since these assumptions seem to be clear if an equal judicial review is giving for each system, in the same manner in which one is given for each individual investigation in criminal procedure. Private employments of such systems are included because of the issues they can still pose, but it is left up to the member states to create adequate rules for them by the Act being silent on the manner.

[9] Other than scaring and hurting the citizens.

3.2.2 Art. 9.

Purpose limitation can prevent effective testing of the AI. Therefore, the following change to Art. 9(6) is suggested: *Testing procedures shall be suitable to achieve the intended purpose of the AI system.* It is not surprising that "intended purpose" is the model of intention, but since the Regulation goes into "actual use" in for example the prohibitions, the Commission should consider whether to change the wording for the Regulation, regardless of its connection to the product legislation in the EU in general[10].

3.2.3 Art. 10.

Two choices are worth taking noticing in this article. Art. 10(2)(a) uses the term "relevant", which poses the question as to whether there are "irrelevant" design choices. Irrelevant design choices can still affect the AI, so why not just write *design choices*? There is little reason not to exclude "relevant" here.

Art. 10(5) does not acknowledge the different ways which bias can be prevented. It could do so to prevent the creation of the expectation that personal data must be used in any way. This can be considered, but otherwise the section is very agreeable.

3.2.4 Art. 14.

Human oversight is not the only solution to the dangers of AI, and while most of the article is worded appropriately, there is one detail that could be changed. Art. 14(2) mentions that "human oversight shall aim", but choosing "shall" is not appropriate, the human oversight must "aim". The use of "shall" conflicts with the rest of the article, and it is suggested that it is excluded, purely in Art. 14(2).

3.2.5 Art. 15.

Why is "state-of-the-art" not used to describe the state of the (cyber)security of the AI? Because of the consequences of failure of the defences when it comes to AI, it is suggested that Art. 15(4), second sentence is changed to the following: *The technical solutions aimed at ensuring the cybersecurity of high-risk AI systems shall be state-of-the-art.*

This, combined with existing guidance on the subject matter, should guarantee a higher level of defences, which is more than adequate for such potentially dangerous tools.

3.2.6 Art. 41.

Art. 41(3) seems either redundant (in that the high-risk AI may fulfill it regardless of this paragraph) or allows for circumvention of fulfillment of Chapter

[10]This could be achieved with a dedicated article on how intention is perceived differently when it comes to AI than other products, but this suggestion is not included here.

2 in the Regulation. Neither seems to have been the intention, so it is encouraged that the Commission deletes this section for sake of clarity and for the prevention of non-conformity.

3.2.7 Art. 47.

Changes in Art. 5 must mean this article will need to be changed as well. The justifications for changing Art. 5 remain the same as to why Art. 47 should not remain in its untouched form either.

3.2.8 Art. 54.

Art. 54(1)(g) and the article in particular does take into consideration the issues that having developed a model and expecting it to not reveal some of the data or its internal models, if it gets "attacked"[11]. This can be fixed by either writing a dedicated paragraph on how every attempt should be made to prevent attacks from revealing personal data in the future or otherwise, but the issue must be addressed in the Act and not just in guidance, especially considering Annex I.

3.3 Annexes

3.3.1 Annex I.

Seeing technology specific (but still generic) techniques listed is not bad, but if the Commission seriously wants to consider everything that the Industry want to sell as AI, they may want to either include more or expand the explanations for each technique. Otherwise, the intention and its contents are very agreeable.

[11] It is known that adversarial attacks to ML/other models, has been documented to be very doable to a majority of all that is currently called "AI". These may reveal the model, reveal the data (personal too) or other types of information, and should perhaps be regulated here - but this has not been done.