etami

ethical and trustworthy artificial and machine intelligence

RESPONSE TO COMMISSION'S PUBLIC CONSULTATION ON THE 'ARTIFICIAL INTELLIGENCE' ACT

INTRODUCTION

etami is the European organisation for "ethical and trustworthy artificial and machine intelligence", set up as a joint project in 2021. Members of etami include ABB, Atos, AVL, Continental, Deutsche Bahn, DFKI, ELTE University Budapest, KU Leuven, Leonardo, Siemens, TU Berlin, UnternehmerTUM, Volkswagen, and Zalando. The consortium is led by Volkswagen.

The goal of etami is to lead on trustworthy and ethical AI, to create an industry standard of the same and pilot its certification strategies.

This document describes joint feedback of the etami consortium with respect to the AI Act, adopted on April 21, 2021.

GENERAL REMARKS

The etami consortium welcomes the European Commission's initiative on AI and its goal to develop a European ecosystem of excellence and trust in AI. etami supports the risk-based approach taken in the legislative framework, since it prevents it to be stifling for low-risk application development, while addressing high-risk applications.

At the same time, there is a large responsibility at the side of the provider of an AI system, which may stifle development esp. for SMEs.

A general issue with the legislative framework is that the wording is often vague and general. It includes several legal provisions that need clarification (e.g., the prohibition of 'real-time' remote biometric identification systems applies to systems used for law enforcement purposes in publicly accessible spaces. It does not apply to systems used by other public authorities or by private actors). Similarly, the prohibition of AI systems used for social scoring purposes is also limited to those deployed by public authorities. Again, private actors are kept out of the line of fire (Art. 5). The Regulation contains notion that will require additional clarification such as 'psychological harm', 'materially distort the behaviour of a person', … This will eventually be a matter of national law/national judges, and therefore lead to a risk of diversity.

The framework gives insufficient attention to position of persons incurring damage/harm. Redress possibilities are not sufficiently discussed, e.g., w.r.t challenging outcomes or obtaining compensations. There should be accountability frameworks as well as rights for EU citizens when confronted with AI system and their consequences. Attention could also be given to (a reversal) of the burden of proof or other procedural elements. Such elements are of particular for victims (e.g., possibility to file complaints at national authorities). We suggest that the legislative proposal be supplemented by the possibility of collective legal protection (class action), especially since limited transparency may lead to citizens not recognising that they are affected.

TECHNICAL REQUIREMENTS

Article 5:

The following artificial intelligence practices shall be prohibited:

§1. (a) the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness 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;

> While the idea is laudable, the wording is vague and may include, e.g., normal advertisement practices. It very likely also includes recommender systems, which may or may not be intentional? Especially as seen by the citizen, it may not be clear when one is affected.

Article 10 – Data and data governance

§1. High-risk AI systems which make use of techniques involving the training of models with data shall be developed on the basis of training, validation and testing data sets that meet the quality criteria referred to in paragraphs 2 to 5.

> Article 10(1): The concept of 'validation and testing' data sets spelled out here is not meaningful for each parameter adaptation methodology. By adding "where applicable" this problem can be addressed.

§3. Training, validation and testing data sets shall be relevant, representative, free of errors and complete. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons on which the high-risk AI system is intended to be used.

> Article 10(3): This is a bit a chicken-and-egg problem. While the general idea is understood and applauded, it is not practicable as formulated. This should not discuss the data, but rather the data in its use with the methodology.

We suggest a wording along the following lines: "*Training, validation and testing data sets shall be relevant and representative. Errors in the data set shall be statistically negligible for the models that use these data sets. The statistical properties of the AI methods trained with the data sets have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons on which the high-risk AI system is intended to be used.*"

Article 12 – Record keeping

§2. The logging capabilities shall ensure a level of traceability of the AI system's functioning throughout its lifecycle that is appropriate to the intended purpose of the system.

The use of the word "lifecycle" is confusing. This can be interpreted as, logging should be permanent for the whole life time? of a product or process. The corresponding logging costs and resource use may be very impactful.

Article 14 – Human oversight

§1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.

Article 14(1): The meaning of "effectively overseen" is not clear. In particular, when involving complex machinery, the "overseeing" by humans may be technically impossible.

Does this indicate that, in each and every time, a qualified human should be able to "overrule" an AI-based system? Or is this a requirement w.r.t. the interface?

Article 17 – Quality management system

Article 17(1): As what was mentioned for Article 10, the quality management system should not overburden the development process, rather focus on proving performance.
Article 17(2): The choice to make the quality management system proportional to the size of the provider's organisation is surprising and probably allows for loopholes. It would be advisable to make the quality management system proportionate to the complexity cq. impact of the AI system instead.

Article 52 – Transparency obligations for certain AI systems

§3: the definition of "manipulation" is only intuitively clear, and that unclarity may lead to different interpretations. A list of examples may address this.

Article 54 – Further processing of personal data for developing certain AI systems in the public interest in the AI regulatory sandbox

§1. In the AI regulatory sandbox personal data lawfully collected for other purposes shall be processed for the purposes of developing and testing certain innovative AI systems in the sandbox under the following conditions:

(a) the innovative AI systems shall be developed for safeguarding substantial public interest in one or more of the following areas: (i) … criminal offences prevention …; (ii) … public safety/health; (iii) … environment protection …

> Article 54(1): While the idea of regulatory sandboxes is applauded, we ask the commission to not restrict the areas of application.

Article 64 – Access to data and documentation

§1. Access to data and documentation in the context of their activities, the market surveillance authorities shall be granted full access to the training, validation and testing datasets used by the provider, including through application programming interfaces ('API') or other appropriate technical means and tools enabling remote access.

> Article 64(1): This provision would imply full storage of all data, which may lead to prohibitively large data storage capacity.
>
> Article 64(2): Giving access to the full data and/or code of an application is not always possible. In particular, when 3rd-party pretrained models are used, data or source code may not be available. Not being able to deploy such systems will lead to a considerable competitive disadvantage.

DEFINITION OF AI

The definition of AI as laid down in Annex I is very broad, as it includes almost all software; it may cover traditional control algorithms as well as any piece of software which is based on statistical approaches.

At the same time, it is applauded that the definition does not refer to undefined concepts such as "intelligence".

Probably the key issue is that we are talking about parameterised methods where the parameters are determined based on data, using non-convex parameter optimisation – that means that, there is no guaranteed optimal solution.

It is well understood that the definition is only relevant in combination with high-risk methodologies as defined in Annex II. It may be worth pointing that out explicitly.

There are two issues. First, the exemplary character of this definition is not to be underestimated. As the AI Act will serve as a basis for many other regulatory frameworks, it is likely that this definition will be adopted by those.

Second, the proposed definition risks to capture in the regulation also traditional software systems that process data and take decisions. These systems are already rigorously tested and covered by current legislation.

Finally, there is no definition of “data”. This means that an every-day interpretation of the word needs to be used.