The European Commission's guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)

Comment by the Advisory Committee on the Ethics of Artificial Intelligence of the Austrian Commission for UNESCO

21st March 2025

Author: Rania Wazir, Ph.D.

Advisory Committee on the Ethics of AI
Austrian Commission for UNESCO
Georg-Coch-Platz 2 | A-1010 Vienna
E: oeuk@unesco.at
T: +43 1 526 13 01 13
www.unesco.at

unesco
Austrian
Commission for UNESCO

The stated purpose of the AI Act is to

> improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation. [Article 1(1)]

It is with this lens, then, that we examine the EC Guidelines on the definition of an AI system.

The AI Act takes a risk-based approach; in fact, there are only two sections of relevance to AI systems in general:

- Article 4 on AI literacy, which affects both providers and deployers
- Chapter VI, measures in support of innovation

While Article 4 is a requirement that must be met, it leaves very much flexibility and comes without sanctions. The burden arising from this requirement is minimal, while the benefits are great. Chapter VI, as the name indicates, comes with measures in support, and hence, there is no reason for providers to seek exclusion from its range of application. In light of this, it is unclear why the Guidelines on the definition of an AI System seek to restrict the breadth of the AI system definition. In fact, all of the examples given in the Guidelines on the definition of an AI system, with the exception of (44), would fall into the category of low-risk systems, and hence not be subject to the requirements of the AI Act (with the exception of Article 4, as noted above). It appears that there is very little to gain from attempts to restrict the AI system definition scope.

However, the attempts to do so leave very much to lose. In order to understand the impact, it is also important to take the Guidelines for definition of an AI system together with the Guidelines on Prohibited AI Practices.

We break down Article 1 of the AI Act into its components:

i) improve the functioning of the internal market
ii) promote the uptake of human-centric and trustworthy artificial intelligence (AI)
iii) ensure a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union
iv) support innovation

For the sake of the logic of the arguments, we will handle them in the order iv), i), ii), iii).

support innovation

Innovation is not hampered by regulation; in fact, good regulation supports innovation and creates the framework of trust that enables healthy experimentation, widespread adoption and overall benefit. However, bad regulation – complex, unclear, and leaving much uncertainty – *is* an obstacle to innovation.

The AI Act follows a clear, risk-based logic. The breadth of the AI system definition allows the regulation to remain technology-neutral and instead focus on those *applications and use cases* that cause the highest risk to health, safety, and fundamental rights.

However, certain concepts within the AI Act are at too high-level for practical implementation. While some are to be operationalised by harmonised European standards (Article 40) or common specifications (Article 41), as foreseen in the New Legislative Framework, others are placed in the remit of the European Commission, with the publication of specific guidelines (Article 96).

One essential lacuna of the AI Act is the lack of a clear demarcation of the boundary of an AI system. In fact, many systems in operation today contain some AI components – but unfortunately, the AI Act does not establish whether the individual AI component is to be considered the AI system subject to AI Act, or the entire encompassing system is to be considered an AI system under the AI Act; or something in-between. Clarity on this aspect could go a long way towards reducing unnecessary burden, but unfortunately the EC Guidelines on the definition of an AI system do not address this issue.

Furthermore, the guidelines do not follow a clearly relatable logic in defining what is an AI system, and what is not. Some concerns are:

- The guidelines are internally contradictory – for example, while paragraph (39) includes mathematical optimisation among the list of AI methods, paragraph (41) seems to exclude it.
- Linear regression and logistic regression are not mathematical optimisation[1] methods, nor are they used to accelerate or approximate mathematical optimization methods (41). Excluding them on that basis does not make technical sense.
- Simple vs complex. The guidelines are trying to draw a distinction between "simple" methods, and "complex" ones. There are several major issues with this.
    - First, there seems to be a belief that a "simple" system cannot be harmful. This is wrong (see, for example, Verma and Rubin[2], as well as the FRA report bias in algorithms[3]): a system can be harmful to your health, safety, or Fundamental Rights, regardless of whether it is simple, or complex. **So why exclude simple?**
    - Second, consider a "simple" algorithm such as logistic regression. This quickly becomes "complex" if you add a lot of independent variables to the equation (who can explain what a logistic regression of 100+ variables is doing?). Also, a neural network is just a whole bunch of logistic regressions strung together (nowadays, many NNs are even simpler - rectified linear units). So, at what point have so many logistic regressions been strung together, that it becomes complex, and AI? The same applies for decision trees (simple) and random forests (complex).
    - Third, the exclusion for systems based on “classical heuristics” (paragraph 48). Most advanced machine learning methods are based on some form of “classical heuristics”.

[1] mathematical optimisation is never defined in the guidelines. As a mathematical concept, it is a set of methods for finding the maximum/minimum of a given function; this can be with, or without, constraints. Linear regression and logistic regression, on the other hand, are statistical methods that attempt to describe the relationship between a dependent variable and one or more independent variables.
[2] Verma and Rubin, “Fairness Definitions Explained”, 2018 ACM/IEEE International Workshop on Software Fairness, https://dl.acm.org/doi/10.1145/3194770.3194776
[3] FRA (2022), Bias in Algorithms – Artificial Intelligence and Discrimination, https://fra.europa.eu/en/publication/2022/bias-algorithm. See Chapter 3 for an example of bias in a logistic regression model; and Chapter 2 for an example of bias in a “simple” Naïve Bayes model

Reinforcement learning is based on “trial-and-error”. And the minimax algorithm mentioned in (48) is the basis of state-of-the-art methods, such as GANs.

- Finally, there is another belief that “simple” systems are less performant (paragraph 49). If they are less performant, and the provider has tested this, then they will not ”place them on the market” - and the simple systems will not fall under the AI Act, no need to make an exception for them here. If they are less performant, but the provider did not test this and places them on the market, then they are more likely to fail – posing greater risks to health, safety, and fundamental rights – so why should they be excluded from the AI Act? And finally, depending on the task, simpler systems might actually perform *as well as, or better* than complex systems. And in this case, the exclusion of simple systems *based on performance* should not apply.

It is difficult to generalise from the guidelines and the examples given, leaving providers and deployers deeply uncertain. Especially for SMEs, which cannot afford legal battles, the safest path forward is to be conservative and avoid novelty. This is the opposite of supporting innovation.

improve the functioning of the internal market

The Guidelines on prohibited practices propose prohibitions, but with several exceptions. The provider or deployer is, however, warned that even if the system is not prohibited under the AI Act, there are other EU laws (for example, GDPR, unfair commercial practices, or specific Member State laws on non-discrimination) which could prohibit them anyway. The operator is once again left facing a highly complex and fragmented market.

If part of the intention of the AI Act was to create harmonized rules, this is not accomplished. Consider further that, in combination with the Guidelines on the definition of an AI system, some practices might be prohibited under the AI Act if performed by a neural network but might *not* be prohibited if using a logistic regression, and the confusion is complete.

promote the uptake of human-centric and trustworthy artificial intelligence

In this context, it is important to note that the goal of the AI Act is not the uptake of AI in general, but the uptake of *human-centric and trustworthy* AI. By excluding potentially high-risk systems from the AI system definition, the importance of *human-centric and trustworthy* is demoted: some systems can be highly risky to health, safety, or fundamental rights, but that is OK, because they are not AI.

ensure a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the European Union

The current Guidelines on the definition of an AI system could be interpreted in such a way as to exclude many potentially harmful systems from the scope of the AI Act (see, for example, FN 2 and 3 on bias when using logistic regression).

A further example in the case of prohibited practices:

> A system to predict the chances of a natural person to commit a crime, based on historical data, and using logistic regression, would not fall under the prohibitions of Article 5. The same system, but using Deep Learning methods, would. However, the fundamental rights harms would be the same, in both cases.

Paragraph (44) provides a further example of a system that falls out of scope of the AI system definition:

> a satellite telecommunication system to optimize bandwidth allocation and resource management

This is an example that would otherwise fall under the critical infrastructure point of Annex III. To understand the implications of this exclusion, replace "satellite telecommunications" with "municipal electricity allocation":

> a municipal electricity allocation system to optimize bandwidth allocation and resource management

The failure of such a system could result in critical shutdowns, including, for example, of hospitals, leading to potential harms for health and safety. But such a system is not AI, and hence, does not fall under the AI Act obligations.

The Guidelines on prohibited AI practices also include some exemptions, which are troubling from the perspective of respect for fundamental rights.

This includes, for example:

- Article 5(1)(b) prohibits AI systems that exploit vulnerabilities based on "age, disability or a specific *social or economic* situation". In the guidelines, however, this becomes vulnerabilities based on "age, disability, or a specific *socio-economic* situation". The motivation for restricting the remit of the prohibition from the broader "social or economic situation" to the narrower "socio-economic situation" is unclear.

- Article 5(1)(c) prohibits "AI systems for the evaluation or classification of natural persons or groups of persons *over a certain period of time* based on their social behaviour or known, inferred or predicted personal or personality characteristics…". However, the Guidelines interpret this as "evaluation or classification [that] *is based on data* that spans over 'a certain period of time'". There is no motivation given for this interpretation, nor is it entirely internally consistent: for example, the prohibition explicitly mentions the use of personal characteristics, and these include features such as gender, nationality, height, or handedness – which are data that can be expected to remain stable for extended periods of time. So, a system that uses these features once to make a classification *would not* be prohibited, but a system that repeatedly, over the space of several months, uses these exact same features, to make the exact same classification, *would* be prohibited? A more rights-friendly interpretation of this prohibition would prohibit any system which evaluates or classifies natural persons, and either does this repeatedly over a period of time, or else does this once, but whose classification *remains valid over a certain period of time* (otherwise, the person could ask to be re-evaluated and reasonably expect the evaluation to change).

- Article 5(1)(f) prohibits "AI systems to infer emotions of a natural person in the areas of workplace and education institutions". However, the Guidelines change "AI systems to infer emotions" into "emotion recognition systems", thus *reducing the scope of prohibition* to only those AI systems that use *biometric data* to infer emotions. This leaves, for example, systems that infer emotion based on written text, out of scope. Since, however, the exact same problems with respect to scientific validity, and fundamental rights impacts, would apply, the reason for excluding such systems from the prohibition is highly problematic.

Conclusion

Rather than focusing on the underlying methods, the guidelines should adhere to the original AI Act strategy of remaining technology-neutral, and focusing on requirements at the level of risk associated with the application that an AI system is intended for.

Many issues highlighted here could be addressed by:

- providing clear and easily implemented guidelines on what constitutes an AI system under the AI Act. An example would be the ELI response on the definition of an AI system: https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Response_on_the_definition_of_an_AI_System.pdf
- providing guidelines establishing the boundary of an AI system under the AI Act
- providing guidelines on AI literacy, as well as providing substantive support for AI literacy initiatives, for example by subsidising AI literacy training, especially for SMEs. Such support would not only reduce the burden of compliance with Article 4 but also help improve the overall level of AI literacy, providing an additional benefit for the EU
- eliminating loopholes in the prohibited practices, creating a consistent framework for protecting health, safety, and fundamental rights

The EU has shown foresight, courage, and leadership in setting up regulations to define how we want to live with AI technology. It should persist in its role as guarantor of citizens' health, safety, and fundamental rights. This should be the field in which innovation happens, to the benefit of all.