2023-02

CEPS EXPLAINER

WHAT'S IN A NAME?

Getting the definition of Artificial Intelligence right in the EU's AI Act

Andrea Renda

Alex Engler

SUMMARY

One of the most significant pieces of EU legislation that will be finalised over the coming months is the AI Act, which is still a source of heated debate among policymakers and stakeholders. One of the key issues being discussed is how the Act should define Artificial Intelligence. In this short CEPS Explainer, Andrea Renda and Alex Engler argue in favour of a broader definition of AI, with a high degree of autonomy given to a dedicated AI Office to tailor the Act's application to the specificities of algorithms in individual sectors and use cases.

Andrea Renda is a Senior Research Fellow and Head of Global Governance, Regulation, Innovation and the Digital Economy (GRID) at CEPS and Alex Engler is a Fellow in Governance Studies at The Brookings Institution.

CEPS Explainers offer shorter, more bite-sized analyses of a wide range of key policy questions facing Europe. Unless otherwise indicated, the views expressed are attributable only to the authors in a personal capacity and not to any institution with which they are associated.

CEPS ▪ Place du Congrès 1 ▪ B-1000 Brussels ▪ Tel: (32.2) 229.39.11s ▪ Tel: (32.2) 229.39.11

Almost two years after the European Commission presented its proposal for an Artificial Intelligence (AI) Act, several crucial elements of the legislation are still subject to a heated debate. Among them is the foundational issue of how to define AI. The issue is not trivial – as an ever-evolving family of techniques with multi-faceted attributes and a plethora of use cases, AI escapes easy definition.

Most regulators have avoided narrow definitions, fearing these would be under-inclusive and become rapidly obsolete. A good example is the U.S. Blueprint for an AI Bill of Rights, which does not constrain its scope based on the type of algorithm used to make a determination, leaving further specifications in the hands of sectoral regulators. International organisations, such as the OECD, the Council of Europe and UNESCO, have provided rather broad definitions (the first two), or stayed away entirely from the exercise (the latter). Similarly, the UK and China refer to key characteristics of AI without providing an explicit definition.

> THE EU CANNOT AFFORD THE LUXURY OF NOT DEFINING AI. PROPOSING A REGULATORY FRAMEWORK WITHOUT PROVIDING A DEFINITION OF THE SUBJECT TO BE REGULATED WOULD BE LEGALLY INFEASIBLE.

The EU cannot afford the luxury of not defining AI. Proposing a regulatory framework without providing a definition of the subject to be regulated would be legally infeasible. At the same time, getting the definition wrong would be a disastrous outcome for a regulation that aims to protect fundamental rights and become a reference framework for future global rules on AI.

But what definition should be adopted in the AI Act?

The original proposal contained a broad definition, and the European Parliament is currently leaning towards a different, but also broad, definition used by the US's NIST. The Council, under the French and Czech presidencies, proposed a much narrower definition focused on a subset of AI techniques that, rather than following step-by-step rules, instead 'infer' the best way to complete a task.

Additionally, the Council's text refers to systems that are designed to operate with 'elements of autonomy', and this may further narrow down the scope of the definition to models capable of adaptive decision-making. This Council definition can be interpreted as more strictly confined to machine learning and similar AI approaches.

So, which is the better approach? This CEPS Explainer argues that a broad definition is key to the success of the AI Act but can be improved upon by allowing the EU's dedicated AI Office to tailor the Act's application to the specificities of algorithms in individual sectors and use cases.

IN PRAISE OF A BROAD DEFINITION OF AI

A broad definition of AI has several advantages. **First, it does not leave any AI technique out of scope**. If it were true that only machine learning algorithms, acting with elements of autonomy, created significant risks to fundamental rights and safety, then a narrow definition would be preferable. However, we see clear evidence that other algorithmic systems, such as rule-based approaches, create comparably meaningful risks.

No matter the maths, all algorithmic decisions are equally obscured without disclosure of the system. Furthermore, if no explanation is attempted, a neural network is equally opaque as linear regression to the affected person. Errors in an individual's data can lead to detrimental results in the most rudimentary automated process. Algorithm discrimination is also possible, with demonstrated examples of racial and gender bias in computer programmes dating back to the 1970s. While more complex algorithms might increase the frequency of these problems or complicate the paths to remediation, there is no case to be heard that simpler algorithms are harmless.

A second, important advantage of a broad definition is that **by being technologically neutral, it avoids creating the perverse incentive to strategically avoid regulatory requirements**.

Operating under a narrow definition, AI providers might end up choosing simpler approaches, to avoid being subject to the regulation. This outcome may end up constraining innovation by making certain approaches more attractive than others, irrespective of their accuracy and effectiveness. As machine learning enables better performance in many circumstances compared to rule-based systems, pushing developers to adopt the latter is not always ideal.

If the AI Act only covered machine learning, AI developers could translate machine learning models into step-by-step code to avoid the requirements. This is not as hard as one might think. Machine learning is fundamentally a process – once the resulting model is deployed, there is often no easy distinction between it and more traditional rules-based or formulaic algorithms.

Since many algorithmic methods defy easy categorisation, a narrow definition of AI would also create constant questions over inclusion. A broad definition therefore offers more **regulatory certainty and legal clarity**. A look at the AI value chain and ongoing market developments reveals that very often, several techniques are blended within the same AI system, with machine learning models and expert systems often co-existing. When this occurs, understanding whether the system, or parts of it, are subject to the AI Act may become an uphill battle. For example, AI systems providers that rely on both regulated and unregulated models would need to carry out *ex ante* and *ex post* periodic conformity assessments only for the part of the software that is covered by a narrow definition.

Fourth, a broad definition has the advantage of being **more future proof**. For example, should new approaches emerge in the future, which do not rely on machine learning or do not ‘infer’ how to complete a task, the broader definition would be more likely to capture them than the one proposed by the Council. In this case, the AI Act would have to be amended, and this could take years to achieve. On the contrary, a broad definition would only require changing Annex I (in the original text of the proposed Act) to incorporate new developments. This can be done through delegated acts, and is therefore a much quicker process.

Finally, one important advantage of a broad definition is that **it would place the AI Act on par with existing international definitions**, such as the one adopted by the OECD, or the one proposed by NIST in the United States. Aligning definitions and scopes with international forums is essential for the ‘Brussels effect’ to materialise, even if both of this Explainer’s authors have expressed doubts on the likelihood that this will happen.

A ‘THIRD WAY’ TO REACH AN AI DEFINITION

Perhaps the most frequent critique of the broad definition of AI is that it will inevitably be overly inclusive, and therefore be overly regulatory. Yet it is important to recall that the AI Act does not subject all systems defined as AI to regulatory requirements.

There are two important filters that significantly narrow down the group of AI applications subject to regulatory requirements. First, the definition is accompanied by an Annex I, which specifies the techniques that are considered to be AI. Second, the Act adopts a risk classification, which imposes regulatory requirements only on those AI systems (less than 10 % of the total according to the Commission) that create high risks for health, safety and fundamental rights. This risk classification includes AI in products already regulated in the EU, and a list of AI services (in Annex III), including applications related to critical infrastructure, education, health, employment, essential services, law enforcement, and more.

IN SHORT, FALLING UNDER THE DEFINITION OF AI DOES NOT MEAN BEING SUBJECT TO THE REGULATION, AND A BROAD DEFINITION DOES NOT IMPLY OVER-REGULATION

In short, falling under the definition of AI does not mean being subject to the regulation, and a broad definition does not imply over-regulation.

This, of course, does not mean that the Commission's approach is perfect. It could still prove flawed in terms of overinclusion and the suitability requirements. If any risk classification is too broad, trivial applications might be included. Furthermore, even well-defined risk classifications will lead to many instances of ambiguity. Courts and regulatory agencies could still clarify and refine the definition in applying the Act, but this would take some time and may also lead to diverging interpretations across Member States.

This challenge can be addressed by coupling the broad definition of AI with a subsequent, more tailored indication of the AI techniques that create higher risks in specific contexts. Accordingly, the definition of AI in the Act could consist of two components:

- A **broad, technology neutral, future-proof definition** of AI, possibly in line with the OECD definition.
- A **mandate for the future AI Office** to cooperate with sectoral regulators at the EU (and possibly national) level to define the specific techniques and applications that would qualify as high risk in specific contexts.

This approach would reconcile the value of a broad definition of AI with the need for more specific guidance within each high-risk classification. This option would necessitate that the AI Office acts (1) with the support of a group of AI and technology policy experts; (2) in cooperation with domain experts in the high-risk domain (e.g. health, educational access, border control etc.); and (3) in a transparent and accountable way, possibly in constant consultation with stakeholders.

If integrated into the broader text of the AI Act, this approach could also leverage the information gained from conformity assessments, offering valuable information to the EU's AI Office and sectoral regulators on the function and risks of these AI systems. As this is a significant responsibility, a better resourced AI Office is likely to be necessary, rather than an 'EU AI Board' that was originally proposed.

This approach to determining what should be included in the AI Act could not only address concerns of over-regulation and legal ambiguity but also enhance the effectiveness of the AI Act. Currently, once an AI application is designated as high risk, it is subject to a full menu of regulatory requirements, which are not tailored to that specific use case and the specific risks raised. This is likely to be a significant implementation challenge.

Early efforts to regulate AI in a specific circumstance — such as the US Equal Employment Opportunity Commission's approach to AI hiring and people with disabilities — shows how deeply nuanced this process can be. However, a network between sectoral regulators and independent experts, organised by the EU's AI Office, would be excellently positioned to help fine-tune these requirements, rather than leaving this entire process to the standards bodies CEN and CENELEC.

Admittedly, this approach would need to carefully consider how to ensure ongoing agreement and harmonisation between EU Member States, but the benefits of specificity in AI governance are worth that effort. The work done in this context would also be instrumental for international regulatory cooperation, starting with the Joint Roadmap for Trustworthy AI and the recent cooperation agreement on AI signed by the European Commission and the Biden administration, as well as extending to the ongoing work of the OECD Network of AI Experts and other relevant international forums.

In adopting this approach to formulate an AI definition, the EU could comprehensively tackle AI governance, while still allowing for the necessary flexibility within sectors. This two-stage approach maintains both the EU's principles and the AI Act's risk-based regulatory framework.

Crucially, it would also be an exemplar of AI governance on the global stage.

CEPS
Place du Congres 1
B-1000 Brussels