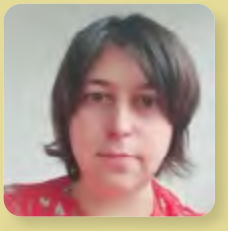
# THE EU AI ACT – BALANCING HUMAN RIGHTS AND INNOVATION THROUGH REGULATORY SANDBOXES AND STANDARDIZATION



**BY**
**KATERINA YORDANOVA**

Researcher, Centre for IT & IP Law, KU Leuven.

# TechREG CHRONICLE
# MARCH 2022



Visit www.competitionpolicyinternational.com for access to these articles and more!

## The EU AI Act – Balancing Human Rights and Innovation Through Regulatory Sandboxes and Standardization

By Katerina Yordanova

EU has invested a lot of efforts into creating a human-centric legislative framework for artificial intelligence, as part of its economy's digital and green transitions. This piece aims to shed light on the main features and the evolution of the proposal for the EU AI Act, as well as critically assess some shortcomings that still need to be addressed. It also concentrates on the new regulatory mechanisms adopted by the proposed regulation as an answer for the dynamic nature of technologies and their effect on society. By concentrating on the regulatory sandboxes and standardization the column aims to explore them in the context of the AI Act and critically evaluate the pros and cons of these tools for the ultimate purpose of balancing innovation and regulation in a manner that fully and effectively protect EU fundamental rights and public interest.

# 01

# BRIEF DESCRIPTION OF THE AI ACT AND ITS EVOLUTION

The EU's ambition to regulate artificial intelligence ("AI") systems has been clearly demonstrated in recent years. The first significant action in that direction was the establishment of the High-Level Expert Group on AI ("HLEG") in 2018 which paved the way for the President of the European Commission, Ursula von der Leyen, to declare the planned adoption of an AI legal instrument as a top priority in her policy agenda.[2] In February 2020, the Commission published a White Paper on AI, presenting different policy options which after public consultation and a number of critical contributions from different stakeholders resulted in the first draft of the Regulation Laying Down Harmonised Rules on Artificial Intelligence ("the AI Act"). The text proposed by the European Commission was discussed by the Council of the EU and the two parts of the Compromise Text were presented in November 2021 and January 2022, respectively, introducing some notable changes.

### A. Scope

The legal basis of the AI Act is Article 114 of the Treaty on Functioning of the European Union. This means that the AI Act pursues four specific objectives – ensuring that AI systems on the Union market are safe and respect fundamental rights and Union values, while safeguarding legal certainty, enhancing governance and effective enforcement of the existing legislation regarding AI systems, and facilitating the development of a single market for lawful, safe, and trustworthy AI and helping to avoid market fragmentation.

Following these four objectives, the rather bulky regulation establishes rules on "placing on the market, putting into service and the use of AI systems in the Union." It attempts to define and classify AI systems adopting a risk-based approach and subsequently regulates them along a spectrum, going as far as prohibiting certain AI practices.

The *ratione paersonae* of the Act is quite broad, encompassing "**providers** placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are physically present or established within the Union or in a third country," **users** of AI systems within the Union and "providers and users of AI systems who are physically present or established in a third country, where the output produced by the system is used in the Union." In addition, the Compromise Text of the Council of the EU amended the text of Article 2 by including as part of the personal scope of the regulation **importers and distributors** of AI systems, **product manufacturers** "placing on the market or putting into service an AI system together with their product and under their own name or trademark" and authorized representatives of providers which are established in the EU.

This extremely wide scope and broad extraterritorial effect resembles somewhat the approach adopted by the General Data Protection Regulation ("GDPR"), showing a prime example of the so-called "Brussels effect"[3] through which EU is striving to regulate global markets. It is evident by the provision of Article 2 of the AI Act in conjunction with recital 10.

To make matters even more complicated, the notion of a "provider" includes:

> [N]atural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed and places that system on the market or puts it into service under its own name or trademark, whether for payment or free of charge.

This definition is problematic in practice because its scope is so large it encompasses big tech companies such as Microsoft but at the same time individual FOSS developers. It is not clear if in such context uploading software to GitHub would constitute "placing it on the market" or "putting it into service" according to the regulation's terminology.

The material scope of the AI Act is limited, for example, by certain regimes that exist in other EU legal acts such as Regulation (EC) 300/2008 on common rules in the field of civil aviation security, or by AI systems developed or used exclusively for military purposes. This, however, encompasses a rather small number of cases, considering the broad scope of the definition of AI system provided by the Act.

The definition itself was a particular focus of criticism throughout the evolution of AI regulation. Article 3 (1) by the original definition proposed by the Commission identified an AI system as "software that is developed with one or more

2 In fact, President von der Leyen committed to a first attempt for regulation of AI during her first 100 days in office.

3 Anu Bradford, *The Brussels Effect* (Columbia Law School, Scholarship Archive, 2012).

of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with." The annex in question contained a rather confusing list of techniques the purpose of which was to make the regulation future-proof.

The Compromise Text of the Council entirely rewrote the definition and got rid of some problematic elements such as defining AI systems as software and as such being protected as copyrighted materials. In the new definition, AI systems are merely referred to as systems that **receive** machine and/or human-based data and inputs, **infer** "how to achieve a given set of human-defined objectives using learning, reasoning or modelling implemented with the techniques and approaches listed in Annex I" and **generate** "outputs in the form of content, predictions, recommendations or decisions, which influence the environments it interacts with." While the new definition seems a little bit clearer, it is also more restrictive, which has already attracted some criticism for leaving out certain types of AI, and also because Annex I, containing a rather large part of the definition, is subject to unilateral amendment by the Commission via delegated acts under Article 73 in conjunction with Article 4 of the AI Act. This approach in recent legislative instruments has been labeled as an attempt to adapt traditional legislation to the dynamic nature of the present times and the effect of disruptive technologies to society. Unfortunately, rather than coming close to the effect of the developing trend of anticipatory regulation[4] tools, it rather contributes to the democratic deficit vis-à-vis the EU and its legislative and regulatory activities.

Article 3 of the AI Act provides plethora of definition for the purpose of the regulation, some with questionable quality. A striking example is the attempted definition of emotion recognition system, as an "AI system for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data." From a legal point of view the work intention" is open to interpretation. Aside from pragmatic questions, such as when a thought becomes intention and how a system would determine this, the use of "intention" in legal acts usually denotes a form of *mens rea*. This is, however, considerably different from the context in which it is used here. Since an EU regulation is directly applicable in the legal systems of Member States this would raise significant problems.

Another problem which was created by the Council's version is the removal of the part "…which allow or confirm the unique identification of that natural person" from the definition of biometric data in Article 3(33). The initial definition was actually a copy of the definition provided by Article 4(14) of GDPR. The changes made by the council created a new scope of the term which is much broader in the AI Act compared to GDPR and thus would create serious problems with regard to the enforcement of both regulations. Unfortunately, similar inconsistency in the language could be found in many places across the AI Act which, together with the lengthy and unnecessary complicated sentences, turns the draft into a very bad example of legislative technique. If it remains unfixed, this would be a significant departure from the rule of law's fundamental principle that legal provisions should be clear and predictable, especially since it is not a problem limited to this particular regulation.

### B. The Risk-based Approach to AI

The AI Act adopts a dynamic risk-based approach for regulation of AI systems, creating different risk tiers depending on the degree of risk for public interest and EU fundamental rights, establishing risk mitigation mechanisms and a detailed governance system.

> ***The definition itself was a particular focus of criticism throughout the evolution of AI regulation***

#### 1. Prohibited AI Practices

The category of prohibited AI practices described in Article 5 provoked heated discussions. On one hand, industrial stakeholders were not happy regarding the existence of prohibited practices on the first place, on the other hand, civil society organizations insisted on a much broader scope than what was envisioned in Article 5, including full prohibition of remote biometric identification. In the Compromise text of the AI Act there were very few rather cosmetic changes in the wording of the article. It is evident that both the Commission and the Council believe that in some specific cases, the risk to human safety and fundamental rights is so great that no mitigation measures would be sufficient. Thus, it is prohibited placing on the market and putting into service of an AI system that for instance:

4 Geoff Mulgan, *Anticipatory Regulation: 10 Ways Governments Can Better Keep up with Fast-Changing Industries, Nesta (blog)* (May 15, 2017) https://www.nesta.org.uk/blog/anticipatory-regulation-10-ways-governments-can-better-keep-up-with-fast-changing-industries/.

> [D]eploys subliminal techniques beyond a person's consciousness with the objective to or the effect of materially distorting a person's behaviour in a manner that causes or is reasonably likely to cause that person or another person physical or psychological harm.

This is rather confusing because the phrase "materially distorting a person's behaviour" is not defined. In fact, this seems more like a spin-off of the "material distortion of the economic behaviour of consumers" criterion, which is well-known to consumer protection lawyers familiar with the Unfair Commercial Practices Directive. It seems, however, judging by the meaning implied in the AI Act, that its use here is broader, but it is not clear how broader precisely. It is indeed concerning to prohibit AI practices EU-wide based on criteria that are anything but clear.

Another interesting example of prohibited AI practices concerns the much-debated biometric identification. Indeed, this topic has been discussed for quite a while; there are serious lobbying efforts advocating a full ban of AI-based biometric identification. It is not surprising they were not happy with the currently proposed ban limited to "the use of 'real-time' biometric identification systems in publicly accessible spaces for the purpose of law enforcement."

*Another interesting example of prohibited AI practices concerns the much-debated biometric identification*

First of all, there are numerous exceptions related to necessity, e.g. for objectives like prevention of "specific, substantial, and imminent threat to the life or physical safety of natural persons of a terrorist attack." While these appear to be valid objectives in principle, the lack of a recognized uniform definition of what constitutes a terrorist attack in both international and European law, coupled with the often intensive *mens rea* requirements, makes it hard to envision how law enforcement authorities would benefit from this exception in a uniform and compliant way.

Secondly, the definition of publicly available space as "any physical place accessible to the public, regardless of whether certain conditions for access may apply" is very broad. When read in conjunction with recital 9, it becomes even less clear which spaces are publicly available. Thirdly, unlike the other two prohibited practices here what is forbidden is 'the use' as opposed to "placing on the market, putting into service or use." Thus, it seems like such "real time" biometric identification systems could be manufactured and installed as a matter of principle, so long as they are not "used" outside the scope of the exception.

**2. High-risk AI systems**

Article 6, defining high-risk AI systems, was completely rewritten in the Compromise Text. In essence the provision remained the same. The change was due to the critiques of the formulation and the language used. Therefore, the AI Act regards as high-risk AI systems those that are in themselves a product covered by the Union harmonization legislation listed in Annex II if they are required by the same pieces of legislation to undergo third-party conformity assessment. These systems are also regarded as high-risk if they are intended as a safety component of a product covered by the aforementioned list of legislation. As a separate sub-category, Article 6 refers to those listed in Annex III. Probably the most notable and discussed such category are AI systems intended to be used for the "real-time" and "post" biometric identification of natural persons. As already stated, a number of stakeholders, especially from civil society, have been advocating a total ban on the use of AI for biometric identification which is currently considered a prohibited AI practice only in the narrow case of real-time biometric identification in publicly accessible spaces and for the purpose of law enforcement, subject to a few exceptions. It is interesting to note that in both cases of Article 5 and Annex III the Council's version of the AI Act changed "remote biometric identification" with "biometric identification" which broadened the scope of both the prohibited and thee high-risk AI systems categories.

Other types of high-risk AI systems that are of particular importance to the business and the sector are those "intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity." This category was broadened by the inclusion of AI systems "intended to be used to control or as safety components of digital infrastructure" and AI systems intended to be "used to control emissions and pollution." Another similar type of high-risk AI systems is indicated to be those used in the context of employment, workers' management and access to self-employment which includes, for example, using AI systems for recruitment purposes or for making decisions regarding promotions or terminations. Both types could have a significant impact on human rights, varying from the right to life and health in the case of management and operation of critical infrastructure, to the right of equality and non-discrimination.

A third group of high-risk AI systems are those used for access to, and enjoyment of, essential private services and public services and benefits, such as AI systems being used

by public authorities to assess someone's eligibility for benefits, or AI systems used for determining access or assigning natural persons to educational and vocational training institutions and assessing natural persons in such institutions.

Finally, Annex III designates as high-risk AI systems those used by law enforcement for various purposes, such as detecting someone's emotional state in order to be used as a lie detector. This particular use of AI systems was also considered in relation to their exploitation for the purpose of migration, asylum and border control management. The final category of high-risk AI systems includes those intended to "be used by a judicial authority or on their behalf for interpreting facts or the law for applying the law to a concrete set of facts." It is worth noting that AI systems intended for purely "ancillary administrative activities," which do not affect administration of justice on the level of an individual case, do not fall into this category.

**3. Limited Risk AI systems**

Article 52 of the AI Act prescribes some special transparency requirements for AI systems that interact in a unique way with humans. This includes AI systems that interact with people, such as chatbots, emotion recognition systems, and systems that generate deep fakes. The transparency obligation aims to ensure that individuals are aware that they interact with a machine, that the system processes their emotions and/or that a certain content has been artificially generated. This is without prejudice to any additional requirements that stem from such AI being additionally classified as high-risk, even though these systems are not considered high-risk *per se,* but they could be if their purpose falls within the scope of Article 6.

**4. Minimal Risk and General Purpose AI systems**

For the remaining AI systems that do not qualify as prohibited, high-risk or requiring high degree of transparency, the Commission proposes a voluntary approach through self-regulatory means, such as codes of conduct. The aim here is apparently to achieve the highest possible level of protection of fundamental rights by representing this voluntary approach as a competitive advantage that would supposedly boost innovation.

This was also the goal of the Council introducing the general purpose AI systems in Article 52a. It was also an attempt of responding to the received criticism regarding the missing regulation of foundation models. Recital 70a defines general purpose AI system as one that "are able to perform generally applicable functions such as image/speech recognition, audio/video generation, pattern detection, question answering, translation, etc." These systems are put in general outside the scope of the AI Act unless its purpose makes it subject to it. Unfortunately, this provision could prove to be ineffective due to the fact that a foundational model does not have intended purpose *per se* and this could be manipulated for certain AI systems to avoid falling under the scope of the AI Act.

## 02
## RISK MITIGATION MECHANISM

The risk-based classification of AI systems in the AI Act is not static. This means that a given AI system could change in type during its life cycle and thus be subject to changing obligations for its providers, users, etc.

High-risk AI systems naturally involve the broadest range of obligations and a good amount of additional costs. To simplify the process, for a high-risk AI system to enter the market it needs to first, be designed and developed following an internal impact assessment by multidisciplinary team. Second, it must undertake a conformity assessment[5] and comply with the requirements set in Chapter II of the AI Act. These requirements vary from establishment of risk management and data governance systems to transparency, human oversight, accuracy, robustness, and cybersecurity. Third, stand-alone AI systems are to be registered in a centralized EU database. Finally, a declaration of conformity must be signed, and the system must bear a CE marking before finally being placed on the market. It is important to note that if the system goes through substantial changes the process must be repeated from step two.

Naturally, this process is regarded to be a huge burden by business, and it could be potentially fatal for certain small and medium enterprises ("SMEs"), which are the backbone of European industry. At the same time, most stakeholders are adamant about keeping fundamental rights at the heart of EU legislation. This is also a unique competitive advan-

5 Certain types of high-risk AI systems must undergo a conformity assessment with the participation of a notified body according to Article 43 of the AI Act.

tage for AI made in Europe.[6] In order to balance fundamental rights protection and innovation the Commission bet on two rather different tools which have one thing in common – they increase predictability for business and have the potential to protect fundamental rights.

### A. Regulatory Sandboxes for AI

It was already mentioned that the AI Act empowers to Commission to use delegated acts quite frequently. While this approach is rightly criticized due to its undemocratic nature, it is also a reaction to the need for more agile ways to effectively regulate dynamic and everchanging fields such as disruptive technologies, including AI.

The term "regulatory sandbox" originates in computer science and was just recently adopted firstly in the area of financial regulation, in particular regarding FinTech.[7] The sandboxes' success allowed their quick adoption in other spheres such as data protection and healthcare. Granted there is no universal definition of the term, the European Securities and Markets Authority ("ESMA") regards regulatory sandboxes as "schemes to enable firms to test, pursuant to a specific testing plan agreed and monitored by a dedicated function of the competent authority, innovative financial products, financial services or business models."[8] This first definition differs from the one provided by the Council of the EU in 2020 where they are described as frameworks. The AI Act adopts a third one in Article 53(1) for specific regulatory sandboxes for AI which are:

> [E]stablished by one or more Member States competent authorities or the European Data Protection Supervisor shall provide a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan. This shall take place under the direct supervision and guidance by the competent authorities with a view to ensuring compliance with the requirements of this Regulation and, where relevant, other Union and Member States legislation supervised within the sandbox.

This specific definition provides some additional and novel elements. First, it explicitly emphasizes the possibility of multi-jurisdictional regulatory sandboxes. The feasibility of this type of sandboxes had been questioned before we even started talking about specific AI sandboxes. It was argued that "the fact that the service lacks the standardization associated with regulation makes the sandboxed activity unfit for cross-border provision of services."[9] It is yet to be found out how this barrier could be overcome.

Furthermore, the scope of the regulatory sandboxes for AI is significantly broadened, encompassing development, testing and validation and therefore combining the traditional function of a regulatory sandbox with those of other tools such as testing and pilots. It is important to note that there is an existing debate on the exact relation between the terminology used to describe these defined safe spaces for testing innovation with or without certain authorities being involved. What is agreed on is that "there is an inherent connection between a regulatory sandbox on the one side, and testing and piloting on the other"[10] and also that usually jurisdictions "with a sandbox approach put certain piloting and testing activities inside the sandbox since this is more convenient."[11] This probably contributes to the spawning of numerous other terms, for example living labs, regulatory testbeds, etc., which are used as synonyms and ultimately addressing areas in which to trial innovation and regulation. Nevertheless, the definition in the draft AI Act seems to incorporate certain testing and piloting elements[12] in addition to the regular sandbox activities, which could be a beneficial element only if it really facilitates the development of

6 Press Release, European Commission, Member States and Commission to work together to boost artificial intelligence "made in Europe" (December 7, 2018).

7 Currently there is not a completely unified definition of FinTech but here we would define it as a new technology aiming to automate and improve financial products and services.

8 ESMA, *Joint Report on Regulatory Sandboxes and Innovation Hubs* (2019).

9 Dirk Zetzsche et al. *Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation,* Fordham Journal of Corporate and Financial Law 23: 31–104 (2017).

10 *Id.*

11 *Id.*

12 The difference between tests and pilots is regarded as tests being a one-time event the outcome of which determines the subsequent development of a product/service/business model, while a pilot is a final test which aims to ensure some missing data before the product/service/business model is finally released to the market.

innovation and ultimately reduces the time to market which has been the primary goal of the tool to begin with.

**B. Standardization**

The other agile method of regulation envisioned by the AI Act is standardization. Recital 61 provides that "[s]tandardization should play a key role to provide technical solutions to providers to ensure compliance with this Regulation." The biggest standard organizations are already working on standards for AI systems (such as IEEE, ISO, ITU, etc.) including on EU level (CEN and CENELEC). Much like with the regulatory sandboxes, standards are seen as a prime tool for promoting "the rapid transfer of technologies from research to application and open international markets for companies and their innovations."[13] Unlike the sandboxes though, standards do not have the scale problem. One of the main issues, however, remains the way human rights protection can actually be implemented in a standard. A prime example is ISO 26000, which provides guidance on social responsibility. It is considered fairly ineffective due to multiple reasons such as sloppy language, price, complexity, the limited scope of social responsibility, etc. This raises some concerns regarding the feasibility of incorporating human rights protection in standards and how effective this could be.

Standards, on the other hand, balance innovation and human rights by contributing to foreseeability and creation of trust. Clear rules increase innovation but there are a number of concerns that need to be taken into consideration. Private standards development organizations are often opaque, and it is unclear if their governance mechanisms and procedural rules follow the procedural principles for standardization such as transparency, openness, impartiality, and balance, etc. Furthermore, incorporating human rights categories in standards is a complicated task, and we are still lacking good know-how on the matter. In conclusion, both regulatory sandboxes and standards, utilized for the purpose of protection of public interest and fundamental rights in the scope of the AI Act have their merits but there is a steep learning curve, and ultimately the one-size-fits-all approach needs to be avoided. Instead, the AI Act should rely on an even broader set of anticipatory regulation tools which would allow a tailor-made response to the challenges presented by the most disruptive technologies up to date.

*The AI Act is still a work in progress. Balancing adequate and comprehensive human rights protection with innovation is not an easy job*

# 03 CONCLUSION

The AI Act is still a work in progress. Balancing adequate and comprehensive human rights protection with innovation is not an easy job. So far, the regulation offers some valuable mechanisms but there is a lot of work to be done regarding its consistency and effectiveness. Recognizing the need for better, more agile tools for regulating technologies is a positive step but it is yet to be determined which ones would work best in the EU context and weather they can really promote innovation. Regulatory sandboxes generated a lot of hype, but their effect is limited due to the small scale of tested products/services/business models. Furthermore, the strong human rights guarantees built into the process hinder their experimental nature and decrease their attractiveness which is primarily based on the lifting of certain legal restrictions during the participation in the sandbox.

13 DIN/DKE, German Standardization Roadmap on Artificial Intelligence, p.4 (November 2020).

# CPI SUBSCRIPTIONS

CPI reaches more than **35,000 readers** in over **150 countries** every day. Our online library houses over **23,000 papers**, articles and interviews.

Visit **competitionpolicyinternational.com** today to see our available plans and join CPI's global community of antitrust experts.

CPI COMPETITION POLICY™ INTERNATIONAL