EU Artificial Intelligence Act: The European Approach to AI

Mauritz Kop[1]

Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021. https://law.stanford.edu/publications/eu-artificial-intelligence-act-the-european-approach-to-ai/

Abstract

On 21 April 2021, the European Commission presented the Artificial Intelligence Act. This Stanford Law School contribution lists the main points of the proposed regulatory framework for AI.

The draft regulation seeks to codify the high standards of the EU trustworthy AI paradigm. It sets out core horizontal rules for the development, trade and use of AI-driven products, services and systems within the territory of the EU, that apply to all industries.

The EU AI Act introduces a sophisticated 'product safety regime' constructed around a set of 4 risk categories. It imposes requirements for market entrance and certification of High-Risk AI Systems through a mandatory CE-marking procedure. This pre-market conformity regime also applies to machine learning training, testing and validation datasets.

The AI Act draft combines a risk-based approach based on the pyramid of criticality, with a modern, layered enforcement mechanism. This means that as risk increases, stricter rules apply. Applications with an unacceptable risk are banned. Fines for violation of the rules can be up to 6% of global turnover for companies.

The EC aims to prevent the rules from stifling innovation and hindering the creation of a flourishing AI ecosystem in Europe, by introducing legal sandboxes that afford breathing room to AI developers.

The new European rules will forever change the way AI is formed. Pursuing trustworthy AI by design seems like a sensible strategy, wherever you are in the world.

***

On 21 April 2021, the European Commission presented the Artificial Intelligence Act. As a Fellow at Stanford University's Transatlantic Technology Law Forum and a Member of the European AI Alliance, I made independent strategic recommendations to the European Commission. President Ursula von der Leyen's team adopted some of the suggestions that I offered them, or has itself arrived to the same conclusions. That is encouraging. This contribution will list the main points of this novel regulatory framework for AI.

[1] Mauritz Kop is Stanford Law School TTLF Fellow at Stanford University and is Managing Partner at AIRecht, Amsterdam, The Netherlands.

Core horizontal rules for AI

The EU AI Act sets out horizontal rules for the development, commodification and use of AI-driven products, services and systems within the territory of the EU. The draft regulation provides core artificial intelligence rules that apply to all industries. The EU AI Act introduces a sophisticated 'product safety framework' constructed around a set of 4 risk categories. It imposes requirements for market entrance and certification of High-Risk AI Systems through a mandatory CE-marking procedure. To ensure equitable outcomes, this pre-market conformity regime also applies to machine learning training, testing and validation datasets. The Act seeks to codify the high standards of the EU trustworthy AI paradigm, which requires AI to be legally, ethically and technically robust, while respecting democratic values, human rights and the rule of law.

Objectives of the EU Artificial Intelligence Act

The proposed regulatory framework on Artificial Intelligence has the following objectives:

1. ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values;

2. ensure legal certainty to facilitate investment and innovation in AI;

3. enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems;

4. facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.

Subject Matter of the EU AI Act

The scope of the AI Act is largely determined by the subject matter to which the rules apply. In that regard, Article 1 states that:

Article 1
Subject matter

This Regulation lays down:

(a) harmonised rules for the placing on the market, the putting into service and the use of artificial intelligence systems ('AI systems') in the Union;

(a) prohibitions of certain artificial intelligence practices;

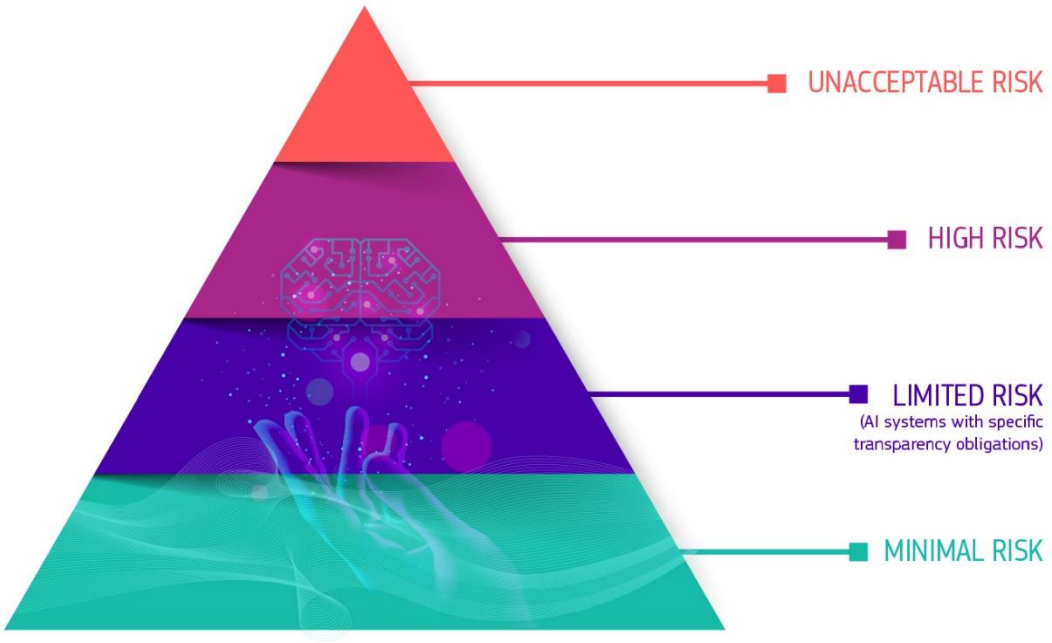
(b) specific requirements for high-risk AI systems and obligations for operators of such systems;

(c) harmonised transparency rules for AI systems intended to interact with natural persons, emotion recognition systems and biometric categorisation systems, and AI systems used to generate or manipulate image, audio or video content;

(d) rules on market monitoring and surveillance.

Pyramid of Criticality: Risk based approach

To achieve the goals outlined, the Artificial Intelligence Act draft combines a risk-based approach based on the pyramid of criticality, with a modern, layered enforcement mechanism. This means, among other things, that a lighter legal regime applies to AI applications with a negligible risk, and that applications with an unacceptable risk are banned. Between these extremes of the spectrum, stricter regulations apply as risk increases. These range from non-binding self-regulatory soft law impact assessments accompanied by codes of conduct, to heavy, externally audited compliance requirements throughout the life cycle of the application.



The Pyramid of Criticality for AI Systems

Unacceptable Risk AI systems

Unacceptable Risk AI systems can be divided into 4 categories: two of these concern cognitive behavioral manipulation of persons or specific vulnerable groups. The other 2 prohibited categories are social scoring and real-time and remote biometric identification systems. There are, however, exceptions to the main rule for each category. The criterion for qualification as an Unacceptable Risk AI system is the harm requirement.

Examples of High-Risk AI-Systems

Hi-Risk AI-systems will be carefully assessed before being put on the market and throughout their lifecycle. Some examples include:

- *Critical infrastructures (e.g. transport), that could put the life and health of citizens at risk*

- *Educational or vocational training, that may determine the access to education and professional course of someone's life (e.g. scoring of exams)*
- *Safety components of products (e.g. AI application in robot-assisted surgery)*
- *Employment, workers management and access to self-employment (e.g. CV sorting software for recruitment procedures)*
- *Essential private and public services (e.g. credit scoring denying citizens opportunity to obtain a loan)*
- *Law enforcement that may interfere with people's fundamental rights (e.g. evaluation of the reliability of evidence)*
- *Migration, asylum and border control management (e.g. verification of authenticity of travel documents)*
- *Administration of justice and democratic processes (e.g. applying the law to a concrete set of facts)*
- *Surveillance systems (e.g. biometric monitoring for law enforcement, facial recognition systems)*

Market Entrance of High-Risk AI-Systems: 4 Steps

In a nutshell, these 4 steps should be followed prior to Hi-Risk AI-Systems market entrance. Note that these steps apply to components of such AI systems as well.

1. A High-Risk AI system is developed, preferably using internal ex ante AI Impact Assessments and Codes of Conduct overseen by inclusive, multidisciplinary teams.

2. The High-Risk AI system must undergo an approved conformity assessment and continuously comply with AI requirements as set forth in the EU AI Act, during its lifecycle. For certain systems an external notified body will be involved in the conformity assessment audit. This dynamic process ensures benchmarking, monitoring and validation. Moreover, in case of changes to the High-Risk AI system, step 2 has to be repeated.

3. Registration of the stand-alone Hi-Risk AI system will take place in a dedicated EU database.

4. A declaration of conformity must be signed and the Hi-Risk AI system must carry the CE marking (Conformité Européenne). Now the system is ready to enter the European markets.

But this is not the end of the story...

In the vision of the EC, after the Hi-Risk AI system haven obtained market approval, authorities on both Union and Member State level '*will be responsible for market surveillance, end users ensure monitoring and human oversight, while providers have a post-market monitoring system in place. Providers and users will also report serious incidents and malfunctioning.*'[2] In other words, continuous upstream and downstream monitoring.

[2] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_en

Since people have the right to know if and when they are interacting with a machine's algorithm instead of a human being, the AI Act introduces specific transparency obligations for both users and providers of AI system, such as bot disclosure. Likewise, specific transparency obligations apply to automated emotion recognition systems, biometric categorization and deepfake/synthetics disclosure. Limited Risk AI Systems such as chatbots necessitate specific transparency obligations as well. The only category exempt from these transparency obligations can be found at the bottom of the pyramid of criticality: the Minimal Risk AI Systems.

In addition, natural persons should be able to oversee the Hi-Risk AI-System. This is termed the human oversight requirement.

Open Norms

The definition of high-risk AI applications is not yet set in stone. Article 6 does provide classification rules. Presumably, the qualification remains a somewhat open standard within the regulation, subject to changing societal views, and to be interpreted by the courts, ultimately by the EU Court of Justice. A standard that is open in terms of content and that needs to be fleshed out in more detail under different circumstances, for example using a catalog of viewpoints. Open standards entail the risk of differences of opinion about their interpretation. If the legislator does not offer sufficient guidance, the courts will ultimately have to make a decision about the interpretation of a standard. This can be seen as a less desirable side of regulating with open standards. A clear risk taxonomy will contribute to legal certainty and offer stakeholders with appropriate answers to questions about liability and insurance.

Enforcement

The draft regulation provides for the installation of a new enforcement body at Union level: the European Artificial Intelligence Board (EAIB). At Member State level, the EAIB will be flanked by national supervisors, similar to the GDPR's oversight mechanism. Fines for violation of the rules can be up to 6% of global turnover, or 30 million euros for private entities.

'*The proposed rules will be enforced through a governance system at Member States level, building on already existing structures, and a cooperation mechanism at Union level with the establishment of a European Artificial Intelligence Board.'*[3]

CE-marking: pre-market conformity requirements

In line with my recommendations, Article 49 of the Artificial Intelligence Act requires high-risk AI and data-driven systems, products and services to comply with EU benchmarks, including safety and compliance assessments. This is crucial because it requires products and services to meet the high technical, legal and ethical standards that reflect the core values of trustworthy AI. Only then will they receive a CE marking that allows them to enter the European markets. This pre-market conformity mechanism works in the same way as the existing CE marking as a safety certification for products placed on the European markets.

Please note that this pre-market conformity regime also applies to machine learning training, testing and validation datasets on the basis of article 10. These corpora need to be representative (I would

[3] ibid

almost say: inclusive), hi- quality, adequately labelled and error-free to ensure non-discriminatory and non-biased outcomes. Thus, the input data must abide to the high standards of trustworthy AI as well.

Pursuant to Article 40, harmonized standards for high-risk AI systems are published in the Official Journal of the European Union:

Article 40
Harmonised standards

High-risk AI systems which are in conformity with harmonised standards or parts thereof the references of which have been published in the Official Journal of the European Union shall be presumed to be in conformity with the requirements set out in Chapter 2 of this Title, to the extent those standards cover those requirements.

The CE marking for the individual types of high-risk AI systems can be applied for via a procedure as described in article 43.

Article 43
Conformity assessment

1. For high-risk AI systems listed in point 1 of Annex III, where, in demonstrating the compliance of a high-risk AI system with the requirements set out in Chapter 2 of this Title, the provider has applied harmonised standards referred to in Article 40, or, where applicable, common specifications referred to in Article 41, the provider shall follow one of the following procedures:

(a)the conformity assessment procedure based on internal control referred to in Annex VI;

(b)the conformity assessment procedure based on assessment of the quality management system and assessment of the technical documentation, with the involvement of a notified body, referred to in Annex VII.

Where, in demonstrating the compliance of a high-risk AI system with the requirements set out in Chapter 2 of this Title, the provider has not applied or has applied only in part harmonised standards referred to in Article 40, or where such harmonised standards do not exist and common specifications referred to in Article 41 are not available, the provider shall follow the conformity assessment procedure set out in Annex VII.

For the purpose of the conformity assessment procedure referred to in Annex VII, the provider may choose any of the notified bodies. However, when the system is intended to be put into service by law enforcement, immigration or asylum authorities as well as EU institutions, bodies or agencies, the market surveillance authority referred to in Article 63(5) or (6), as applicable, shall act as a notified body.

...

Article 43 paragraph 6 aims to prevent or avoid risks with regard to health, safety and fundamental rights:

6. The Commission is empowered to adopt delegated acts to amend paragraphs 1 and 2 in order to subject high-risk AI systems referred to in points 2 to 8 of Annex III to the conformity assessment procedure referred to in Annex VII or parts thereof. The Commission shall adopt such delegated acts taking into account the effectiveness of the conformity assessment procedure based on internal control referred to in Annex VI in preventing or minimizing the risks to health and safety and protection of fundamental rights posed by such systems as well as the availability of adequate capacities and resources among notified bodies.

Article 48 paragraph 1, EU declaration of conformity indicates that:

Article 48
EU declaration of conformity

1.The provider shall draw up a written EU declaration of conformity for each AI system and keep it at the disposal of the national competent authorities for 10 years after the AI system has been placed on the market or put into service. The EU declaration of conformity shall identify the AI system for which it has been drawn up. A copy of the EU declaration of conformity shall be given to the relevant national competent authorities upon request.

...

Further, Article 49 CE marking of conformity determines that:

Article 49
CE marking of conformity

1.The CE marking shall be affixed visibly, legibly and indelibly for high-risk AI systems. Where that is not possible or not warranted on account of the nature of the high-risk AI system, it shall be affixed to the packaging or to the accompanying documentation, as appropriate.

2.The CE marking referred to in paragraph 1 of this Article shall be subject to the general principles set out in Article 30 of Regulation (EC) No 765/2008.

3.Where applicable, the CE marking shall be followed by the identification number of the notified body responsible for the conformity assessment procedures set out in Article 43. The identification number shall also be indicated in any promotional material which mentions that the high-risk AI system fulfils the requirements for CE marking.

Finally, Article 30 of the draft regulation on notifying authorities provides that:

Article 30
Notifying authorities

1.Each Member State shall designate or establish a notifying authority responsible for setting up and carrying out the necessary procedures for the assessment, designation and notification of conformity assessment bodies and for their monitoring.

2.Member States may designate a national accreditation body referred to in Regulation (EC) No 765/2008 as a notifying authority.

3.Notifying authorities shall be established, organised and operated in such a way that no conflict of interest arises with conformity assessment bodies and the objectivity and impartiality of their activities are safeguarded.

Self-assessment too non-committal (non-binding)?

First, it is crucial that certification bodies and notified bodies are independent and that no conflicts of interest arise due to a financial or political interest. In this regard, I wrote elsewhere that the EU should be inspired by the modus operandi of the US FDA.

Second, the extent to which companies can achieve compliance with this new AI 'product safety regime' through risk-based self-assessment and self-certification, without third party notified bodies, determines the effect of the Regulation on business practices and thus on the preservation and reinforcement of our values. Internally audited self-assessment is too non-committal given the high risks involved. Therefore, I think it is important that the final version of the EU AI Act subjects all high-risk systems to external, independent third party assessments requirements. Self-regulation in combination with awareness of the risks via (voluntary or mandatory) internal ai impact assessments is not enough to protect our societal values, since companies have completely different incentives for promoting social good and pursuing social welfare, than the state. We need mandatory third party audits for all High-Risk AI Systems.

In this regard, it is interesting to compare the American way of regulating AI with the European approach. In America people tend to advocate free market thinking and a laissez faire approach. For example, the Stanford University, Silicon Valley group The Adaptive Agents Group recently proposed The Shibboleth Rule for Artificial Agents. Their proposal is reminiscent of the EU Human oversight requirement, and maintains that:

'Any artificial agent that functions autonomously should be required to produce, on demand, an AI shibboleth: a cryptographic token that unambiguously identifies it as an artificial agent, encodes a product identifier and, where the agent can learn and adapt to its environment, an ownership and training history fingerprint.'[4]

Their modest proposition contrasts strongly with the widely scoped European legal-ethical framework. However, history has already taught us dramatically that the power and social impact of AI is too great to be left largely to the companies themselves.

In addition, it is key that international standard setting bodies like ISO and IEEE adopt and translate the norms and values of the EU Act in their own technical standards, so that they are in line with each other. Such harmonized standards will encourage sustainable innovation and responsible business practices. In other words, worldwide adoption of such technical standards increases the chance that leading firms will adjust their behavior vis-a-vis AI.

Moreover, a harmonized global framework prevents forum shopping. With forum shopping I mean finding the most favorable possible regime to achieve one's own rights, motivated by financial interests that are often at the expense of consumers, competition, the environment and society.

[4] https://hai.stanford.edu/news/shibboleth-rule-artificial-agents

Innovation Friendly Flexibilities: Legal Sandboxes

In line with my recommendations, the draft aims to prevent the rules from stifling innovation and hindering the creation of a flourishing AI ecosystem in Europe. This is ensured by introducing various flexibilities and exceptions, including the application of legal sandboxes that afford breathing room to research institutions and SME's. Thus, to guarantee room for innovation, the draft establishes AI regulatory sandboxes. Further, an IP Action Plan has been drawn up to modernize technology related intellectual property laws.

'Additional measures are also proposed to support innovation, in particular through AI regulatory sandboxes and other measures to reduce the regulatory burden and to support Small and Medium-Sized Enterprises ('SMEs') and start-ups.'[5]

The concept thus seeks to balance divergent interests, including democratic, economic and social values. That irrevocably means that trade-offs will be made. It is to be hoped that during its journey through the European Parliament, the proposal will not be relegated to an unworkable compromise, as happened recently with the Copyright Directive, under the influence of the lobbying power of a motley crew of stakeholders.

Sustainability

Moreover, the explanatory memorandum pays attention to the environment and sustainability, in the sense that the ecological footprint of technologies should be kept as small as possible and that the application of artificial intelligence should support socially and environmentally beneficial outcomes. This is in line with article 37 of the EU Charter of Fundamental Rights ('the Charter'), and the EU Green Deal, which strives for the decarbonization of our society.

Sector specific rules

On top of the new AI rules, AI infused systems, products and services must comply with sector-specific regulations such as the Machinery Directive and the Regulations for medical devices (MDR) and in vitro diagnostics (IVDR), as well. Furthermore, besides the General Data Protection Regulation (GDPR) for personal data, the FFD Regulation for non-personal data and both GDPR and FFD for mixed datasets, the upcoming Data Act will apply. This applies, among other things, to B2B and B2G data sharing (depending on the types of data used), the use of privacy-preserving synthetic dataset generation techniques, and the use of machine learning training and validation data sets. In addition, audits of products and services equipped with AI must fit into existing quality management systems of industries and economic sectors such as logistics, energy and healthcare.

Regulations versus Directives

In the EU, regulations result in unification, in unification of legal rules. Member States have no discretion here for their own interpretation of the Brussels regulations. Member States do have that room for directives. Directives on the other hand, lead to harmonization of legal rules. Regulations such as the new Artificial Intelligence Act are directly applicable in the national legal orders of the member states, without the need for transposition or implementation. As was necessary, for

[5] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_en

example, with the recent Copyright Directive. As soon as the European Parliament and the Council of Europe agree with the final text in mid-2022 and if it is adopted, the AI Regulation will be immediately applicable law in all countries of the European Union.

AI Governance: trans-Atlantic perspectives

It is understandable that the European Union considers AI to be part of European strategic autonomy. Moreover, a degree of strategic European digital sovereignty is needed to safeguard European culture. Nevertheless, it is of existential importance for the EU to work together in concert with countries that share our European digital DNA, based on common respect for the rule of law, human rights and democratic values. Against this background, it is essential to stimulate systematic, multilateral transatlantic cooperation and jointly promote and achieve inclusive, participatory digitalization. The transatlantic and geopolitical dialogue on transformative technology, together with the development of globally accepted technology standards and protocols for interoperability, should be strengthened.

Setting Global Standards for AI

It takes courage and creativity to legislate through this stormy, interdisciplinary matter, forcing US and Chinese companies to conform to values-based EU standards before their AI products and services can access the European market with its 450 million consumers. Consequentially, the proposal has extraterritorial effect.

By drafting the Artificial Intelligence Act and embedding our norms and values into the architecture and infrastructure of our technology, the EU provides direction and leads the world towards a meaningful destination. As the Commission did before with the GDPR, which has now become the international blueprint for privacy, data protection and data sovereignty.

Methods also useful for other emerging technologies

While enforcing the proposed rules will be a whole new adventure, the novel legal-ethical framework for AI enriches the way of thinking about regulating the Fourth Industrial Revolution (4IR). This means that - if proven to be useful and successful - we can also use methods from this legal-ethical cadre for the regulation of 4IR technologies such as quantum technology, 3D printing, synthetic biology, virtual reality, augmented reality and nuclear fusion. It should be noted that each of these technologies requires a differentiated horizontal-vertical legislative approach in terms of innovation incentives and risks.

Trustworthy AI by Design

Responsible, Trustworthy AI requires awareness from all parties involved, from the first line of code. The way in which we design our technology is shaping the future of our society. In this vision democratic values and fundamental rights play a key role. Indispensable tools to facilitate this awareness process are AI impact and conformity assessments, best practices, technology roadmaps and codes of conduct. These tools are executed by inclusive, multidisciplinary teams, that use them to monitor, validate and benchmark AI systems. It will all come down to *ex ante* and life-cycle auditing.

The new European rules will forever change the way AI is formed. Pursuing trustworthy AI by design seems like a sensible strategy, wherever you are in the world.

**** fin ****