The EU's New Rules on Artificial Intelligence: A Risky Endeavour

Dimitar Lilkov

14 June 2021

With its April 2021 proposal for a European Artificial Intelligence Act, the European Commission makes a clear step beyond general ethical frameworks by proposing binding rules and mechanisms for the placement on the European market of Artificial Intelligence systems. The Act should be regarded as a balanced attempt to welcome new technologies while also embedding fundamental European values in a rapidly expanding and increasingly complex digital landscape.

In its striking exploration of Artificial Intelligence (AI), Nick Bostrom's book *Superintelligence* [1] addresses the potential impact of highly advanced AI systems on our society and weighs some of the most pressing potential risks and threats posed by sentient machines. The focus remains the *control problem* – how to create intelligent agents that are beneficial to their creators, while making sure they do not cause societal harm. This same focus appears to have guided reflections leading to the European Commission's (EC) proposed Artificial Intelligence Act.[2] The Act is built around the need to address the potential risks and harms that could be generated by specific uses of AI. It is important to note that the Commission does not aim to regulate the technology itself. Rather, the goal is to create future-proof harmonised rules on the use of AI that would ensure trust among users, increase uptake, and provide legal certainty for public bodies and private companies.

Background and pyramid of risk

In the last few years, the debate on AI governance was mostly centered on ethical frameworks. With its April 2021 legislative proposal, the EC makes a clear step beyond such frameworks by introducing binding rules for the placement on the EU market of AI systems. The Act is a proposed Regulation in the sense of European Law, directly applicable to both public and private sector actors across the Union. It addresses providers and users of AI systems in the Union, irrespective of whether they are established in the Union or in a third country. Given the size and importance of the European internal market, EU policy makers hope to create a policy spillover effect, nudging other countries to develop similar norms. It is noteworthy that the horizontal rules proposed by the Commission were not developed in a vacuum: there is an upcoming update on EU rules to address liability[3] issues related to new technologies, as well as a revision of sectoral safety legislation (e.g. Machinery Regulation).

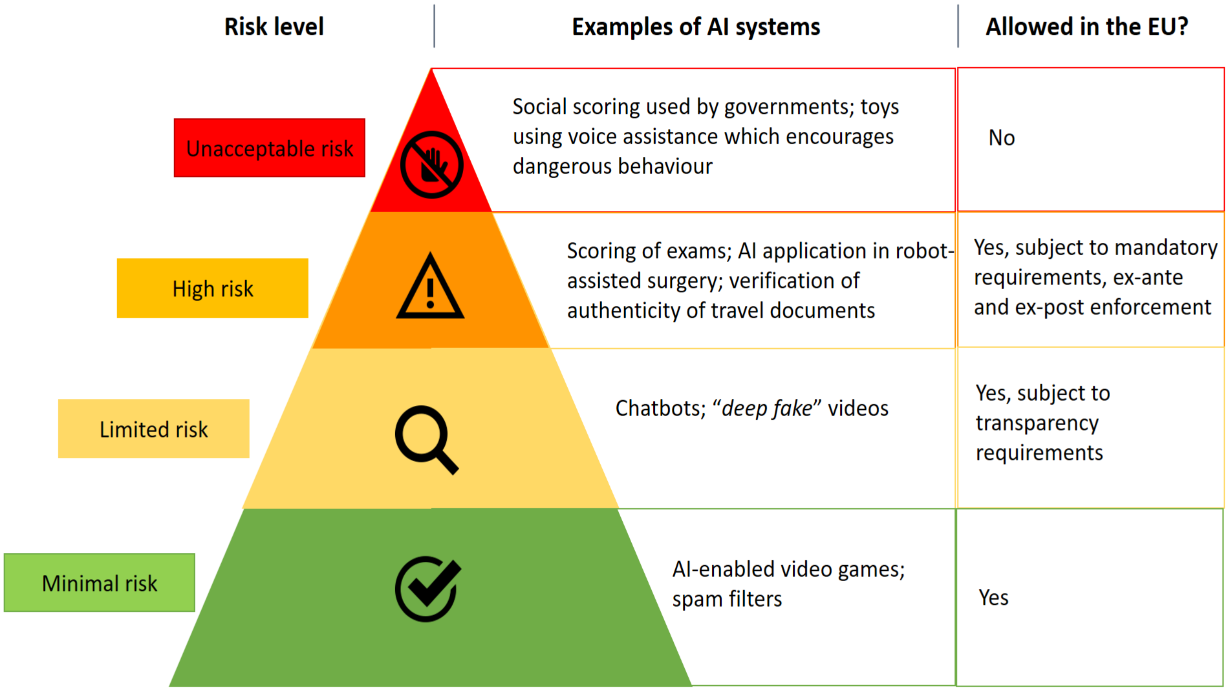
While the EC proposal does not give a detailed definition of AI, it defines *AI systems* as software developed using machine learning, knowledge-based, or statistical approaches (see Annex I of the Act[4]). The Commission proposal recognises that AI systems can be both a driver for economic and societal advancement but can also '*bring about new risks or negative consequences'*. Consequently, the centerpiece of the AI proposal is the introduction of four categories of risk, which set different obligations for the providers of AI systems.

At the top of this 'risk pyramid' the draft Regulation establishes an explicit list of prohibited AI practices, which create '**unacceptable risk'** as they can violate fundamental rights and go against EU values. This includes AI applications that have a high potential to '*manipulate persons through subliminal techniques* *beyond their consciousness*' or exploit the susceptibilities of specific vulnerable groups in a manner that could cause them or others physical or psychological harm. This provision is laudable, but open to interpretation. What is the objective threshold for psychological harm? How will this explicit prohibition apply to certain algorithms for behavioral nudging[5] that are commonly met online? Some of these techniques go beyond simple digital marketing tools for generating additional clicks. For example, complex recommendation algorithms in social media and video-sharing platforms can lead users to malign disinformation[6] or extremist online content.

An additional prohibition applies to AI-enhanced social scoring by public authorities for measuring the trustworthiness of individuals based on their social behaviour or predicted personality characteristics. The EU wants to signal that data-driven societal management, similar to China's nascent social credit[7] system, would pose unacceptable risks for European citizens. It seems that this provision also aims to prohibit future experiments which may lead to the creation of 'digital welfare states'. For instance, in 2020 a Dutch court ruled against attempts to optimise the country`s social welfare system through an algorithmic risk scoring system.[8]

Real-time remote biometric identification systems in public spaces for the purpose of law enforcement also appear in the list of banned AI systems. However, the Commission has envisaged several exceptions to the ban when this type of technology is used for the search of crime victims, prevention of imminent threats to the life of natural persons or the identification of criminal perpetrators. This provision remains contentious for many stakeholders, for instance civil society campaigners, who advocate[9] for a stronger ban on facial recognition in public spaces, as it is allegedly prone to misuse and potential discrimination of certain societal groups.

At the heart of the proposal are high-risk AI systems as the second layer of the pyramid (Fig. 1). The EC proposal considers an AI system to be '**high risk'** if (1) it is used as a safety component of a product falling under certain EU harmonised legislation (machinery, toys, medical devices) or (2) a stand-alone AI system in 8 defined sectors which can cause a high risk to the safety or health of EU citizens (full list can be found in Annex III). The Commission would be able to update this list through delegated acts.



Risk level	Examples of AI systems	Allowed in the EU?
Unacceptable risk	Social scoring used by governments; toys using voice assistance which encourages dangerous behaviour	No
High risk	Scoring of exams; AI application in robot-assisted surgery; verification of authenticity of travel documents	Yes, subject to mandatory requirements, ex-ante and ex-post enforcement
Limited risk	Chatbots; "*deep fake*" videos	Yes, subject to transparency requirements
Minimal risk	AI-enabled video games; spam filters	Yes

Fig. 1 AI pyramid of risk based on the European Commission proposal

High-risk AI systems would need to comply with a set of stringent ex-ante and ex-post requirements. Such AI products need to go through a prior conformity assessment before their rollout within the EU. A bundle of mandatory provisions is envisaged, e.g. high quality datasets, detailed technical documentation and appropriate level of human oversight among others. This assessment should be conducted by an independent national 'notified body' or by the provider of the AI systems depending on their specific type. AI systems which directly interact with natural persons (e.g. chatbots) and pose a risk for potential manipulation would need to comply with transparency obligations. Such systems would be considered as presenting '**limited risk'** and users should be fully aware that they are interacting with a machine.

At the bottom of the risk pyramid remain the rest of the AI systems considered as posing '**minimal risk'.** According to the Commission, these are the vast majority of future AI system. However, providers of such systems could voluntarily apply the mandatory requirements for high-risk AI systems or adhere to voluntary codes of conduct.

Governance and Enforcement

EU member states would have to designate one or more national competent authorities to oversee the implementation of the new Regulation. Each EU country has to appoint a national supervisory authority, which would act in two important capacities. First, it would act as a 'notifying authority' which would be responsible for drawing up the necessary procedures and appointing the independent bodies that verify the myriad of requirements for high-risk AI systems discussed above (the draft text refers to them as 'notified bodies'). The notified bodies would be able to issue certificates for compliance with the mandatory requirements valid for a period no longer than five years. All of the official notified bodies need to be registered in a designated list created by the European Commission.

Second, the national supervisory authority would also act as a market surveillance authority which controls the national market and investigates compliance with the necessary rules for high-risk AI systems. This would ensure the ex-post enforcement of the rules and provide public authorities with the necessary powers to 'intervene in case AI systems generate unexpected risks, which warrant rapid action'.

Additionally, the draft rules provide for the creation of a European AI Board comprised of representatives from the member states and chaired by the Commission. It would be responsible for facilitating the harmonised application of the Regulation across the EU and ensuring smooth cooperation between the national supervisory authorities.

For breaches of the Regulation's provisions, the EC has proposed certain thresholds for sanctions, similar to the General Data Protection Regulation (GDPR). Infringement of prohibited practices or non-compliance related to requirements on data should lead to administrative fines up to 30 million euros or 6 % of the worldwide annual turnover of the preceding financial year (whichever is higher). Non-compliance with any other requirement or obligation can lead to fines up to 20 million euros or 4 % of the worldwide annual turnover.

Discussion and Future Considerations

The Act will be subjected to the EU's legislative process, leading to scrutiny and amendments by the European Parliament and the European Council. Assuming a compromise version of the Act is passed, the final Regulation will enter into force after a further transition period of two years after the text is officially adopted.

One of the biggest challenges for the EU's future AI rules is the set-up of a coherent and effective governance framework across the continent. The intricate web of national bodies entrusted with the implementation of the Regulation might face budgeting or technical capacity issues. As a comparison, a number of national data protection authorities across EU member states have struggled to enforce[10] the GDPR due to problems with inadequate staffing[11] or limited resources at their disposal. In a similar way, EU member states might diverge in the way they supervise and enforce AI rules within their jurisdictions. At the same time, the Commission has tried to boost its own role and promote a more supranational approach toward governance. For instance, it is expected that the EC will have a strong role in the proposed European AI Board in charge of coordinating the cooperation among national bodies. Additionally, the EC would be able to officially challenge the competence of national notified bodies if there are substantial reasons to doubt whether the respective body complies with the necessary requirements. These are valid steps but a general concern remains about the adequate implementation and enforcement of the new AI rules across the EU.

The newly created European AI Board will function in parallel with the already existing European Data Protection Board. Additionally, the draft Digital Services Act[12] (DSA) proposes the creation of a European Board for Digital Services, which would contribute to the oversight of large online platforms. All of these structures will be comprised of national representatives with the support of the Commission. It would be interesting to observe how these quasi-federalist structures will interact in the future and whether this unique governance framework will yield the necessary results.

The AI proposal tries to address the potential regulatory burden on small businesses and start-ups. The EC has put forward the creation of regulatory sandboxes which would foster innovation by providing a controlled environment for development and testing of novel AI systems. National competent authorities will be in the lead for creating such sandboxes, i.e. smaller companies would rely on the proactivity of the respective national administration involved. It is uncertain if this would involve time constraints and would actually be beneficial. European policy makers should explore additional options for supporting SMEs and start-ups and reducing their costs for ensuring compliance with the new rules.

In parallel, the EC has also put forward the *2021 Coordinated Plan on AI*.[13] This plan maps out ways to accelerate private and public investment and foster better synergies between member states. The EC is committed to ensuring the EU's 'global leadership in trustworthy AI' even though Europe is still lagging behind in comparison to actors like the US or China.[14]

The EC's AI proposal joins ranks with other draft pieces of legislation (such as the DSA and DMA) in search of a fragile regulatory balance. How do you foster innovation and allow for novel services, while also protecting users from the negative externalities of invasive data-driven technologies? The draft AI Regulation is neither a panacea that will automatically guarantee Europe's leadership in AI nor should it be regarded as a draconian legislative over-reach that will stifle innovative potential across the continent. It should be regarded as a balanced attempt to welcome new technologies while also embedding fundamental European values in a rapidly expanding and increasingly complex digital landscape. It is essential that European policy makers manage to learn from previous regulatory mistakes and develop a flexible-enough binding framework with the help of all concerned stakeholders. This is a risky endeavour but a much needed one as it sends a clear signal to the global community that the AI race should not be allowed to become a race to the bottom.

How to cite this blog article (APA style):
Lilkov, D. (2021, June 14). The EU's New Rules on Artificial Intelligence: A Risky Endeavour. *AI Policy Blog*.

Notes

[1] Bostrom, N., *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press (2014)
[2] European Commission, *Proposal for a Regulation laying down harmonised rules on artificial intelligence*, 21 April 2021, COM (2021) 206 final
[3] European Commission, *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*, 19 February 2020, COM (2020) 64 final
[4] European Commission, Annexes to the Proposal for a Regulation of the European Parliament and of the Council, 21 April 2021, COM (2021) 206 final
[5] Matz, S.C. et al., *Psychological targeting in digital mass persuasion,* PNAS 28 November 2017 114 (48) 12714-12719, DOI: 10.1073/pnas.1710966114
[6] Bradshaw, S. et al, *2020 Global Inventory of Organized Social Media Manipulation,* Oxford Internet Institute (2020), accessed at: https://demtech.oii.ox.ac.uk/wpcontent/uploads/sites/127/2021/02/CyberTroop-Report20-Draft9.pdf
[7] Drinhausen, K., Brusse, V., *China's Social Credit System in 2021: From fragmentation towards integration,* Mercator Institute for China Studies (2021), accessed at: https://merics.org/en/report/chinas-social-credit-system-2021-fragmentation-towards-integration
[8] Natasha, L., *Blackbox welfare fraud detection system breaches human rights, Dutch court rules,* Techcrunch, 06 February 200, accessed at: https://techcrunch.com/2020/02/06/blackbox-welfare-fraud-detection-system-breaches-human-rights-dutch-court-rules/?renderMode=ie11

[9] European Digital Rights Initiative, *EU`s AI law needs major changes to prevent discrimination and mass surveillance* (April 2020), accessed at https://edri.org/our-work/eus-ai-law-needs-major-changes-to-prevent-discrimination-and-mass-surveillance/
[10] Lomas, N., *GDPR`s two-year review flags lack of 'vigorous' enforcement,* Techcrunch, 24 June 2020, accessed at https://techcrunch.com/2020/06/24/gdprs-two-year-review-flags-lack-of-vigorous-enforcement/
[11] Ryan, J., and Toner, A., *Europe`s governments are failing the GDPR,* Brave (2020), accessed at https://brave.com/wp-content/uploads/2020/04/Brave-2020-DPA-Report.pdf
[12] European Commission, Proposal for a Regulation on a Single Market for Digital Services and amending Directive 2000/31/EC, 15 December 2020, COM (2020) 825 final
[13] European Commission, *Annexes to the Communication from the Commission on Fostering a European Approach on Artificial Intelligence*, 21 April 2021, COM (2021) 205 final
[14] Castro, D. and McLaughlin, M., *Who Is Winning the AI Race: China, the EU, or the United States? — 2021 Update*, Information Technology and Innovation Foundation, 25 January 2021, accessed at: https://itif.org/publications/2021/01/25/who-winning-ai-race-china-eu-or-united-states-2021-update