One Act to Rule Them All

What Is At Stake In the AI Act Trilogue?

Filip Konopczynski

18.08.2023 doi: 10.17176/20230818-062853-0

The AI Act is marketed as the world's first comprehensive law regulating the development and implementation of artificial intelligence (AI). According to the EU officials, it will establish a legal framework for companies and institutions that create or use "intelligent" algorithms that is aligned with the system of fundamental rights and democratic values. From the market point of view, as AI became one of the fastest-growing sectors in the emerging technologies, EU regulations will also impact the outcome of the battle for dominance over the single market fought over by the largest global information and communication technology (ICT) corporations. For regular people, the ultimate judgment of the act will depend on whether the new law manages to forbid, restrict or contain the abusive ways governments and large digital platforms develop, and apply AI-based tools capable of surveilling, discriminating, and manipulating citizens and consumers.

Although much of the initial hype over ChatGPT has already faded away, it would be a mistake if the media, academia and civil society let the AI Act off the hook just yet. As we enter the very final stages of the negotiations, Big Tech and Member State governments can still undermine much of the (relative) progress that has been achieved in the parliamentary improvements to the original draft by the Commission. The lack of transparency of the trilogue itself, intense corporate lobbying, and the electoral kerfuffle that the Spanish Presidency leading the work on the file has found itself to be, mean that the battle over European AI is still raging. Fortunately – thanks to the efforts of civil society – there are also reasons for optimism.

In order to better understand what's at stake, who are the main actors and their motivations, and how to make one's mind about all the conflicting claims we need to dive into the legal, economic and political aspects of the AI Act. The aim of this piece is to contextualize major milestones in the negotiations, showcase some of its critical features and flaws, and present challenges it may in the near future pose to people affected by "smart" models and systems.

The Legislative Context: From Nerds to ChatGPT to Apocalypse

At the start of the legislative process, the Act was of interest mainly to start-ups and ICT companies, lawyers, administrators, and a handful of politicians who wanted to ride the AI wave. Everything changed last fall thanks to OpenAI and ChatGPT, when applications based on large language models such as GPT3, Midjourney, and Stable Diffusion that allow for the generation and analysis of any text, code, image, or sound were made available to the wider public. ChatGPT quickly became the fastest-downloaded application in the history of the internet (only recently to be overtaken by Threads). It also achieved the unthinkable: transformed AI regulation into a fancy, almost populist topic for politicians.

Soon after the realization of the promising, often outstanding benefits of LLMs – the limits for their use in medicine, science, or process automation are still to be discovered – came awareness of the scale and gravity of the risks posed by autonomous systems. From the rule of law and democratic values angle, AI programs and devices allow for the mass creation of generated content and deep fakes ideal for propaganda or disinformation. Current LLMs also violate our privacy laws (the data on which the GPT model was trained on was very likely collected in breach of the GDPR). The list of legal, economic, or democratic threats is extensive, and includes copyright infringe-

ment, cybersecurity incidents, fraud and forgery, as well as discrimination encoded in algorithms, or their overall negative impact on our mental health.

The application of such models and systems by governments, particularly when fed by citizens' data already uploaded to the digital administrative services, could bring about an even more perilous future. Thanks to AI (particularly GPT and other LLMs), government agencies now can quickly access and process information from various sources, which previously was either impossible or at least extremely time- and cost-consuming. Profiling, surveillance, discrimination, or persecution of political opponents have never been so easy and cheap, which is bad news for citizens but great for authoritarian governments around the world.

An incomplete Commission Proposal

To prevent such dark scenarios, the European Commission put forward a draft regulation in 2021 that would establish a legal framework for the development and use of AI in the EU. The proposed regulation, the Artificial Intelligence Act (AIA), is based on the principle of risk assessment.

When risk is insignificant, as in the case of most AI products and services, the obligations of companies and institutions that develop or introduce them are correspondingly minimal. In the case of generative AI or conversational chatbots, the AI Act puts an obligation to inform end users that they are not dealing with a person. Greater requirements appear when certain, invasive AI is to be used, particularly in education, critical resources and public utilities management, public administration, the judiciary, or the police. Such systems are to be considered high-risk, and their introduction to the market will require compliance with legal and technical standards and procedures. Finally, some AI applications have been considered so dangerous that they are to be completely prohibited on the market. Programs that manipulate our decisions, assess our behavior as citizens ("social scoring"), or discriminate on the basis of biometric health or other sensitive data, as well as (most) "smart" systems designed to surveil us on the streets, airports, or malls are to be outright banned. Companies and institutions that violate these rules will have to face the removal of the product from the market, as well as financial penalties. The harshest of them are up to 7% of global annual turnover- an amount that can shake even the largest global tech companies.

The European Commission's (EC) draft was followed by the member states' General Approach hammered out in the Council. The version proposed by the Council at the end of 2022 postulates a different, narrower and more technical definition of AI. It explicitly mentions specific statistical, algorithmic and computational techniques ("using machine learning and/or logic- and knowledge based approaches"), which critics say may narrow the potential use of AI Act by making it less technologically future-proof. Regarding the approach towards high-risk AI, an additional criteria of classification was added requiring the systems to pose a "significant risk" for people's health, safety and fundamental rights. This provision could constitute a potential gateway for a systemic lack of liability and accountability in the context of financial, information, and legal asymmetry between private or public institutions and consumers and individuals. The EU's ambition to pave the way for a global, human-centered, ethical, and responsible AI framework is difficult to reconcile with an admission that a "moderate" risk for fundamental rights violation complies with the Union's *aquis communitaire*.

The Council's perspective was shaped through and eventually managed to incorporate first lessons learned about generative AI. At the Council level, at the very last moment, the category of "General Purpose AI" entered the frame, however without profound alterations to the overall structure of the text.

Another major change adopted by the Council was an inclusion of a strong, blanket exemption of the area of "national security" from the scope of the Act. Thus, activity of intelligence agencies and, in certain cases (e.g., anti-terrorist forces) the police, the prosecution and other law enforcement agencies was treated by the Governments of the Member States in the same manner as the military, and science and R&D.

A Motivated Parliament

The explosion of interest in AI after the premiere of ChatGPT led to lively discussions in the European Parliament, which resulted in the submission of several thousand amendments during the parliamentary work. The text adopted by the Members of the European Parliament (MEPs) differs significantly from the first draft of the Commission and the text accepted by Member State governments in the Council in the fall of last year. Thanks to the engagement of civil society organizations, MEPs equipped the Act with provisions aimed at ensuring that users are properly protected.

The list of prohibited AI applications has been extended to include criminal prediction, systems based on emotion recognition, and biometric categorization of individuals. Additionally, the catalogue of high-risk AI has been extended to include recommendation algorithms of large online platforms (VLOPs according to the DSA), systems utilized in electoral campaigns, and AI-powered tools used in public administration for legalistic, procedural purposes. Transparency and monitoring obligations over high-risk AI systems are also to be increased by a mandatory registration in national, and EU (public institutions, VLOPs, systems operating in more than one country) database administered by the Commission.

High risks and foundation models

Even more importantly, high risk AI systems are to be assessed before being put on the market in terms of their impact on fundamental rights, health, or financial situation of affected persons. The introduction of the Fundamental Rights Impact Assessment (FRIA) framework, as well as major procedural rights for citizens into the text is a major improvement that would not be possible if not for the advocacy campaigns coordinated by NGOs within the European Digital Rights (EDRi) network.

The final stages of the European Parliament (EP) works also brought about changes addressing generative, multimodal AI such as GPT, Bard, and Bloom. MEPs agreed upon the introduction of a new term – "Foundation models", an alternative to the "General Purpose AI" in the Council's General Approach. Providers of such models would have to comply with the EU law, including in the field of privacy, prohibition of discrimination or copyright. The modification appeared at the very end of the negotiations, and (perhaps not surprisingly) the new category ("models" instead of "systems") is relatively favorable for the leading AI companies. For instance, the catalog of penalties was revised, and in the worst case the provider responsible for infringements would be administered the lowest possible fine – 2% of the annual global turnover.

What's still missing?

Even though the EP's amendments to the AI Act are a significant improvement, concerns prevail. The biggest risk to the effective implementation of the AI Act is the potential inefficiency of the mechanisms that shall protect fundamental rights, especially civil and political rights.

Real-time biometric identification

The lack of consistency in the ban on biometric surveillance is an especially grave reason for concern. Real-time biometric identification (RBI), although still erroneous, is a dangerous technology. It allows people to be identified, and traced in real time based on their appearance, movement, or other distinctive features. In order to be fully operational, the model on which RBI operates must be trained on huge datasets, including data of ordinary people. This technology itself is far from flawless, since there have already been several reports of innocent people have been detained or accused as a result of RBI 's mistakes. A total ban on RBI is vital, for even minor exceptions could lead to governments developing or buying such systems "just in case". For that reason alone it is vital to block the adoption of this technology, even if it's scope of application would be limited. The use of RBI in the area of migration policy or to categorize political dissenters as suspicious would be particularly dangerous.

Accountability, assessment, redress

Another concerning issue relates the agreement with Council's changes to the categorization of high-risk systems, and the insertion of an ambiguous criterion of "significance" of potential harm to people. To add to the worse, It remains unclear how effective the

risk assessment procedure can actually be given the EP's draft equals third-party audits with the ones carried out in-house. Simply put, AI developers and organizations using AI shall self-police. In contrast, supervisory bodies will have limited time to deny the deployer's registration application of potentially dangerous systems before they hit the market. As a result, the enhancement of the high-risk list may in the end prove irrelevant due to a (perhaps deliberately) erroneous classification, registration and oversight mechanisms. It is also not certain whether citizens will be effectively able to file complaints with the authorities or the courts against the decisions of AI systems, especially in cases when the deployer is a public entity. At the stage of implementing the new law, another major problem may be the lack of possibility for social organizations to represent people who have been harmed by AI, as well as the right for organizations to act in the public interest to prevent them from doing actual harm. Without the access to detailed, technical information as well as support of ICT specialists, exercising one's rights may prove to be cumbersome.

Procedural Rights

The right to an explanation of a decision made using AI is also too narrow. Even in the most citizen-friendly version of the AI Act, an explanation is only possible in the case of high-risk systems, such as those that analyze emotions, candidates in school recruitment, or AI used in the judiciary. As AI becomes an increasingly common element of products and services, such a legal limitation could significantly weaken consumer protection.

The too narrow scope of transparency obligations for providers and users of high-risk AI used by private entities is also a bad omen for the future. The current versions of the AI Act assume that high-risk systems created by private entities and used only within one country will be registered, and supervised in national databases. This not only reduces transparency, and provides an easy administrative route towards an abusive use of invasive AI systems capable of undermining the fundamental rights of individuals at a mass scale.

The Trilogue

The trilogue formally began immediately after the vote in the European Parliament. This means that the debate has been moved from the democratic premises (although the transparency and openness of the process leave much to be desired) to the realm of political haggling between national and EU-level politicians. The access to this process is strictly limited not only for the media, but also for representatives of civil society, academia, or the AI start-up community. The situation got even more complicated due to the snap elections in Spain, a country that holds the EU Presidency until the end of the year. Even if Sánchez' government manages to hold onto power, its ambitions regarding the AI Act may fall victim to internal politics, leading to the process being yet again delayed.

What is at stake?

The stakes in the trilogue from the perspective of rights of EU citizens are twofold. Some governments continue to opt for a blanket exemption for national security use cases from the scope of the AI Act at large. In practice, this would mean that whenever government officials consider someone or something a "threat to the national security", it will have the discretionary power to decide whether or not the Act is applicable, thus denying individuals and groups affected by AI systems access to legal means securing their fundamental rights.

National Security Exemption

Weakening provisions protecting citizens' rights can be also achieved, should the parties reduce the list of prohibited systems (especially in the case of tech used by migration and border control agencies or the police), exclude certain areas from the scope of the regulation altogether (intelligence agencies, law enforcement agencies), narrow down obligations for public institutions implementing (particularly high risk) systems, or allow the oversight mechanisms to remain solely within the control of national governments. Either way, such alterations would have profound implications, particularly for migrants, refugees and other non-EU citizens, marginalized communities and the underprivileged. For citizens of

countries struggling with authoritarian tendencies, democratic backsliding, or human rights or rule of law violations the lack of effective legal protections, and independent oversight over AI systems used by governments could also translate to an even more perilous situation for minorities, opposition politicians, or civil society activists.

Big Tech's Lobbying

Secondly, big business has not yet had its final say. The very moment the trilogue officially took off, an open letter attacking the AI Act signed by over 150 representatives from European companies such as Airbus, Renault, Siemens (as well as Heineken, and Danone) was published. Its message is clear: we do not want what they portrayed as restrictive and innovation-stifling regulations. However, the main role in the private sector advocacy campaign regarding the Act is played by non-European companies: Microsoft, Alphabet and OpenAI. Officially, tech giants proclaim deep commitment to developing procedures based on ethical standards, building responsible artificial intelligence, and programs designed for the benefit of users. Behind the doors, however, as we learned from a report by the European Corporate Observatory, Google, Microsoft, and OpenAI are simultaneously spending big money on large-scale lobbying efforts aimed at ensuring that the AI Act imposes the least possible and least costly requirements on companies.

The main focus points of Big Tech lobbyists are the generative AI legal framework, and the obligations for importers, providers and deployers of high-risk systems. In both cases the ultimate goal is to limit the scope of responsibilities and legal liability and weaken compliance mechanisms (classification of systems as high risk, fundamental rights impact assessment, mandate and competences of the Commission and the national supervisory authorities). Or, at the very least, complicate them to a point of procedural ineffectiveness or inaccessibility. Lobbying activities peaked during the tournee by none other than Sam Altman, who in June visited, and met with government officials in Brussels, Warsaw, Madrid, and Munich. In addition to theorizing about the future apocalypse caused by sentient machines, the head of Open AI appealed for "global AI regulation". What distinguishes the Act from international law within the UN framework is that as an element of the EU law it applies directly both to States, and the private sector companies. Although a global treaty on AI is desirable, its importance as well as feasibility should not be overestimated, especially now, on the verge of the introduction of a first international legal act dedicated to AI as technology.

Conclusion

The AI Act is not a brainchild of revolutionary activists and critical academics but a result of the Commission's attempt to foster the development of the European AI sector. As an element of the EU's new digital policy agenda, its main goal is to meet expectations of EU entrepreneurs by providing legal certainty within the single market and encourage public and private sector institutions to implement AI services fast, at a large scale, and without compromising security. Thus, most AI systems will not be subject to scrutiny, and large part of the Act is devoted to sandboxes – regulatory spaces in which a given model, system or AI program can be safely experimented with and tested before meeting all requirements. Given new, significant funds allocated towards AI research and development in recent years by the Commission it is all the more incomprehensible why businesses are so unambiguously and vehemently critical of the Act.

Instead of pondering about far-fetched scenarios reminiscent of Stanley Kubrick's '2001: Space Odyssey', Sam Altman and other CEOs of leading AI companies, should address more realistic, short-term questions. Were datasets that GPT models were trained on compiled in breach of the GDPR? Does OpenAI or Microsoft (which at this point has incorporated GPT into its business model) intend to supply AI technology to governments that use AI to EU states with an ongoing record of violating democratic principles and the rule of law? Will Alphabet or Meta (continue to) export or just enable its potentially harmful products to countries without any effective human or civil rights protection?

An outcome of the Trilogue, in which the parties strike a deal based on limiting obligations related to the security, safety, and liberties of individuals affected by AI algorithms for both the government and the private sector is easy to imagine. Such agreement would make several tech executives happy, but come at the expense of citizens and consumers. For people living in countries that have been struggling with maintaining rule of law or human rights violations, and undergoing societal polarization even before the advent of generative AI, these consequences will be dire, if not irreversible.

ABOUT THE AUTHOR

Filip Konopczynski

Filip Konopczyński is a lawyer and analyst specializing in AI and emerging technologies regulation. Member of the Panoptykon Foundation, a Polish watchdog working towards more social control over surveillance practices and policies, particularly in the realm of rights in digital spaces.Filip Konopczyński is a lawyer and analyst specializing in AI and emerging technologies regulation. Member of the Panoptykon Foundation, a Polish watchdog working towards more social control over surveillance practices and policies, particularly in the realm of rights in digital spaces.

SUGGESTED CITATION

Konopczynski, Filip: One Act to Rule Them All: What Is At Stake In the AI Act Trilogue?, Verfassungsblog, 18.08.2023, doi: 10.17176/20230818-062853-0.

Licensed under CC BY-SA 4.0.

Langzeitarchiviert bei IntRechtDok unter: doi.org/10.17176/20230818-062853-0.