Article

Wilfried Martens Centre for European Studies

European View
2021, Vol. 20(2) 166–174

DOI: 10.1177/17816858211059248
journals.sagepub.com/home/euv
SAGE

Regulating artificial intelligence in the EU: A risky game*

Dimitar Lilkov

Abstract

The EU continues its quest to draw the contours of innovative legislation for the digital domain. The European Commission's draft Regulation on artificial intelligence (AI) is a clear departure from previous 'soft' attempts to set the rules through ethical principles and industry pledges. The EU aspires to be the first global player to adopt a comprehensive framework that classifies and regulates the roll-out of AI software and hardware within its internal market. The draft rules try to provide legal certainty for public and private bodies across the EU, while making sure that potential risks to its citizens are minimised. This article sketches out some of the most important provisions of the draft Regulation and tries to critically assess its potential shortcomings related to implementation and enforcement. The final version of the AI proposal should avoid the mistakes of previous attempts to draft transnational rules for the online space and establish a sufficiently flexible legal framework.

Keywords

Artificial intelligence, Ethics, European Commission, Regulation, Risk

Introduction

The advent of complex algorithms, machine learning and automated decision-making processes is no longer a prospect of the distant future. Smart software and hardware are steadily being rolled out across the EU in order to provide convenience, connectivity and new services to users. From a regulatory perspective, however, this opens the door for new challenges and frictions. How do you reap the benefits of artificial intelligence (AI) and advanced technology, while ensuring that they do not cause societal harm? It would

Corresponding author:
D. Lilkov, Wilfried Martens Centre for European Studies, 20 Rue du Commerce, Brussels, B-1000, Belgium.
Email: dli@martenscentre.eu

*An earlier version of this article was published on the AI Policy Consulting Blog and can be found at www.aipolicyconsulting.com/the-eu-s-new-rules-on-ai

appear that this was the foundational question that guided the experts behind the European Commission's latest regulatory proposal on AI (European Commission 2021).

In the last few years, the debate on AI governance has mostly been centred on ethical frameworks and pledges for foundational AI principles. Leading international organisations (OECD 2019) and international fora (*G20 Information Centre* 2019) have issued their considerations for the ethical and trustworthy application of AI, while a number of multinational digital companies have raced ahead to announce the principles of AI application and development in their services too (Pichai 2018; Amershi et al. 2019). The EU's own initial efforts led to the creation of a high-level expert group on AI, which presented its *Ethics Guidelines for Trustworthy AI* in 2019. The EU's initiative received a positive welcome, even though there were allegations that private industry had had too much of an influence within the expert group and that the final result was one of 'ethics-washing' (*Tagesspiegel* 2019).

All of these documents or pledges share a similar commitment to human dignity and protecting basic rights, as well as to providing unbiased algorithms, accountability and the safeguarding of user privacy. Interestingly, similar noble commitments to ethics and trustworthiness can be found in the white papers of Chinese scientific organisations, private companies and government-affiliated Chinese think tanks (MERICS 2021). One can only wonder if any of these commitments is tenable, given China's grim human rights track record and its unique model of state-led digital authoritarianism (Lilkov 2020).

For EU policymakers, it has become apparent that neither self-regulation by the industry nor ethical standards will suffice in such a complex field. With its April 2021 legislative proposal (European Commission 2021), the Commission takes a clear step beyond ethics by introducing binding rules for AI systems placed on the EU market. The draft AI Regulation would apply to both public and private actors, providers and users of such systems, irrespective of whether they are established within the Union or in a third country. Given the size and importance of the European internal market, EU policymakers intend to create a spillover effect and nudge other countries to develop similar frameworks. 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 issues related to new technologies (European Commission 2020b), as well as a revision of the relevant sectoral safety legislation.

This article tries to provide a brief overview of the voluminous draft text and its innovative framework for regulating the potential risks and harms generated by specific uses of AI. It is important to note that the Commission does not aim to regulate the advanced technology itself. The goal is to create future-proof harmonised rules on the use of AI that will ensure trust among users, increase its uptake, and provide legal certainty for public bodies and private companies. The following sections try to assess some of the most important provisions of the Commission's proposal.

Pyramid of risk

While the Commission's draft does not give a definition of the term 'artificial intelligence', it defines AI systems as software that is developed with certain machine learning, knowledge-based or statistical approaches (the full list can be found in European Commission 2021, Annex I). The proposal recognises that while AI systems can be a driver for economic and societal advancement, they can also 'bring about new risks or negative consequences' (European Commission 2021, 1). Consequently, the centrepiece 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' (European Commission 2021, Title II) as they violate fundamental rights or go against EU values. The prohibitions cover AI applications that have a high potential to 'manipulate persons through subliminal techniques beyond their consciousness' (European Commission 2021, 12) 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 behavioural nudging that are commonly found online (Matz et al. 2017)? Some of these techniques go beyond simple digital marketing tools for generating additional clicks. For example, complex recommendation algorithms in social media and on video-sharing platforms can lead users to polarising disinformation or extremist online content (Bradshaw et al. 2020). Given the likely challenges related to ascertaining a direct link between algorithmic manipulation and obvious harm to the user, it is questionable whether this outright prohibition will actually have practical value in the future.

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 system (Drinhausen and Brusse 2021), 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 the use of an algorithmic risk-scoring system (Lomas 2020a).

Real-time remote biometric identification systems (i.e. facial recognition 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 could be used to search for the victims of crime, to prevent imminent threats to life or to identify criminal perpetrators (European Commission 2021, art. 5). This provision remains contentious for many stakeholders, for instance civil society campaigners, who advocate a stronger ban on facial recognition in public spaces, as it is

allegedly prone to misuse and could potentially discriminate against certain societal groups (*EDRI* 2020). There are also shortcomings related to the scope of the prohibition—it applies only to 'real-time' systems, which excludes systems that could biometrically analyse footage from a recording and could have a 'chilling effect on the exercise of fundamental rights and freedoms' (EDPB and EDPS 2021, para. 31). An additional caveat is that the prohibition does not preclude actors from using such biometric identification for non-law enforcement purposes, such as crowd control or public health (Veale and Borgesius 2021, 9).

Title III of the draft can be seen as an essential section of the Regulation as it deals with the classification of AI systems which are considered to be 'high risk'. The Commission's proposal considers an AI system to be high risk if it is either a safety component of a product falling under certain EU harmonised legislation (machinery, toys and medical devices) or a stand-alone AI system in other defined sectors which could pose a high risk to the safety or health of EU citizens (the full list can be found in European Commission 2021, Annex III). The Commission would be able to update this list through delegated acts.

High-risk AI systems would need to comply with a set of stringent *ex ante* requirements, including a conformity assessment, before their rollout within the EU. A bundle of mandatory obligations is also envisaged *ex post*; for example, the provision of high-quality datasets, detailed technical documentation, comprehensive record-keeping and an appropriate level of human oversight. Chapters two and three of the section on high-risk systems provide an exhaustive list of all the responsibilities and obligations that fall on the providers of such systems. The assessment should be conducted by an independent national 'notified body' or by the provider of the AI system, depending on its specific type.

What would be the cost of compliance with such a stringent set of requirements? A study requested by the European Commission suggests that the estimated cost for an AI product to comply with the potential new requirements would be close to €10,000, while the purchase of additional services or staff could increase the cost to €30,000 (Renda et al. 2021). These are, of course, approximate figures, which depend on many future factors that might push costs up or down. Here, one can make the valid supranational argument that the absence of harmonised legislation might bring about a fragmented system of national requirements, which would lead to even higher compliance costs for European businesses as they would have to deal with a patchwork of AI regimes and national standards.

At the bottom of the 'risk pyramid', the draft text sets out the obligations for AI systems which are regarded as posing 'limited' or 'minimal' risk. AI systems which directly interact with real people (e.g. chatbots) and present a risk of potential manipulation would need to comply with transparency obligations. Such systems would be considered of 'limited risk' and users should be fully aware that they are interacting with a machine or software algorithm.

Lastly, all other AI systems will be viewed as posing 'minimal risk'. According to the Commission, these will comprise the vast majority of future AI systems. 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 (European Commission 2021, art. 59). Each EU country would have to appoint a national supervisory authority, which would act in two important capacities. First, it would act as the 'notified body' which would be responsible for drawing up the necessary procedures and appointing the independent authorities that would verify the myriad requirements for the high-risk AI systems discussed above. The notified bodies would be able to issue certificates for compliance with the mandatory requirements. These certificates would be valid for a period of no longer than five years. All official notified bodies would need to be registered on a list created by the European Commission.

Second, the national supervisory authority would also act as a market surveillance authority, controlling the national market and investigating compliance with the necessary rules for high-risk AI systems. This would ensure *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' (European Commission 2021, 15).

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 Commission has proposed certain thresholds for sanctions, similar to those for the General Data Protection Regulation. Infringement of prohibited practices or non-compliance related to data requirements would lead to administrative fines of up to €30 million or 6% of the company's worldwide annual turnover in the preceding financial year (whichever is higher) (European Commission 2021, art. 71). Non-compliance with any other requirement or obligation could lead to fines of up to €20 million or 4% of global annual turnover (European Commission 2021, art. 71).

Future considerations

These are just some of the most noteworthy provisions of the proposed Regulation. The text has yet to be subjected to the legislative scrutiny of the European Parliament and Council. The final Regulation will apply after a further transition period of two years after the text is officially adopted.

One of the biggest challenges to implementing the EU's future AI rules is the setting 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 the EU member states have struggled to enforce the General Data Protection Regulation (Lomas 2020b) due to problems with inadequate staffing or limited resources (Ryan and Toner 2020). In a similar way, EU member states might diverge in the way that 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 to governance. For instance, it is expected that the Commission will have a strong role in the proposed European AI Board. Additionally, the Commission would be able to officially challenge the competence of national notified bodies if there were substantial reasons to doubt whether they were complying 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.

Something curious can be observed here. The newly created European AI Board will function in parallel with the existing European Data Protection Board (European Parliament and Council 2016, art. 68). Additionally, the draft Digital Services Act (European Commission 2020a) 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 will be interesting to observe how these quasi-federalist structures will interact in the future and whether this unique governance framework will yield the required results.

The AI proposal tries to address the potential regulatory burden on small businesses and start-ups. The Commission (2021, art. 53) has suggested the creation of regulatory sandboxes which would foster innovation by providing a controlled environment for the development and testing of new AI systems. The national competent authorities would be at the forefront of creating such sandboxes, meaning that smaller companies would rely on the proactivity of their respective national administration. It is unclear whether this would create time constraints or would actually be beneficial. European policymakers should explore additional options for supporting small and medium-sized enterprises and start-ups and reducing their costs for compliance with the new rules.

In terms of AI funding or improving the EU's standing globally, the European Commission has put forward the 2021 Coordinated Plan on AI. The plan maps out ways to accelerate private and public investment and foster better synergies between member states. The Commission is committed to ensuring the EU's 'global leadership in trustworthy AI' even though Europe is still lagging behind in comparison to actors such as the US or China (Castro and McLaughlin 2021). The draft Regulation is the first global attempt to legislate on AI and set up the necessary obligations for the providers of complex systems. Even though these issues are also being debated in the US, Washington has

placed more emphasis on the implications of AI for national security and how the country can claim leadership in a domain heavily contested by China (NSCAI 2021).

In closing, it should be mentioned that the Commission's AI proposal joins ranks with other draft pieces of EU legislation in search of a fragile regulatory balance in the digital domain. How can innovation be fostered and allow for new services, while also protecting users from the negative externalities of invasive data-driven technologies? The draft AI Regulation is neither a panacea which will automatically guarantee Europe's leadership in AI nor a draconian legislative over-reach that will stifle innovative potential across the continent. It should be seen as a unique attempt to welcome new technologies but also embed basic European values in a chaotic digital universe that is expanding beyond our comprehension. It is essential that European policymakers learn from previous regulatory mistakes and develop a flexible-enough but binding framework with the help of all concerned stakeholders. This endeavour is risky but much needed as it will send a clear signal to the global community that the AI race should not be allowed to become a race to the bottom.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. *Guidelines for human-AI interaction*. Microsoft Research. doi.org/10.1145/3290605.3300233.

Bradshaw, S., Bailey, H., & Howard, P. (2020). *2020 global inventory of organized social media manipulation*. Oxford Internet Institute. https://demtech.oii.ox.ac.uk/wpcontent/uploads/sites/127/2021/02/CyberTroop-Report20-Draft9.pdf. Accessed 6 October 2021.

Castro, D., & McLaughlin, M. (2021). Who is winning the AI race: China, the EU, or the United States? — 2021 update. *Information Technology and Innovation Foundation*, 25 January. https://itif.org/publications/2021/01/25/who-winning-ai-race-china-eu-or-united-states-2021-update. Accessed 6 October 2021.

Drinhausen, K., & Brusse, V. (2021). *China's social credit system in 2021: From fragmentation towards integration*. Mercator Institute for China Studies. 3 March. https://merics.org/en/report/chinas-social-credit-system-2021-fragmentation-towards-integration. Accessed 6 October 2021.

EDPB & EDPS (European Data Protection Board & European Data Protection Supervisor). (2021). *Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. 18 June. https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en. Accessed 6 October 2021.

EDRI (European Digital Rights Initiative). (2020). EU's AI law needs major changes to prevent discrimination and mass surveillance. April. https://edri.org/our-work/eus-ai-law-needs-major-changes-to-prevent-discrimination-and-mass-surveillance/. Accessed 6 October 2021.

European Commission. (2020a). *Proposal for a Regulation on a single market for digital services (Digital Services Act)*. COM (2020) 825 final, 15 December 2020. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en. Accessed 6 October 2021.

European Commission. (2020b). *Report on the safety and liability implications of artificial intelligence, the Internet of Things and robotics*. COM (2020) 64 final, 19 February. https://eur-lex.

europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX%3A52020DC0064. Accessed 6 October 2021.

European Commission. (2021). *Proposal for a Regulation laying down harmonised rules on artificial intelligence*. COM (2021) 206 final, 21 April 2021. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF. Accessed 6 October 2021.

European Parliament and Council. (2016). Regulation (EU) no. 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). OJ L119 (27 April), 1. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN. Accessed 6 October 2021.

G20 Information Centre. (2019). Ministerial statement on trade and digital economy. 9 June. http://www.g20.utoronto.ca/2019/2019-g20-trade.html. Accessed 6 October 2021.

Lilkov, D. (2020). *Made in China: Tackling digital authoritarianism*. Wilfried Martens Centre for European Studies. Brussels, 11 February. https://www.martenscentre.eu/publication/made-in-china-tackling-digital-authoritarianism/. Accessed 6 October 2021.

Lomas, N. (2020a). Blackbox welfare fraud detection system breaches human rights. Dutch court rules. *Techcrunch*, 6 February. https://techcrunch.com/2020/02/06/blackbox-welfare-fraud-detection-system-breaches-human-rights-dutch-court-rules/?renderMode=ie11. Accessed 6 October 2021.

Lomas, N. (2020b). GDPR's two-year review flags lack of 'vigorous' enforcement. *Techcrunch*, 24 June. https://techcrunch.com/2020/06/24/gdprs-two-year-review-flags-lack-of-vigorous-enforcement/. Accessed 6 October 2021.

Matz, S. C., Kosinski, M., Nave, G., & Stillwel, D. (2017). Psychological targeting in digital mass persuasion. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1710966114.

MERICS (Mercator Institute for China Studies). (2021). *Lofty principles, conflicting incentives: AI ethics and governance in China*. 24 June. https://merics.org/en/report/lofty-principles-conflicting-incentives-ai-ethics-and-governance-china. Accessed 6 October 2021.

NSCAI (National Security Commission on Artificial Intelligence). (2021). *Final report: National Security Commission on Artificial Intelligence*. 19 March. https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf. Accessed 6 October 2021.

OECD. (2019). *Recommendation of the Council on artificial intelligence*. 22 May. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Accessed 6 October 2021.

Pichai, S. (2018). AI at Google: Our principles. *Google Blogs*, 7 June. https://www.blog.google/technology/ai/ai-principles/. Accessed 6 October 2021.

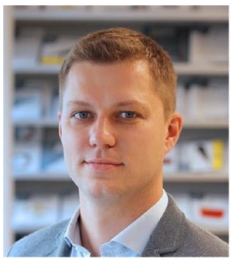
Renda, A., Arroyo, J., Fanni, R., Laurer, M., Sipiczki, A., Yeung, T., Maridis, G., Fernandes, M., Endrodi, G., & Milio, S. (2021). *Study to support an impact assessment of regulatory requirements for artificial intelligence in Europe*. 21 April. https://op.europa.eu/en/publication-detail/-/publication/55538b70-a638-11eb-9585-01aa75ed71a1/language-en/format-PDF/source-204305195. Accessed 6 October 2021.

Ryan, J., & Toner, A. (2020). *Europe's governments are failing the GDPR*. Brave. https://brave.com/wp-content/uploads/2020/04/Brave-2020-DPA-Report.pdf. Accessed 6 October 2021.

Tagesspiegel. (2019). Ethik-Waschmaschinen made in Europe. 8 April. https://background.tagesspiegel.de/ethik-waschmaschinen-made-in-europe. Accessed 6 October 2021.

Veale, M., & Borgesius, F. (2021). Demystifying the draft EU Artificial Intelligence Act. *Computer Law Review International*, *22*(4). https://ssrn.com/abstract=3896852. Accessed 6 October 2021.

Author biography



Dimitar Lilkov *is a research officer at the Wilfried Martens Centre, where he is responsible for matters involving the digital economy, energy and the environment. Dimitar is the host of the Martens Centre's 'Brussels Bytes' podcast series on technology and European policy.*