The European Parliament's AI Regulation: Should We Call It Progress?

Meeri Haataja
CEO and Founder of Saidot

Joanna J. Bryson
Hertie School, Centre for Digital Governance

Abstract

The European Union (EU) has been leading the world with its influential digital regulation. However, the EU's legislative process is sufficiently complex and careful that some national legislation clearly influenced by the EU's AI Regulation is already in place in other countries, before the law has even been finalized in the EU. Meanwhile, other states and regions are just beginning to develop AI policy. For both the EU and such others, we here describe the outcomes of the first round of legislative action by one of the EU's two legislative bodies, the European Parliament, in terms of modifying the Artificial Intelligence Act. The Parliament has introduced a number of changes we consider to be enormously important, some in a very good way, and some in a very bad way. At stake is whether the AI Act really brings the power and strength of product law to continuously scale improved practice on products in the EU with intelligent components, or whether the law becomes window-dressing aimed only at attacking a few elite actors *post hoc*. We describe here the EU process, the changes and our recommendations.

Keywords: AI Act; digital governance; AI regulation; parliamentary processes; European Union; transnational regulation.

[A] INTRODUCTION

Roughly two years ago, the European Commission announced a regulatory proposal for artificial intelligence, the Artificial Intelligence (AI) Act (European Commission 2022b). This Act is already undoubtedly one of the most influential regulative proposals for AI globally, with clear echoes in law from Brazil to China, as well as impacting on regulatory discourse in the United States. In August 2021, we analysed the core policy concepts of the AI Act, its strengths and weaknesses, to provide

input to policy-makers and anyone else affected by the coming regulation (Haataja & Bryson 2022). Since then, the AI Act has proceeded through the legislative process of the European Union (EU). A key step in this process is the adaptation of the Act by the Commission's key stakeholders: the two legislative bodies of the EU—the Council of the European Union (also known as the Council of Ministers) and the European Parliament. The Council of Ministers is a rotating body of ministers from whichever of the EU's 27 constituent states currently holds its presidency, whereas the Parliament is directly elected by the 375 million eligible voters of the EU. Now the legislative process is nearly completed: the Council adopted its final compromise version of the AI Act on 21 November 2022, and the European Parliament is expected to adopt its own final version in June 2023. These "final" versions are critical to the true final outcome, which results from trilateral talks between the two legislative bodies and the Commission.

In this article, we again aim to influence these final outcomes. Here, we consider the key provisions of the European Parliament's new proposals. We briefly analyse the amendments prepared during the parliamentary process and adopted by the two leading committees in May 2023. This analysis is intended to help pave the way for the next and final stage, trilateral negotiations between the European Parliament, the Council and the Commission, which will soon be underway, with an expected completion before the end of the year. We hope this analysis can help readers understand the strengths and weaknesses of the European Parliament's present proposal and its position in anticipation of the coming negotiations. Our perspective, as earlier in our assessments, is based primarily on the expected practical impacts of the proposed adaptations on the providers and deployers of AI-based systems, though also, of course, where relevant considering these products' ultimate consumers.

[B] THE EUROPEAN PARLIAMENT IS HIGHBALL-ANCHORING

It seems that the proposal on the table is set intentionally high in expectation of coming compromises needed in the trilateral negotiations. Readers should be mindful of this and certainly not consider the proposal as a collectively adopted version fully agreed by all key stakeholders.

Nevertheless, some of these extraneous additions raise questions about the legitimacy of the requirements overall. For example, additions related to general principles (article 4a) and AI literacy (article 4d) in chapter 4 seem to target strengthening the influence of the Act beyond the high-risk

systems. While we are strong proponents of AI transparency, the European Parliament's version seems to both unnecessarily and impractically require the extension of AI literacy and educational obligations to the industry. We suggest instead *setting the requirements to transparency to be in line with established literacy*. Rather than putting AI players in the role of government in themselves ensuring education, this would simply motivate players with sufficient capacity to consider contributing to those endeavours in order to reduce their own efforts and liabilities in achieving transparency.

We are particularly worried that such extraneous requirements beyond what is necessary can motivate an avoidance of being classified as AI. As we have argued previously (Bryson 2022; Haataja & Bryson 2022), the regulatory considerations the Commission has chosen to address in this Act are broad and the compliance burden rather light-weight, relative to other sectors. Ideally, any system capable of generating actions deemed "high risk" (that is, essentially, altering human lives, such as medical devices or welfare decisions) would provide adequate documentation for such actions or decisions to be auditable and adequately explainable for courts to determine if the decisions were correct, or at least the product of due care and diligence. The Commission itself set the precedent of recognizing that liability—a significant component of the earlier White Paper (European Commission 2020)—as a horizontal concern better handled by updating product liability law (an excellent draft of this update and the new AI Liability Directive are already available: European Commission 2021, 2022a).

[C] DEPLOYER OBLIGATIONS ARE CLEARER, BUT NEW ONES GO BEYOND THE NECESSARY

One of the significant changes by the European Parliament is when it comes to the role of deployers. First, it hugely clarified the Act's text by adopting the term "deployer" rather than "user" for those institutions deploying AI in their products or services. The term "user" was ambiguous as it is ordinarily applied to the end-users with no role in the design of the AI system (Haataja & Bryson 2022). Now, however, the responsibilities and obligations of deployers of high-risk systems are a clear change from any previous versions. Earlier versions set expectations for the deployers not only to use the systems as instructed by their providers, but also to exercise human oversight, keep automated logs, and monitor their

systems once active. But the European Parliament has now gone on to ask for more.

The deployers under the present European Parliament draft should also inform the end-users that they are subject to the use of a high-risk AI system and offer complaint-handling and redress procedures for affected individuals. This seems a logical step, assuming the levels of capacity and costs are proportionate to the amount of potential harm caused. Deployers can if they choose demand these capacities as part of procurement, and providers can compete in providing efficient and effective tools for such processes. However, the biggest change for deployers clearly comes from a new obligation to conduct a fundamental rights assessment of a high-risk AI system before deploying the system into use. So, for example, a recruiter using human resources (HR) tools in its recruitment process would be obliged not only to ensure that the provider of the system complies with the AI Act, and to use the system according to the instructions provided by the provider, but also to conduct their own individual fundamental rights assessment of the use of such a system (article 29a). In addition, the deployer of an HR AI system would be required to consult workers' representatives to reach an agreement and inform the affected employees that they will be subject to the system (article 29(5a)).

Hiring clearly has life-changing consequences, so certainly it falls under the category of "high-risk" AI that requires sufficient oversight to ensure that no errors or illegal biases are introduced through such a system. But this critical impact on lives and communities is true of hiring *whether or not* AI is utilized. In fact, with a well-written digital system, potential applicants and employees are likely to have more access to explanation and more recourse to remedies than if decisions were being made over thousands of applicants by overworked humans. AI trained through machine learning by default reflects the same biases as the humans who would otherwise be doing the procedures without AI augmentation (Caliskan & Ors 2017). In practice, we hear anecdotally that HR departments have found AI an excellent way to bypass implicit biases and reveal diverse candidates previously overlooked.[1] Thus adding the extra burden of a fundamental rights review only in the case when software is secured by the AI Act makes no sense. The AI Act should not introduce burdens unless they are directly relevant only to AI, not to the

[1] Unfortunately, for perhaps obvious liability reasons, we have failed to find anyone willing to go on the record about such improvements, but we have heard this from multiple sources and no counter-narratives. Famously, Amazon caught such an error in its AI HR system, though what was a triumph of *ex ante* AI auditing has often been unfortunately presented in the press as a failure. The biased system was never permitted to go live with that fault.

process in general. Similarly, there is no question that strong worker representation and consultation is excellent practice. This may well explain, for example, why AI deployment in Germany tends to increase workers' wages (Battisti & Ors 2023). However, this labour arrangement is part of Germany's sovereign law and applies to employer behaviour far beyond only introducing artificially intelligent systems. These are excellent regulations, but they do not belong in the AI Act.

Putting undue burdens in the AI Act encourages people to engage in regulatory avoidance by pretending that the systems they deploy are not "intelligent". This brings us back again to the question of the *definition of intelligence*, which is one of the most reworked pieces of the AI Act. We again advocate for the simplest, broadest definition possible. Really, all software should be subject to product law, which is largely the impact of the AI Act. The kinds of diligence we are asking for in high-risk systems should be applied to any system that might "decide" something life-changing, whether that system uses Excel macros, large language models, or steam-punk clockwork. If there is a possibility that a human might not be present at the point of the decision, then humans need to do due diligence on that decision-making *ex ante*, and humans need to be able to go back and ensure that a decision was made justly in retrospect.

The European Parliament's proposal goes on to require that deployers notify national supervisory authorities, consult relevant stakeholders and involve representatives of the affected people in providing input to such impact assessments. Such representatives could include but are not limited to, for example, equality bodies, consumer protection agencies, social partners and data protection agencies (article 29a(4)). While again we recognize the value of systematic stakeholder involvement in the deployment of AI systems and are generally in support of such processes, regulating such mechanisms as obligatory for every deployment is clearly overly demanding and provides the basis for serious complaints of over-regulation, and perhaps successful challenges in court.

Looking at the initial proposal, the European Commission has prioritized proportionality and sought to avoid over-regulation by several means. One such means is, for example, leaning towards options that rely predominantly on self-assessment processes rather than obliging providers to have independent audits. The Council, in its own compromise version, seems to build on these same premises (European Commission 2022b). However, the European Parliament is taking a major step in the opposite direction. The deployers of high-risk systems are not trusted to the same extent as providers, but instead, are required to undertake

obligatory consultations with stakeholders, and, further, to notify supervisory authorities. To some extent, this is sensible, since it is the specific application or deployment of AI that determines what harms are available and likely. However, where such considerations are sensible and truly specific to AI, it is important that the costs are kept proportionate to the potential harms. Where they are not specific to AI, it is important that they are in legislation appropriate to their sector. If such requirements are allowed to inflate the number of operators facing serious obligations under the Act, this would increase the costs enormously and decrease the probability of compliance with the Act. In order to properly evaluate the impacts of such a major change, it is essential that the impacts assessment of such amendments are reported. We would like to see all such language carefully reconsidered and reframed in a way that is most likely to benefit the European digital economy.

This same reasoning applies to providers of high-risk systems, who are also burdened with added accessibility and environmental reporting requirements. In line with our previous suggestion, we would prefer the scope of accessibility requirements to be controlled in the accessibility regulations to avoid misalignment of the scope of products and services that would face the accessibility requirements. Just as we do not want to motivate evasion of the label “AI”, we do not want to generate justifiable reasons for evasion of the label “high-risk”. We believe that the inclusion of AI should not be the defining factor for whether or not a system should be green or accessible. We affirm that there can literally be no greater concern than sustainability, and that human rights are rightly central to all EU legislation as well as agreed international law because they are essential to not only our ethics but our survival. Fundamental rights are why we are here: they are what it means to be here. We are a social species that cares for one another and can persist only in a vibrant ecosystem. But burdening some but not all products (annex IV(3b))—digital or otherwise—with these concerns begs regulatory evasion.

[D] WHAT IS A HIGH-RISK SYSTEM?

Moreover, perhaps the European Parliament’s most worrying change renders the question of what systems even need to comply with the most stringent regulatory burden highly debatable. The proposal introduces an extra layer of consideration when it comes to classifying a system as high-risk. The systems used in high-risk domains, as in annex III, would only be classified as high-risk if they pose a “significant risk of harm” to the health and safety or the fundamental rights of persons (or, in some cases, the environment). Practically, this would mean that a provider could—

based on their own assessment of the risk's severity, intensity, probability of occurrence and duration—object to the regulatory requirements by notifying a supervisory authority with a one-page notification letter and would be free to place their system on the market without such obligation. In response, if necessary, the supervisory authority could object to such a claim within three months of such notification.

We find this process extremely problematic for a variety of reasons. Firstly, it questions the rule-maker's capability to create credible categories of high-risk systems. Surely, the list of high-risk areas is based on a plethora of previous evidence of (and often research into) harm to health and the safety of fundamental rights to back up the classifications? Second, it opens considerable wiggle room for highly compromised interpretations of risk levels by providers who prefer not to comply. Finally, such an amendment would require the significant additional administrative capacity necessary for assessing these notifications. Taken in combination with our previous concern, it seems almost as if the European Parliament is seeking to make the Act a weapon to be used only in extreme circumstances, perhaps always *ex post*, where a very high regulatory burden can be asserted against only a small number of hand-picked companies. This is in complete opposition to the vision of an EU and digital economy full of safe, fair, trustworthy AI products. We strongly encourage sticking to a narrow enough but definite list of high-risk categories in order to avoid the harmful effects of such a new layer in the classification of an AI system as a high-risk system.

[E] NECESSARY CLARIFICATIONS ON FOUNDATION MODELS

On the positive side, the European Parliament takes both a clear and specific approach when it comes to ensuring the fair sharing of responsibilities along the AI value chain. The Council's approach suggests that general-purpose AI systems should be classified as high-risk if they could be used for high-risk use cases, which of course, if they are truly general-purpose, would encompass all such systems. Contrary to this, the European Parliament prefers an alternative approach: anyone who modifies a general-purpose AI system such that it becomes part of a high-risk AI system (for example, by fine-tuning a general-purpose model specifically for recruiting purposes) becomes the provider of the high-risk system. This makes sense, as we strongly believe that the severity of the limitations and risks of general-purpose systems are specific to their applications. General-purpose AI providers will still be motivated to

provide "hooks" or application programming interfaces to support high-risk systems' audits because those who make this process easiest will be the ones whose product deployers will choose to deploy.

The European Parliament's take on a special type of general-purpose AI, so-called *foundation models*, is particularly interesting. Its proposal defines foundation models as AI models that are trained on broad data at scale, designed for generality of output, and that can be adapted to a wide range of distinctive tasks (article 3(1c)). Furthermore, it rightly notes that such models can be "reused in countless downstream AI or general-purpose AI systems", and they "hold growing importance to many downstream applications and systems" (recital 60e).

While the Council of the EU has focused on regulating such models in the same way as high-risk systems, the European Parliament takes better account of their special nature. Its suggestion is threefold. First, providers of foundation models, including open-source models, would be obliged to comply with specific guardrails related to data governance, risk management, model evaluation, energy efficiency and quality management (article 28b). Secondly, providers of foundation models would be expected to provide "extensive technical documentation and intelligible instructions for use" to help downstream providers of high-risk systems comply with their regulatory obligations, and to register all of this in a central EU database (article 28b(2e, g)). Finally, as a special obligation for the providers of generative AI foundation models, they would be required to ensure safeguards against the generation of content in breach of existing laws and make publicly available a "sufficiently detailed summary of the use of training data protected under copyright law" (article 28b(4b-c)). We welcome this transparency-centred approach that is highly aligned with what we have previously suggested. We also applaud the European Parliament's suggestions for the development of capabilities for the benchmarking of foundation models, in collaboration between national and international metrology and benchmarking authorities (Haataja 2022; article 58a, article 15(1a)).

For the sake of clarity, contrary to claims by some earlier commentators (Technomancers.ai 2023), the proposal does not suggest any prohibitions or bans on foundation models, nor does it hold them accountable for their applications, provided they function correctly as specified in their documentation.

[F] SUPPORTING THE RIGHTS OF AFFECTED INDIVIDUALS AND GROUPS AND FURTHER PROHIBITIONS

An amendment that has been long awaited by many (including us) is the right of individual persons or groups of persons to lodge complaints of infringement of the AI Act to supervising authorities. The European Parliament wishes to protect the rights of affected individuals by granting them a right to request from a deployer a clear and meaningful explanation of the role of the AI system in the decision-making procedure, including the main parameters of the decision taken and the related input data (article 68c(1)). This approach to complaints processes seems to align closely with that of the General Data Protection Regulation 2016 (GDPR), though it includes a minor extension. The European Parliament adds the same right not only for individual persons affected but also for collectives of affected people. When coming to remedies, the proposal should be analysed in conjunction with the suggested AI Liability Directive, which (again, in our minds rightly) would take the role of ensuring fair remedy for any individuals harmed by AI systems (European Commission 2022a). We welcome these additions and believe the actionable recourse is what must become an increasingly important ingredient of good AI governance.

Throughout the legislative process, the unacceptable use cases (ie prohibitions) have divided opinions. As expected, the European Parliament is continuing the push for a full ban on biometric identification systems other than the ones used solely for biometric verification and authentication. To understand its logic, it is worth remembering that biometric data is considered sensitive personal data under the GDPR. Furthermore, the European Parliament is worried about the combination of potentially uncontrolled power of the deployers of AI-based biometric categorization systems with well-known biases of the same systems. It is understandable that the European Parliament is seeking a complete ban on AI-based biometric identification in publicly accessible spaces. In contrast to such potentially pervasive surveillance applications, biometrics used for verification and authentication are necessarily consensual. These are systems like passports where the document is matched to user-provided biometric information. Such narrow, consensual uses of biometrics are permitted. Some claim that banning wide-scale face recognition in public spaces disadvantages blind people who might not be able to use their devices to "see" their friends walking by, but so long as friends consent to sharing their photos, a blind person through such a device would be able recognize them but not strangers, just like anyone else. Other additions

to the suggested list of prohibitions are AI systems used for assessing a person's risk for committing criminal offences (predictive crime) and emotion recognition systems for law enforcement, border management, workplaces and in an education setting. The bans are in line with established failings and abuses seen in other countries, such as the "re-education" camps in China.

[G] CONCLUSION

Based on our assessment, it looks as if the European Parliament has taken some important steps forward, but some quite surprising and large steps back. These odd combinations of moves could make the trilateral negotiations dance an interesting one. We applaud the sensible division of responsibilities between deployers and general-purpose providers, the specific transparency requirements for the foundation models, and particularly the new mechanisms for supporting the actionable recourse by affected persons. The European Parliament's clear recognition that the severity of limitations and risks of AI systems can only fully be assessed and mitigated with a clear-use case in mind is essential to good governance, and the suggested clarifications on the roles of providers of foundations models, as well as the role of deployers, deserve positive remark. We believe these are suggestions worth fighting for in the coming trilateral negotiations. We hope our short assessment encourages further impact assessment though for various of the other suggestions, which in our opinion, go beyond necessary and, at worst, carry disproportionately regulative burdens—some for the entire AI ecosystem, others only on the providers of high-risk systems. Particularly concerning (almost incomprehensible) is the suggestion that the relatively light-weight regulatory burden proposed in the AI Act, which should help ensure due diligence, might only apply to sufficiently risky ("significant risk of harm") high-risk systems. This almost makes a joke of the long years of effort to ensure that all AI in the EU is responsibly deployed. Nothing should motivate more providers to position their systems as non-AI, or "mostly harmless". While the market desperately needs clarity, the European Parliament's suggestion for the extra layer in classifying systems as high-risk seems an antithesis, and potentially dangerous to all the good attempts to establish regulative clarity that the AI market truly needs.

About the authors

Joanna Joy Bryson *is a globally recognized leader and expert in intelligence broadly, including AI policy and AI ethics. Her original academic focus was the natural sciences, using artificial intelligence for scientific simulations*

of natural cognitive systems. Her present research focuses on the impact of technology on economies and human cooperation, transparency for and through AI systems, interference in democratic regulation and the future of labour, society and digital governance in general. She also consults frequently on policy. In 2020, she was nominated for the first cohort of experts for the Global Partnership of AI, chairing an AI Governance committee. She holds two degrees each in psychology (BA Chicago & MPhil Edinburgh) and AI (MSc Edinburgh & PhD MIT). From 2002-2019 she was Computer Science faculty at the University of Bath; she has also held postdoctoral, sabbatical, or visiting positions at Harvard, Oxford, Nottingham, Mannheim, Konrad Lorenz Institute for Evolution and Cognition Research, and Princeton Center for Information Technology Policy. She has been Professor of Ethics and Technology at Hertie School, Berlin, since February 2020.

Email: jjb@alum.mit.edu.

Meeri Haataja *is the CEO and Founder of Saidot, a company providing a leading enterprise SaaS platform and services for AI governance and transparency. She is also the chair of IEEE's AI Impact Use Cases and a former Ethics Lead of the Finnish AI Program. An affiliate at the Berkman Klein Center for Internet & Society at Harvard University, Haataja is also active in public communication concerning responsible AI, AI governance, data, innovation and entrepreneurship.*

Email: meeri@saidot.ai.

References

Battisti, Michele, Christian Dustmann & Uta Schönberg. "Technological and Organizational Change and the Careers of Workers," *Journal of the European Economic Association* jvad014 (2023).

Bryson, Joanna. "Europe is in Danger of Using the Wrong Definition of AI," Wired (2022).

Caliskan, Aylin, Joanna J Bryson & Arvind Narayanan. "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases," *Science* 356 (6334) (2017): 183-186.

European Commission. White Paper on Artificial Intelligence—A European Approach to Excellence and Trust. 19 February 2020.

Haataja, Meeri. "3 Ways to Tame ChatGPT," *Wired* (2022).

Haataja, Meeri & Joanna Bryson. "Reflections on the EU's AI Act and How We Could Make It Even Better," *TechREG™ Chronicle* (March 2022).

Technomancers.ai. "EU AI Act to Target US Open Source Software" (13 May 2023).

Legislation, Regulations and Rules

European Commission. Proposal for a Regulation of the European Parliament and of the Council on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council. 30 June 2021

European Commission. Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive). 28 September 2022a

European Commission. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts – General Approach 25 November 2022b

General Data Protection Regulation 2016