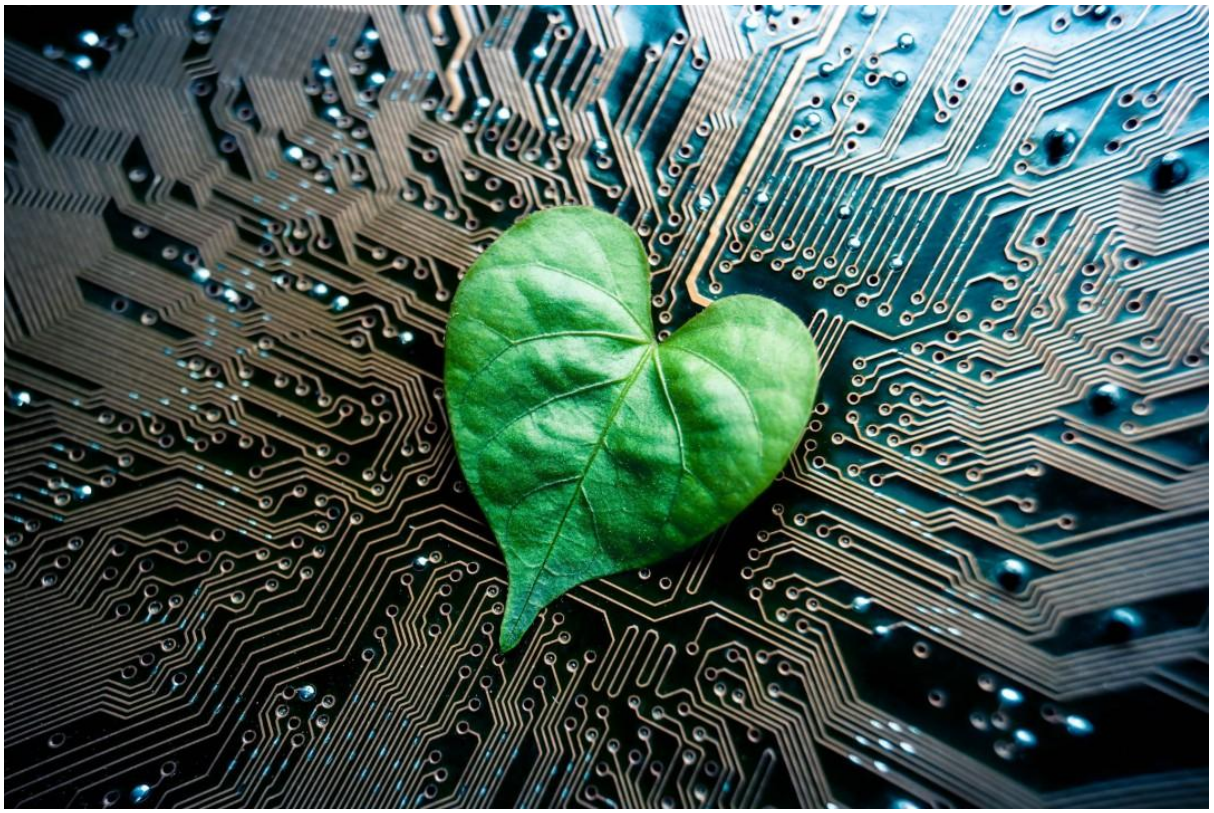
Does the EU AI Act promote socially sustainable AI?



Policy Brief and Project Report

by Isai Josue Canedo, Luis Martinez,
Chi Khánh Phan & Sophia Spornraft

July 2023

Table of contents

Policy Brief

Introduction

The EU AI Act is a pioneering legislative proposal to comprehensively regulate artificial intelligence (AI) systems and establish an unified framework across the European Union (EU) by being the first of its kind worldwide. It aims to ensure a responsible design, development and deployment of AI technologies and their safe and ethical use. In addition, the regulatory framework seeks to promote transparency and accountability as well as trustworthiness by safeguarding fundamental rights and ethical principles of AI within the EU using a risk-based approach (European Commission, 2021). Therefore, the EU AI Act is a central legislative piece of the European regulatory environment for Europe's digital and sustainable future.

Hence, the AI Act can play an important role in promoting and fostering social sustainability but only if fundamental human rights are protected and potential harms especially to vulnerable groups are prevented. Social sustainability thereby contains multiple concepts including and not limited to social justice, participation and co-creation, intergenerational justice, access to housing, goods and fair work as well as the quality of living (Colantonio, 2009). That is why despite the efforts of the European Parliament to safeguard human rights and prohibit AI practices that pose unacceptable risks to European citizens, a variety of civil society organizations have raised criticism of the current AI Act.

By taking their concerns seriously and tackling existing shortcomings and loopholes, the following policy recommendations aim to ensure a socially sustainable framework that ensures the protection of fundamental rights of European citizens by enhancing transparency, accountability and trust while allowing for responsible innovation in the development and deployment of AI systems.

The following policy recommendations aim to amend Recitals as well as Articles of the EU AI Act. At first, the main problem, potential risks and negative effects of each amendment are highlighted before the solution is proposed and the underlying rationale is explained in detail. Additionally, the positive benefit of the suggested change in wording of the legal text and its implications for social sustainability are described in-depth. At the end, a further outlook with regard to the upcoming trilogue negotiations between the European Commission, Parliament and Council is provided.

Amendment 1 - Recital 58 (a)

Recital 58a with Amendment 92	Our Amendment
Whilst risks related to AI systems can result from the way such systems are designed, risks can as well stem from how such AI systems are used. Deployers of high-risk AI system therefore play a critical role in ensuring that fundamental rights are protected, complementing the obligations of the provider when developing the AI system. Deployers are best placed to understand how the high-risk AI system will be used concretely and can therefore identify potential significant risks that were not foreseen in the development phase, due to a more precise knowledge of the context of use, the people or groups of people likely to be affected, including marginalised and vulnerable groups. Deployers should identify appropriate governance structures in that specific context of use, such as arrangements for human oversight, complaint-handling procedures and redress procedures, because choices in the governance structures can be instrumental in mitigating risks to fundamental rights in concrete use-cases. In order to efficiently ensure that fundamental rights are protected, the deployer of high-risk AI systems should therefore carry out a fundamental rights impact assessment prior to putting it into use. The impact assessment should be accompanied by a detailed plan describing the measures or tools that will help mitigating the risks to fundamental rights identified at the latest from the time of putting it into use. If such plan cannot be identified, the deployer should refrain from putting the system into use. When performing this impact assessment, the deployer should notify the national supervisory authority and, to the best extent possible relevant stakeholders as well as representatives of groups of persons likely to be affected by the AI system in order to collect relevant information which is deemed necessary to perform the impact assessment and are encouraged to make the summary of their	Whilst risks related to AI systems can result from the way such systems are designed, risks can as well stem from how such AI systems are used. Deployers of high-risk AI system therefore play a critical role in ensuring that fundamental rights are protected, complementing the obligations of the provider when developing the AI system. Deployers are best placed to understand how the high-risk AI system will be used concretely and can therefore identify potential significant risks that were not foreseen in the development phase, due to a more precise knowledge of the context of use, the people or groups of people likely to be affected, including marginalised and vulnerable groups. Deployers should identify appropriate governance structures in that specific context of use, such as arrangements for human oversight, complaint-handling procedures and redress procedures, because choices in the governance structures can be instrumental in mitigating risks to fundamental rights in concrete use-cases. In order to efficiently ensure that fundamental rights are protected, the deployer of high-risk AI systems **must** therefore carry out a fundamental rights impact assessment prior to putting it into use **by an independent party. It is further mandatory for** the impact assessment to **be** accompanied by a detailed **strategy** describing the measures **and** tools **to mitigate** the risks to fundamental rights identified at the latest from the time of putting it into use **and needs to be regularly updated**. If such plan cannot be identified, the deployer **must** refrain from putting the system into use. When performing this impact assessment, the deployer **must** notify the national supervisory authority and, to the best extent possible relevant stakeholders as well as representatives of groups of persons likely to be affected by the AI system **using easy language** in order to

fundamental rights impact assessment publicly available on their online website. This obligations should not apply to SMEs which, given the lack of resources, might find it difficult to perform such consultation. Nevertheless, they should also strive to involve such representatives when carrying out their fundamental rights impact assessment. In addition, given the potential impact and the need for democratic oversight and scrutiny, deployers of high-risk AI systems that are public authorities or Union institutions, bodies, offices and agencies, as well deployers who are undertakings designated as a gatekeeper under Regulation (EU) 2022/1925 should be required to register the use of any high- risk AI system in a public database. Other deployers may voluntarily register.	collect relevant information which is deemed necessary to perform the impact assessment and are **required** encouraged to make **a detailed** summary of their fundamental rights impact assessment publicly available on their online website. This obligations should not apply to SMEs which, given the lack of resources, might find it difficult to perform such consultation. Nevertheless, they are **highly encouraged** to also strive to involve such representatives when carrying out their fundamental rights impact assessment. In addition, given the potential impact and the need for democratic oversight and scrutiny, deployers of high-risk AI systems that are public authorities or Union institutions, bodies, offices and agencies, as well deployers who are undertakings designated as a gatekeeper under Regulation (EU) 2022/1925 are required to register the use of any high-risk AI system in a public database. Other deployers may voluntarily register.

Our proposed amendment for Recital 58 (a) is based on an identified shortcoming in the legislative AI Act. While we welcome the conduction of fundamental rights impact assessments of AI systems prior to their market release, it is highly problematic to conduct them by AI technology service providers themselves. As this self-check can result in a subjective judgment about the discriminatory effects of high-risk AI systems, it could lead to incentive providers to conceal potential problems arising with the technology, especially with regard to vulnerable groups. Therefore, in our opinion, a self-assessment on the impact on human rights does not sufficiently address the transparency about potential discriminatory effects of an AI system.

Hence, we suggest that providers of high-risk AI systems in Annex III are obligated to conduct ex-ante human rights impact assessments by an independent third-party organization prior to the market launch of the AI. In addition, a detailed strategy to mitigate the identified risks should be made mandatory and updated on a regular basis. Furthermore, AI system providers are obliged to inform national supervisory authorities and persons who are likely to be negatively affected by the AI system in easy language. The results of the assessments are required to be summarized and included in the documentation submitted to the EU database. While this requirement remains not binding

for small and medium-sized companies (SMEs), we highly encourage them to conduct ex-ante human rights impact assessments.

This amendment thus aims to protect the independence of fundamental human rights audits, by making them transparent to supervisory authorities, vulnerable groups that might be affected and the general public. By doing so the comparability of impact assessments through the transparency requirements of high-risk AI systems could further be enhanced. In addition, the requirement for easy language of the assessments enables citizens to comprehend the effects of the AI systems which is a prerequisite for taking action against discrimination in the first place. Finally, the requirement to transfer the impact assessment reports to the EU database makes the comparison of multiple AI systems and their effects also easier to research and assess patterns.

Amendment 2 - Recital 84

Recital 84	Our Amendment
The development of AI systems other than high-risk AI systems in accordance with the requirements of this Regulation may lead to a larger uptake of trustworthy, ***socially responsible and environmentally sustainable*** artificial intelligence in the Union. Providers of non-high-risk AI systems should be encouraged to create codes of conduct intended to foster the voluntary application of the mandatory requirements applicable to high-risk AI systems. Providers should also be encouraged to apply on a voluntary basis additional requirements related, for example, to environmental sustainability,accessibility to persons with disability, stakeholders' participation in the design and development of AI systems, and diversity of the development teams. The Commission may develop initiatives, including of a sectorial nature, to facilitate the lowering of technical barriers hindering cross-border exchange of data for AI development, including on data access infrastructure, semantic and technical interoperability of different types of data.	The development of AI systems other than high-risk AI systems in accordance with the requirements of this Regulation may lead to a larger uptake of trustworthy, *socially responsible and environmentally sustainable* artificial intelligence in the Union. Providers of non-high-risk AI systems should be encouraged to create codes of conduct intended to foster the voluntary application of the mandatory requirements applicable to high-risk AI systems. **Providers are strongly advised to actively consider and, where feasible, apply additional requirements on a voluntary basis, such as measures to enhance environmental sustainability, improve accessibility for persons with disabilities, and engage stakeholders in the design and development of AI systems. A deliberate effort to increase the diversity of development teams should be seen as a high priority, promoting varied perspectives and reducing the potential for bias. The Commission will regard the use of diverse development teams as a measure for eliminating or reducing risks through adequate design and development during the conformity assessment process.** The Commission may develop initiatives, including of a sectorial nature, to facilitate the lowering of technical barriers hindering cross-border exchange of data for AI development, including on data access infrastructure, semantic and technical interoperability of different types of data.

Our proposed amendment emerges from a recognized gap within the existing AI Act: the insufficient emphasis placed on promoting diversity within AI development teams. As it stands, the Act merely encourages providers to diversify their teams, a stance we believe is too passive and does not sufficiently address the integral role that diverse perspectives play in the development of equitable and unbiased AI systems.

Research consistently demonstrates that diverse teams contribute to more innovative solutions and reduced biases in AI systems, fostering a broader applicability and inclusivity in AI technology. This diversity encompasses not just gender, race, and ethnicity, but also different experiences, disciplines, and cognitive perspectives.

Currently, without a more assertive stance in the Act, many providers have little motivation to strive for diversity within their teams. This lack of motivation can perpetuate existing biases in AI systems, irrespective of whether the AI is deemed high-risk or not. The consequence is a potentially limited and biased AI system that fails to represent and serve the wider community effectively.

Our amendment aims to rectify this by urging providers, particularly larger organizations, to adopt concrete strategies to enhance diversity within their development teams. While we are not suggesting making this a mandatory requirement, a stronger push from the Act can incentivize organizations to commit to these diversity-enhancing practices.

This amendment also acts as an incentive for providers, since the Commission will regard the use of diverse development teams as a measure for eliminating or reducing risks through adequate design and development during the conformity assessment process. As a consequence, diverse teams will have higher chances to pass the conformity test described in the EU Act.

The ultimate goal is to foster a culture of inclusivity and social sustainability within AI practices, ensuring AI technology developed is representative, fair, and beneficial to all users. By doing so, we believe this amendment will make significant strides toward an AI landscape that is not only technologically advanced but also socially equitable and sustainable.

Amendment 3 - Article 15 (3)

Article 15 with Amendment 327	Our Amendment
High-risk AI systems that continue to learn after being placed on the market or put into service shall be developed in such a way to ensure that possibly biased outputs ***influencing*** input for future operations ('feedback loops') ***and malicious manipulation of inputs used in learning during operation*** are duly addressed with appropriate mitigation measures.	High-risk AI systems that continue to learn after being placed on the market or put into service shall be developed in such a way to ensure that possibly biased outputs ***influencing*** input for future operations ('feedback loops'), *malicious manipulation of inputs used in learning during operation and* ***biased inputs due to the origin, education, characteristics, behavior and preferences of the user group*** are duly addressed with appropriate mitigation measures.

Our proposed amendment seeks to address a critical issue: biased inputs due to user group characteristics, which can lead to skewed outputs and perpetuate biases within AI systems. This problem arises from the data-driven nature of AI, which continuously learns and adapts based on the input it receives from its users.

The current legislation does not sufficiently highlight or address this aspect. It does not draw enough attention to the fact that the demographic profile, behaviors, and other attributes of the user group can significantly influence an AI system's learning process.

For instance, an AI system primarily used by a specific demographic group will learn predominantly from the data reflecting the behaviors, preferences, and characteristics of that group. This learning may inadvertently exclude or not adequately represent the experiences and needs of users outside this demographic. Consequently, such AI systems could perform suboptimally or in a biased manner for these other users.

Our amendment proposes the incorporation of explicit measures to mitigate these potential sources of bias. These measures could include diversifying the user group, implementing techniques to detect and correct for bias in the input data, and establishing safeguards to ensure the system's performance is unbiased and effective for all users.

The ultimate objective is to foster a more inclusive AI landscape, where AI systems perform equitably and optimally for diverse user groups, instead of being inadvertently biased towards a specific demographic. We believe this amendment will make significant strides

towards AI technology that is fairer, more representative, and beneficial to all users, thereby aligning with the social sustainability goals of the AI Act.

Outlook

The future of AI systems based on democracy and social sustainability rests on the decisions we make today. Therefore, the above-outlined amendments aim to protect fundamental human rights and promote European values in the responsible design, development and deployment of AI. To maintain a regulatory framework that aligns with these values, it is crucial to foster open and transparent discussions with multiple stakeholders in the process. This will enable the AI framework to remain up-to-date and adaptable to new developments and societal needs while placing human rights at its core. As AI technologies continue to evolve and integrate into various aspects of society, continuous monitoring and evaluation become essential to assess their impact on social sustainability. Additionally, regular assessments of the effectiveness of the EU AI law will help to identify gaps or areas that require further improvement. This iterative approach to an European AI regulation ensures that the Act remains dynamic, responsive, and capable of addressing new challenges and opportunities that arise within the ever-evolving AI landscape. By upholding these principles, the EU can establish itself as a leader in socially sustainable AI development, promoting and preserving fundamental human rights and democratic principles for the long-term benefit of its citizens and society as a whole.

Project Report

As part of the sustAInability course, we were given the challenge with the question "Does the EU AI Act promote socially sustainable AI?" by the Deutsches Forum für Ethisches Maschinelles Entscheiden e.V. (EME). The forum brings together scientific experts from various disciplines, companies and politicians to shape ethical guidelines for dealing with machine-based decisions by upholding human norms and values thereby promoting the ethical use of artificial intelligence and machine learning (EME, 2023). Accordingly, the aim of the challenge was to recommend amendments with regard to social sustainability of AI systems to the latest legal version of the EU AI Act which was adopted by the European Parliament on 14th June 2023. Our policy recommendations are later on shared with Axel Voss (EVP), MdEP for Germany and shadow rapporteur for the EU AI Act in the Parliament.

On the first day of our challenge, we had an initial meeting with our supervisor Felix Rank who was a great support throughout the whole project week. During this meeting, he provided us with an overview of the EU AI Act before we discussed the main objectives and key results of the challenge and his expectations of the outcome. Once we set the goals of the challenge, he shared the key documents with us to start conducting our research which served as a basis for our policy recommendations.
Afterwards, our team started by assessing the initial proposal of the AI Act, the adopted amendments by the European Parliament as well as the critique by civil society organizations. We quickly realized two major challenges ahead. First, as all our team members were from different continents and had diverse educational backgrounds, we had a different level of understanding about the ordinary legislative procedure by which the European Union proceeded for the EU AI Act. In order to tackle this challenge and to get us all on the same page, we supported each other by giving us a crash course on the functioning of the EU's legislative process. This enabled each team member to better understand the current negotiated version of the AI Act and contextualize it in the bigger picture of the European Union's goals and values. Second, as the initial AI Act proposal alone had over 700 pages while there were several thousand amendments proposed by various stakeholders, and a limited time frame for the challenge as well as resource capacities of our team members, we needed to organize our team in the most optimum manner to conduct our research. Thus, we divided the responsibilities among us into four different working packages. While some of our team members searched for social sustainability key terms out of a predefined catalog within the adopted amendments by the Parliament to identify relevant paragraphs within the legal text, another team member concentrated on the initial proposal to review potential important chapters with regard to

social sustainability. Additionally, the last team member researched and screened the critique such as shortcomings and loopholes by civil society organizations and thought leaders who advocate for social sustainability values. We conducted this research strategy on the first and second day of the challenge in order to identify and select amendments which we would suggest to adjust and improve to promote social sustainability within the AI Act.
On the third day of the challenge, we came back together with our collected data and a shortlist of all potential amendments. In this next phase, we proceeded by collaboratively going through each amendment and explained our reasoning for choosing it whereby all other team members were invited to critically challenge the validity of the argumentation. In addition, we consulted Felix again to clarify our open questions and let our definition of social sustainability be challenged by our professors. This threefold pre-check enabled us to be confident in our final choice of selected amendments for the pitch and policy brief.

On day four, we gathered our preliminary amendments into a single document and opened a discussion about the key amendments that we would be presenting for the pitch night. As we were a diverse team where each of us grew up in different countries, we had culturally other values and therefore opinions on limiting the usage of AI systems in the high-risk category proposed in the EU AI Act. To solve this team challenge we consulted our seminar lecturers Helene and Charlotte who advised us to choose the amendments for the final pitch which all team members agree upon. Moreover, it was important for us to present amendments which our audience resonate the most with and which enable them to understand the implications and positive benefits of our policy recommendations in the most comprehensible way. Hence, we chose to present our amendment on Recital 84 to tackle the issue of non-diverse AI development teams which could perpetuate existing biases in AI systems by setting incentives to promote diverse AI developing teams to reduce or eliminate this risk. In addition, we chose to show our amendment on Recital 58a on the problem of assessing high-risk AI systems by providers themselves to avoid subjective judgment about discriminatory effects. Instead, we suggest conducting ex-ante assessments by independent third parties.
Before the final pitch night on day five, our team spent the majority of the work on building the pitch deck, dividing the pitch among us and practicing our individual part of the presentation. Following the project week, we summarized our complete list of amendments in a policy brief and provided it to our supervisor Felix.

We would like to thank Helene, Charlotte and Felix again for motivating our team, providing us with useful advice and encouraging us during the time of the seminar week!

Bibliography

Colantonio, A. (2009). Social sustainability: a review and critique of traditional versus emerging themes and assessment methods.

European Commission. (2021). *A European approach to artificial intelligence.* https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

Deutsches Forum für Ethisches Maschinelles Entscheiden e. V. (2023, June 26). Über das Forum. Deutsches Forum Für Ethisches Maschinelles Entscheiden. https://df-eme.org/das-forum