Technische Universität München

How does the EU AI Act promote socially sustainable AI?

Özge Demir
Silvia Cirolini
Fatma Ben Yahia
Marco Mata
Lisa Putter

July 2023

Contents

1

Introduction: The EU AI Act in a social sustainable framework

The EU AI Act represents a pivotal step in the European Union's efforts to establish ethical and trustworthy standards for artificial intelligence (AI) systems. It serves as a comprehensive regulatory framework designed to ensure the responsible development and deployment of AI technologies within the EU. Furthermore, it plays a crucial role in fostering social sustainability, as it protects individuals' rights, enhances transparency and accountability, and promotes the responsible use of AI across various sectors by addressing the potential risks and challenges posed by AI technologies. By establishing clear rules and guidelines, it seeks to safeguard fundamental rights, prevent harm, and promote trust and confidence in AI systems. Through its provisions, the Act aims to strike a balance between innovation and accountability, creating an environment where AI can flourish while ensuring the well-being and dignity of individuals.

However, despite the efforts to draft the EU AI Act, questions and concerns have been raised, even from within the European Parliament itself. Civil society organizations, comprising diverse stakeholders with a vested interest in AI's impact on society, have expressed their reservations and highlighted areas where they believe the Act falls short. In a joint statement, these civil society organizations have evaluated the draft report on the EU AI Act, emphasizing the need for further improvements and amendments. They aim to ensure that the Act adequately protects individuals and their fundamental rights in the face of fast-paced AI development and innovation.
The concerns raised by civil society organizations reflect the importance of continuous scrutiny and refinement to create a robust and effective regulatory framework. Through their statement, they call for amendments addressing issues such as the flexibility of the Act's risk-based approach, including biometric systems, and coverage of high-risk AI systems in migration management [1].

In addition to the concerns raised by civil society organizations, other voices express reservations about certain aspects of the EU AI Act. The undersigned organizations and individuals have highlighted key areas where they believe the original proposal needs to adequately address the harms and risks associated with AI systems in the migration context [2]. Their recommendations for amendments include prohibiting specific uses of AI systems, expanding the list of high-risk systems, ensuring the application of the AI Act to all relevant systems, and incorporating transparency and oversight measures. They emphasize the need to protect fundamental rights and ensure accountability for the impacts of AI on migration. While they did not explicitly refer to specific amendments or articles, their concerns encompass various aspects of the AI Act, calling for significant changes to enhance its effectiveness and safeguard the rights of all individuals, including those on the move.

The EU AI Act plays a vital role in promoting social sustainability by establishing ethical standards and regulations for AI systems. However, there is still an ongoing need for improvement in order to create an AI regulatory framework that fully addresses the challenges and ensures the well-being of individuals and society as a whole, which is why we are now going to fully focus on the social sustainability goals spoken of in the Article 5 where it mainly focuses on the prohibited artificial intelligence practices. By giving particular attention to these goals, we can work towards ensuring that AI technologies are developed and deployed in a manner that aligns with the principles of social sustainability and protects the rights and well-being of all individuals.

2

Challenge and Project Week Progress

For this course, we were assigned the challenge by the EME (Deutschen Forums für Ethisches Maschinelles Entscheiden), an NGO based in Munich founded by a group of experts from various disciplines to promote the ethical use of artificial intelligence. They provide the institutional framework for debates around AI and machine learning (ML). The goal of this challenge was to find strong and sound amendments to the latest version of the EU AI Act, especially related to social sustainability in AI systems. The EME will later share these findings and policy recommendations with Alex Voss, German MEP, and shadow rapporteur for the EU AI Act.

The constant feedback and guidance from Felix Rank (EME) was an important opportunity for the team. He proposed changes, gave useful advice, and provided an external and objective point of view regarding our work without forgetting to support and encourage us during the whole seminar week. The amendments elaborated during the seminar week with related justification can be requested through our supervisor Felix Rank.

During our initial meeting with Felix Rank, we opened a discussion about the project's main objectives. We deliberated on the best method to commence our work and identify the key focus areas for our report. After that, we started by delving into the extensive documentation, including the report of the amendments adopted by the European Parliament and the 2021 Proposal for the AI Act. However, as we realized the tremendous time commitment required to review each chapter and amendment thoroughly, we divided the chapters to read per person to select amendments that should be improved in order to include social sustainability.

On the second day of our collaboration, we recognized the challenges posed by the sheer volume of material and decided to pivot our strategy. We collectively decided to concentrate solely on the Amendments adopted by the European Parliament and made it our new method. Consequently, we established an objective of addressing four specific topics based on our selection procedure of the chapters and amendments and selected a subset of articles to analyze in depth instead of reviewing all the titles. These topics included Title II - Prohibited Artificial Intelligence Practices, Title V - Measures in Support of Innovation, Title VI - Governance, and Title VIII - Post-Market Monitoring, Information Sharing, and Market Surveillance.

On the third day, our focus shifted to meticulously reading through the titles and identifying the necessary modifications and key points that attracted our attention. For example, in Title II, which focuses on the Prohibited Artificial Intelligence Practices, we decided to highlight the lack of clarity regarding the areas of migration and asylum concerning the prohibition of AI usage in Article 5. Our progress on analyzing the various titles continued throughout Saturday before our main focus switched to preparing for the Sunday pitch night presentation.

Following the workshop week, we provided Felix Rank with a policy memo that can be utilized in upcoming discussions. Our list of amendments includes our social sustainability modifications to the latest EU AI Act amendments.

3

Prohibited artificial intelligence practices

The EU AI Act can encourage social sustainability by highlighting how AI systems affect migrants. Artificial intelligence (AI) systems are being used more frequently to control people in motion. To evaluate people's claims for immigration, AI is employed to produce predictions, assessments, and evaluations about them. it poses serious threats to persons who are crossing borders, applying for refugee status, or immigrating in terms of their human rights. Such programs violate the EU Charter of Fundamental Rights' ban on discrimination and are inherently discriminatory because they create profiles of people based on their nationality, ethnicity, and race. This is racial discrimination, as it is regarded in international human rights law.

Some AI systems used in migration technology include predictive analytics, biometrics, emotion detection, border enforcement, and data sharing. The EU AI Act divides AI application risk into four categories, ranging from minimal risk to unacceptable/prohibited risk. Even though there is the potential for irreparable harm, the proposed AI Act does not forbid some of the most destructive uses of AI in migration and border control. EU AI Act categorizes "migration, asylum, and border control management" as high-risk AI systems in Annex III, and policymakers recently added a new amendment in Annex III (Amendment 712) (new) stating that " AI systems intended to be used to make inferences about personal characteristics of natural persons based on biometric or biometrics-based data, including emotion recognition systems, with the exception of those mentioned in Article 5;" are listed as high-risk AI systems.

As seen from the amendment, except for those mentioned in Article 5, EU AI Act still categorizes emotion recognition systems as high-risk. Yet, there is still space for discrimination and biases while making inferences about persons' characteristics and emotions, especially for marginalized groups including refugees, immigrants, and asylum seekers. The exception to this is made in Article 5, and the European Parliament prohibited the placing on the market, putting into service, or using AI systems to infer a natural person's emotions in law enforcement, border management, workplace, and education institutions.

It is important to note that the EU AI Act is a complex legal framework, and its interpretation might vary. Border management and immigration/asylum are not specifically described as separate categories in this prohibition of AI usage. The control and monitoring of borders, including operations including law enforcement, security, and the flow of people across borders, is referred to as border management. It focuses on controlling how people enter and leave countries at borders. On the other hand, migration and asylum refer to the procedures and processes involved in the movement and settlement of people across borders, particularly those who are migrating or seeking international protection for various reasons. This covers issues including determining if a person is a refugee, requesting asylum, enforcing immigration laws, and making integration efforts.

While making inferences about personal characteristics of persons based on biometric data is generally called a high-risk AI application in the area of "migration, asylum, and border control management", the exception that is mentioned in Article 5 on the prohibition of emotional inference by using AI systems are only limited to "border management". Thus, there is a lack of clarity on the areas of migration and asylum. They should be included in the prohibited AI list as well.

With these changes made in Amendment 226, the goal is to draw attention to the great threat that artificial intelligence technologies have posed and will pose in the field of migration and to reduce it as much as possible. While Europe protects its borders thanks to artificial intelligence applications, it should prevent people in motion from being harmed by these applications.

4

Including biases in the sandbox environment

In a regulatory sandbox environment, it is crucial to ensure that A.I. systems do not unfairly discriminate against individuals based on their personal data. This means that sandboxes and A.I. development should actively work to avoid negative biases, errors, and lack of transparency during training. Failing to address these issues is unacceptable and poses a significant risk.

According to Yordanova [3], a sandbox is a secure, isolated space for testing and learning programs without damaging computer systems or resources. Similarly, a regulatory sandbox is an experimental environment for testing regulations. A regulatory sandbox is like a data science sandbox in the A.I. Act. Still, its purpose is to develop, test, and validate innovative A.I. systems for a limited time before they are put on the market or used according to a specific plan (A.I. Act, Art. 53(1)). However, specific techniques and systems that pose an unacceptable risk, such as subliminal techniques, exploitative systems, and social scoring systems that include negative biases, are strictly prohibited. A negative bias results in direct or indirect discrimination against an individual (A.I. Act, Recital 44). The AI Act requires immediate action to mitigate identified risks and holds all participating actors accountable for any harm caused to third parties during experimental phases.

Algorithms, systems, and programs may unintentionally contain negative biases, especially racial and gender biases, facial recognition biases, and criminal justice algorithm biases [4]. These biases cannot be avoided if there is insufficient training data. Therefore, the A.I. office should protect against potential negative biases in algorithms, companies, and other operators using A.I. Additionally, it is crucial to prevent biases in datasets that could lead to risks to fundamental rights or discriminatory outcomes for individuals affected by high-risk A.I. systems. Biases can be inherent in underlying datasets, introduced by algorithm developers, or generated during implementation in real-world settings [4]. Inherent biases can perpetuate and amplify discrimination, particularly for vulnerable or racialized communities. Therefore, it is crucial to ensure that under-represented groups are not systematically ignored when training algorithms [5]. Sandboxes and regulatory sandboxes can assist in training algorithms before they are released. Companies could use these sandboxes to prop the technology in a way that passes a testing phase or only share data that they want to share with the sandbox while excluding certain aspects, such as biased learning behavior [5]. However, it is essential to note that sandboxes do not necessarily replicate the complexity of society and reality. The E.U. A.I. Act Article 53-55 regulates algorithm training within sandboxes by establishing authorities and specific conditions. To further prevent negative biases during the learning process, amendments to the E.U. A.I. Act should be considered, including potential negative biases and safeguarding the public interest. Ultimately, the true impact of technology on society can only be measured once it is placed on the market.

To conclude, sandboxes may not replicate society's complexity, so amending the E.U. A.I. Act Article 53 to incorporate negative biases is necessary. Preventing biases in datasets and algorithms is essential to avoid discrimination and protect fundamental rights.

5 The AI Office

Enhance the European Artificial Intelligence Office (The 'AI Office')'s capacity to address and incorporate social, environmental, and economic considerations into its decision-making processes. By including legitimate social, environmental, and economic organizations as authorities within the Office, the legislation would be able to ensure that these critical perspectives were taken into account when evaluating and regulating Artificial Intelligence (AI) systems.

The initial EAIO focused primarily on technical aspects of AI, such as research, development, and standardization, which led it to be mainly staffed by experts in computer science, engineering, and related fields who focused on the technical aspects of AI systems. However, as the understanding of AI's impact on society, the environment, and the economy grows, it becomes apparent that a comprehensive approach is required to address the broader implications of AI. Recognizing that AI technologies have far-reaching social, environmental, and economic consequences, it is necessary to broaden the expertise and perspectives within the office.

These organizations, with expertise and knowledge in multidisciplinary areas such as sustainability, social impact assessment, environmental, ethical considerations, and economic implications, would allow for a more comprehensive understanding of AI technologies' potential benefits, risks, and trade-offs associated with AI technologies.

Overall, this amendment aims to promote responsible and sustainable AI development by granting legitimate social, environmental, and economic organizations a voice in the decision-making processes of the European Artificial Intelligence Office. By doing so, it allows for a more comprehensive understanding of policies and guidelines that align with societal values and promote sustainable development, seeking to create a more inclusive and balanced approach to AI governance that addresses the broader societal and environmental considerations associated with AI systems.

6

The potential shutdown of malfunctioning AI systems

Mandatory establishment of a robust mechanism to include potential shutdown or access revocation of high-risk AI systems in cases where they exhibit erroneous behavior or produce incorrect answers with significant implications, ensuring that any AI system operating in contexts where human rights are at stake can be subjected after critical errors or deviations from expected behavior to be taken down in order to mitigate potential harm. This requirement applies to instances where errors are encountered during the testing and verification phase of a specific version of an AI system that is already in operation or available on the market.

The absence of the amendment compromises patient safety, undermines the effectiveness of AI systems, and erodes trust in their reliability. By incorporating the proposed amendment, prompt reporting, investigation, and resolution of errors discovered in the lab environment can be enforced, preventing such harmful consequences in real-world scenarios. This amendment safeguards human rights and ensures compliance with the EU AI Act, promoting a responsible and secure AI ecosystem.

Errors detected in the lab can indicate potential flaws or risks that could manifest in harmful consequences when the AI system is deployed in operational settings. This proactive risk mitigation stems from the need to uphold trust and reliability in AI systems.

Including this amendment in the regulation is essential to prevent the potential negative consequences of risk behaviors detected solely in the lab environment. By requiring immediate reporting and potential shutdown of AI systems with identified errors, we can mitigate harm, protect human rights, and maintain the integrity and reliability of AI technologies in line with the EU AI Act.

7 Outlook

Moving forward, it is crucial to continue engaging with stakeholders, including civil society organizations, experts, policymakers, and industry representatives. The feedback and insights gained from their interactions will contribute to a more comprehensive understanding of the challenges and potential solutions related to social sustainability in AI systems. For this reason, it is important to foster open and transparent discussions to ensure that the regulatory framework remains up-to-date and adaptable to emerging technologies and societal needs. As AI technologies continue to evolve and become increasingly integrated into various aspects of society, ongoing monitoring and evaluation will be essential. Furthermore, regular assessments of the EU AI Act's impact and effectiveness will help identify any gaps or areas for further improvement. This iterative approach to regulation will ensure that the framework remains dynamic, responsive, and able to address new challenges and opportunities that arise in the AI landscape.

Bibliography

[1] Civil society. Civil society reacts to ep ai act draft report. *PDF Report*, November 2021. Available online: https://www.accessnow.org/wp-content/uploads/2022/05/Civil-society-reacts-to-EP-AI-Act-draft-report_FINAL.pdf.

[2] Accessnow. Joint statement: The eu ai act must protect people on the move. *Web Article*, December 2022. Online website: https://www.accessnow.org/press-release/joint-statement-ai-act-people-on-the-move/.

[3] Katerina Yordanova. The shifting sands of regulatory sandboxes for ai. *Web Article*, July 2019. Online article: https://www.law.kuleuven.be/citip/blog/the-shifting-sands-of-regulatory-sandboxes-for-ai/.

[4] Genie Barton Nicol Turner Lee, Paul Resnick. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. *Article*, May 2019. Online article: https://rb.gy/tv8g5.

[5] Dan McCarthy. To regulate ai, try playing in a sandbox. *Article*, May 2021. Online article: https://www.emergingtechbrew.com/stories/2021/05/26/regulate-ai-just-play-sandbox.