Fairness in the AI Act proposal

Pier Luigi Parcu

Natalia Menendez Gonzalez

2023 CDS Autumn Conference Fairness in the Digital Age

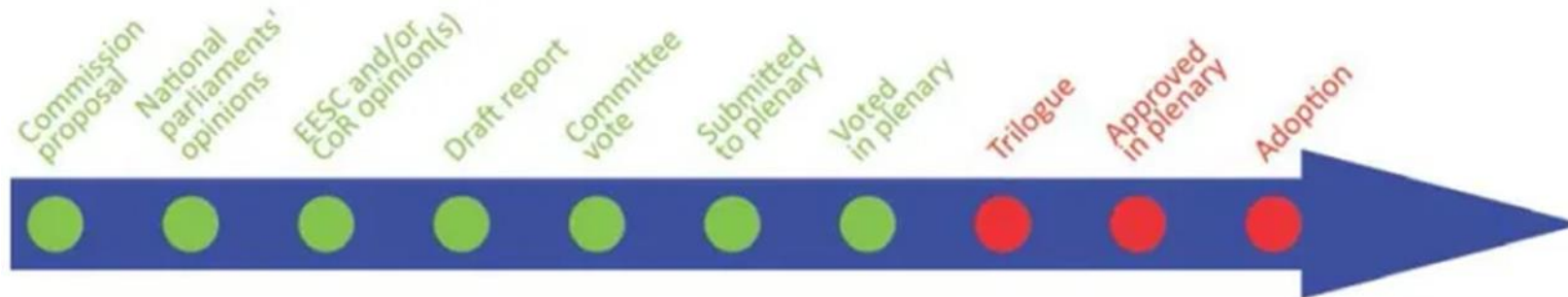European University Institute, 19th October 2023

www.eui.eu

Where are we at with the AI Act?

- 14 June 2023 – European Parliament adopted its negotiating position with 499 votes in favour, 28 against, and 93 abstentions.

- The Spanish Presidency aims to reach a deal by November with three trilogue meetings, but two have also been planned as a backup (POLITICO, Jun 2023).

- End 2023 / Beginning 2024 – Final text?

Commission proposal · National parliaments' opinions · EESC and/or CoR opinion(s) · Draft report · Committee vote · Submitted to plenary · Voted in plenary · Trilogue · Approved in plenary · Adoption

Source: European Parliamentary Research Service

Fairness in the AI Act

The word fairness is surprisingly absent from the original proposal of the AI Act (although concepts like (non)-discrimination and equality appear several time). However, after the institutional interaction, is clear that fairness has become one of the key aspects of the AI Act.

Fairness as a principle

Fairness in contractual terms

Fairness in procedures

Fair access to regulatory sandboxes

Fairness as a principle (I)

Recital 9 AI Act European Parliament (**negotiating position text)**

*"It is important to note that **AI systems should** make best efforts to **respect general principles** establishing a high-level framework that **promotes a coherent human-centric approach to ethical and trustworthy AI in line with the Charter** of Fundamental Rights of the European Union **and the values** on which the Union is founded, including the protection of fundamental rights, human agency and oversight, technical robustness and safety, privacy and data governance, transparency, nondiscrimination and **fairness** and societal and environmental wellbeing."*

Fairness as a principle (II)

Article 4 a AI Act European Parliament (**negotiating position text)**

General principles applicable to all AI systems

e) 'diversity, non-discrimination and
***fairness**' means that AI systems shall be*
developed and used in a way that includes
diverse actors and promotes equal access,
gender equality and cultural diversity,
*while **avoiding** discriminatory impacts*
*and **unfair biases** that are **prohibited by***
Union or national law

When to implement fairness?

Fairness by design throughout the AI lifecycle:

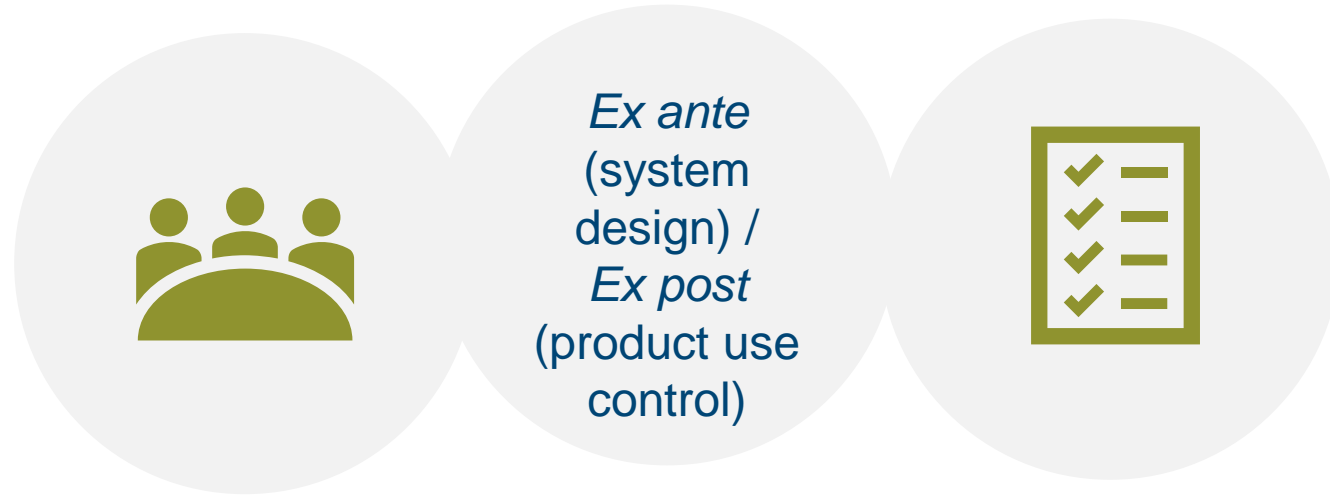| Training datasets (algorithmic bias)	Accuracy (false positives/negatives)	Feedback loops

Fair use of AI systems

Not to discriminate or oppress vulnerable people and minorities

How to implement fairness?

Stakeholder/public participation
(Impact assessments)

Ex ante (system design) / *Ex post* (product use control)

Quality Standards

Questions for the debate

- Is unfairness of AI a specific type of risk? Could enlarge the high risk categories?

- What are the differences between fairness, equality, and non-discrimination?

- How do we practically operationalize fairness for AI systems?

- Is it possible to achieve fair AI systems when those systems operate within an unfair world?

- What tools and enforcement mechanisms should the AI Act incorporate to ensure fairness within AI systems?

- What notion of fairness should be predominant when dealing with AI systems: ethical, mathematical, legal…?

- What are the main current obstacles to achieving fairness within AI systems?

- What solutions have AI providers put in place to ensure fairness within their AI products and services?

Audrey Scozzaro Ferrazzini

Qualcomm

Professor Daniel Innerarity

European University Institute

Professor Michal Gal

University of Haifa

J. Scott Marcus

Bruegel

The Panelists

Bibliography

- 'MEPs Ready to Negotiate First-Ever Rules for Safe and Transparent AI | News | European Parliament' (14 June 2023) <https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai> accessed 27 September 2023.

- 'AI Act Enters Final Phase of EU Legislative Process – EURACTIV.Com' <https://www.euractiv.com/section/artificial-intelligence Act/news/ai-act-enters-final-phase-of-eu-legislative-process/> accessed 27 September 2023.

- 'Artificial Intelligence Act [EU Legislation in Progress][Policy Podcast]' (Epthinktank, 18 November 2021) <https://epthinktank.eu/2021/11/18/artificial-intelligence-act-eu-legislation-in-progress/> accessed 27 September 2023.

- 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 COM/2021/206 final

- Council of the European Union Interinstitutional File: 2021/0106(COD) No. Cion doc.: 8115/21 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

- Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))

- Ondrej Bohdal, Timothy Hospedales, Philip Torr, Fazl Barez. Fairness in AI and Its Long-Term Implications on Society. *Proceedings of the Stanford Existential Risks Conference* 2023

- European Commission, Directorate-General for Communications Networks, Content and Technology, The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment, Publications Office, 2020, https://data.europa.eu/doi/10.2759/002360

- Anna Schmitz. 2023. Towards formalizing and assessing AI fairness. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES '23), August 08–10, 2023, Montréal, QC, Canada*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3600211.3604762

Thank you!

@PLParcu

linkedin.com/in/pier-luigi-parcu

Centre for a Digital Society
@EUI_DigitalSociety

Co-funded by the Erasmus+ Programme of the European Union

www.eui.eu