A Reflection on How Cross-Cultural Perspectives on the Ethics of Facial Analysis AI Can Inform EU Policymaking

Chiara Ullstein[1], Severin Engelmann[1], Orestis Papakyriakopoulos[2] and Jens Grossklags[1]

[1]*Technical University of Munich, Chair of Cyber Trust, Munich, Germany*
[2]*Sony AI, Zurich, Switzerland*

Abstract

The EU AI Act proposal addresses, among other applications, AI systems that enable facial classification and emotion recognition. As part of previous work, we have investigated how citizens deliberate about the validity of AI-based facial classifications in the advertisement and the hiring contexts (N = 3745). In our current research, we extend this investigation by collecting laypeople's ethical evaluations of facial analysis AI in Japan, Argentina, Kenya and the United States (N ~4000). Our project serves as a motivation to ask how such cross-cultural AI ethics perspectives can inform EU policymaking regarding AI systems, which enable facial classification and emotion recognition. We refer to suggestions on achieving policy impact and aim to discuss this topic space with workshop participants.

Keywords

Facial analysis AI, participatory AI ethics, EU policymaking, AI Act, cross-cultural perspectives in AI ethics

1. Contextualization of facial analysis AI and the EU's AI Act

In computer vision AI ethics, a key challenge is to determine how digital systems should classify human faces in images. In response, there has been an ongoing interdisciplinary scholarly debate about normative guidelines that inform policymaking for facial classification. A number of research papers, including research contributions from the FAccT community [e.g., 1, 2, 3], have highlighted ethical challenges with regard to computer vision AI inferences. At the same time, facial analysis AI techniques have become more pervasive and attempt to derive a broad range of characteristics from faces such as demographics, emotion expression as well as personality traits. With our empirical work, we study the public perception of facial analysis systems and aim to inform AI policymaking. In our discussion we focus, in particular, on the AI Act proposal [4], which addresses these systems and can, therefore, serve as a suitable policymaking use case.

EWAF'23: European Workshop on Algorithmic Fairness, June 07–09, 2023, Winterthur, Switzerland
chiara.ullstein@tum.de (C. Ullstein); severin.engelmann@tum.de (S. Engelmann); orestis.papakyriakopoulos@sony.com (O. Papakyriakopoulos); jens.grossklags@in.tum.de (J. Grossklags)
0000-0002-4834-4537 (C. Ullstein); 0000-0001-8368-4484 (S. Engelmann); 0000-0003-4680-0022 (O. Papakyriakopoulos); 0000-0003-1093-1282 (J. Grossklags)

In the amendments to the AI Act proposal, members of the European Parliament have criticized facial analysis AI pointing out its limited theoretical foundation [5]. Most recently, in May 2023, their lead committees on the AI Act adopted a draft of compromise amendments [6] for the European Commission's AI Act proposal [4]. The amendments propose to add to the list of prohibited AI practices the use of "[b]iometric categorisation systems using sensitive characteristics" and "[e]motion recognition systems in law enforcement, border management, workplace, and educational institutions" [7]. While gaps remain, these amendments respond to demands by civil society organizations that have been pushing for amendments to the AI Act proposal to better protect the fundamental rights of citizen [e.g., 8, 9]. In contrast, the Council of the European Union suggests an additional transparency obligation to inform natural persons when exposed to emotion recognition systems [10]. The members of the European Parliament are expected to vote on the amendments in June 2023 (marking the end of the Parliament's first reading [11]), before entering inter-institutional negotiations (trilogue) with the Council of the European Union and the European Commission to debate over the final details of the act [12].

2. Our findings on citizens' perceptions of reasonable inferences

Results from our research studies on the public's perception of the ethics of facial analysis AI *support* the amendments suggested by the European Parliament [6], in that the majority of participants in our studies reject the use of AI systems that categorize people based on inferred protected or sensitive attributes and emotion expression in high-risk application areas [13, 14]. Our previous work investigated how laypeople and AI competent people deliberate about the validity of AI-based facial classifications in the advertisement and the hiring application contexts [13, 14]. We found that US laypeople (N = 3745) reject facial AI inferences such as trustworthiness and likability in both contexts. In contrast, they show more agreement with inferences such as skin color or gender in the low-stakes advertisement than in the high-stakes hiring context. Analyzing 29,760 written justifications using the transformer-based language model roBERTa, we found that laypeople assess the ethical status of a facial AI inference based on whether they think faces warrant sufficient or insufficient evidence for an inference (evidentialists) or whether making the inference results in beneficial or detrimental outcomes (pragmatists) [13]. We found further support for these results in a second study with AI-competent participants from Germany [14]. Our studies quantified the normative complexity behind classifying human faces. We also found justificatory pitfalls that legitimized evidentially invalid facial AI classifications. These justifications reflected an over-confidence in AI capabilities, while others appealed to narratives of bias-free technological decision-making or cited the pragmatic benefits of facial analysis for decision-making contexts in advertisement or hiring. Thus, contrary to popular justifications for facial classification technologies among more technology-focused communities, these results suggest that there is no such thing as a "common sense" facial classification that accords simply with a general, homogeneous "human intuition."

Research on the ethics of AI commonly focuses on a largely WEIRD (Western, educated, industrialized, rich, and democratic) [15, 16] participant pool and pays little attention to voices from other cultures. In *ongoing* work, we add such missing cross-cultural perspectives and extend the research project to an analysis of laypeople's justifications of facial AI classification from

Japan, Argentina, Kenya, and the USA (N ~4000). Across all countries, survey participants were recruited from urban areas with more than half a million inhabitants to ensure comparability between the country samples. We aimed at understanding whether cultural commonalities and differences in the ethical evaluation of facial AI classifications exist. Moreover, we ran several qualitative analysis workshops with researchers from these countries and researchers who have lived in these countries to discuss cultural specifics and cross-cultural commonalities in the context of face perception and classifications. We believe our research can, in principle, support critical policymaking by documenting cross-cultural perceptions and judgments of computer vision AI classification projects with the goal of developing ethical digital systems that work in the public's interest.

3. Informing EU policymaking

With regard to the impact of our research, we pose the following question: *How can cross-cultural perspectives on the ethics of facial analysis AI inform EU policymaking?*

When we rolled out this research project in 2019 our main interest was to explore how citizens think and deliberate about AI that classifies faces in diverse contexts. However, we did not cocreate our research question with policymakers (see [17] for cocreation of research questions for a policy context) and conducted our research with less of an emphasis on policymaking. Post the research phase, if the uptake of evidence in policymaking was defined as a goal, the question arises of how research results could be meaningfully 'translated' in order to be useful evidence to EU policymakers.

To produce our research results, we deployed an embedded research design and analyzed data using mixed methods by integrating qualitative and quantitative data. Our mixed-methods studies, despite "only" consulting citizens and not involving them in a more participatory manner [18], are able to demonstrate the complexity of ethical justifications (qualitatively) as well as indicate preferences for rejecting facial analysis AI for certain inferences and application areas (quantitatively). Our approach is a relatively straightfoward method to obtain nuanced insights into the perceptions of citizens across geographic regions. As such, our approach can, in principle, provide policymakers quick access to citizens' perspectives and uncover a (de)legitimization of specific technologies and/or related policies by society.

Besides the value of high-quality evidence, achieving policy impact requires strategic planning. Based on learnings from their Science for Policy 2.0 model that overcomes the science-policy demarcation tradition, the European Commission's science service (JRC) formulated suggestions for policy impact planning: to engage relevant audiences with research results, both by participating in the policymaking environment and by gaining allies through the presentation of results in nonscientific audiences who spread messages; to build relationships and gather policy intelligence by understanding "drivers, actors and their different roles, the importance of timing and appropriate language, and the ways to increase visibility in policy and stakeholder circles" [19]. A similar summary of suggestions and related discussion is provided by academics [20]. While evidence on whether these How-Tos necessarily lead to policy impact is lacking and while How-Tos are not without limitations, it is well known that policymakers are on the one side confronted with competing information providers representing different

interests in a complex policy environment and on the other side have limited attention and capacities and are, as all humans, subject to biases [19, 20]. In turn, this highlights the necessity to speak the language that policymakers pay attention to and understand [19].

Taking up the question why perspectives from citizens outside the EU should be considered for EU policymaking, we would like to underline that facial classification AI is developed, operationalized and applied in research and commercial institutions around the world. Analyzing perspectives from other geographic regions or countries with different levels of technology adoption can show more comprehensively what impact the technology can have within the EU and why it is or is not accepted. This also means learning about discriminatory effects on specific demographic groups that have to be avoided and that may manifest without regulation. We believe the uptake of citizens' perspectives from outside the EU can enhance ethical considerations and contribute to inclusive policies that are considerate of the implications of technologies on diverse societies such as those that constitute the EU.

With regard to the current stage of the AI Act development and given the limited influence individuals usually have past the first reading of the ordinary legislative process of EU policymaking [11], we acknowledge that at a later stage of policymaking data is only welcomed if it is easy to integrate [19]. Given that our findings are in support of the European Parliament lead committee's amendments on biometric classification, we perceive it to be timely to provide our evidence as supporting material for policymakers. We recognize that in the area of biometrics, both classification and identification, NGOs and societal initiatives have had a leading role in the early public critique of the AI Act proposal – in most matters successfully. We reflect that building relationships with policymakers and the policy environment in general or communicating results takes time and resources that not all researchers might have. Therefore, we appreciate initiatives that support, first, opening the policy environment to researchers and, vice versa, opening the academic research environment to policymakers. Second, we appreciate initiatives that support researchers in acquiring relevant skills to communicate their research to relevant stakeholders beyond the academic sphere.

Acknowledgments

We deeply thank our collaborators Hiromi Yokoyama, Yuko Ikkatai, Naira Paola Arnez Jordan, Tilman Hartwig, Rose Caleno, and Brian Mboya for their invaluable time and effort on this project. We are also grateful to all participants that took part in our surveys.

References

[1] J. Buolamwini, T. Gebru, Gender shades: Intersectional accuracy disparities in commercial gender classification, in: Proceedings of the 2018 Conference on Fairness, Accountability and Transparency, 2018, pp. 77–91.

[2] M. Miceli, T. Yang, L. Naudts, M. Schuessler, D. Serbanescu, A. Hanna, Documenting computer vision datasets: An invitation to reflexive data practices, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021, pp. 161–172.

[3] M. K. Scheuerman, K. Wade, C. Lustig, J. R. Brubaker, How we've taught algorithms to see identity: Constructing race and gender in image databases for facial analysis, Proceedings of the ACM on Human-Computer Interaction 4 (2020) 1–35.
[4] European Commission, Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, 2021. URL: https://tinyurl.com/43eap7bz.
[5] P. Vitanov, B. Sippel, B. Vollath, T. Penkova, J. F. L. Aguilar, M. Grapini, B. Benifei, Amendment 460 recital 18b (new), in: B. Benifei, D. Tudorache (Eds.), European Parliament Amendments Draft Report, 2022, pp. 122–123.
[6] European Parliament, Draft Compromise Amendments on the Draft Report (version 1.1), 2023. URL: https://tinyurl.com/2rfwvw57.
[7] European Parliament, AI Act: A step closer to the first rules on Artificial Intelligence, 2023. URL: https://tinyurl.com/r9wxbj2d.
[8] European Digital Rights (EDRi), EU Parliament sends a global message to protect human rights from AI, 2023. URL: https://tinyurl.com/yc289rau.
[9] Access Now, Big wins, but gaps remain: European Parliament Committees vote to secure key rights protections in AI Act, 2023. URL: https://www.accessnow.org/press-release/european-parliament-committees-vote-ai-act/.
[10] Council of the EU, Artificial intelligence act: Council calls for promoting safe AI that respects fundamental rights, 2022. URL: https://tinyurl.com/52uetdsj.
[11] European Parliament, Ordinary legislative proposal, n.d. URL: https://www.europarl.europa.eu/infographic/legislative-procedure/index_en.html#.
[12] European Union Publication Office, Trilogue, n.d. URL: https://eur-lex.europa.eu/EN/legal-content/glossary/trilogue.html.
[13] S. Engelmann, C. Ullstein, O. Papakyriakopoulos, J. Grossklags, What people think AI should infer from faces, in: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022, pp. 128–141.
[14] C. Ullstein, S. Engelmann, O. Papakyriakopoulos, M. Hohendanner, J. Grossklags, AI-competent individuals and laypeople tend to oppose facial analysis AI, in: Equity and Access in Algorithms, Mechanisms, and Optimization, 2022, pp. 1–12.
[15] J. Henrich, S. J. Heine, A. Norenzayan, The weirdest people in the world?, Behavioral and brain sciences 33 (2010) 61–83.
[16] S. Linxen, C. Sturm, F. Brühlmann, V. Cassau, K. Opwis, K. Reinecke, How WEIRD is CHI?, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021.
[17] M. Sienkiewicz, From a policy problem to a research question: Getting it right together, in: V. Šucha, M. Sienkiewicz (Eds.), Science for Policy Handbook, Elsevier, 2020, pp. 52–61.
[18] A. Cornwall, Unpacking 'participation': Models, meanings and practices, Community Development Journal 43 (2008) 269–283. doi:10.1093/cdj/bsn010.
[19] M. Sienkiewicz, P. van Nes, M.-A. Deleglise, Achieving policy impact, in: V. Šucha, M. Sienkiewicz (Eds.), Science for Policy Handbook, Elsevier, 2020, pp. 44–51.
[20] P. Cairney, K. Oliver, How should academics engage in policymaking to achieve impact?, Political Studies Review 18 (2020) 228–244.