Mapping policymakers' and laypeople's perceptions of genAI and FPT

Chiara Ullstein*1†*, Michel Hohendanner*1,2†* and Jens Grossklags*1*

1Technical University of Munich, Chair of Cyber Trust, Germany

2HM Hochschule München University of Applied Sciences, Munich Center for Digital Sciences and AI, Germany

Abstract

The recently enforced EU AI Act has advanced the state of AI regulation and is largely perceived as a valuable step toward regulating AI. Nevertheless, some articles of the AI Act remain contested and raise societal concerns, highlighting that the AI Act's aim to support the development of trustworthy AI is a continuous endeavor, also in terms of discourse with society. Two highly discussed AI application areas with great impact on societies worldwide are facial processing technologies (FPT) and generative AI (genAI). For a socially sustainable regulatory approach, discourse between policymakers and citizens is important. This requires, on the one side, understanding policymakers' general opinions on citizen participation. On the other side, there is the need to know which touchpoints and perceptions laypeople worldwide currently have with and about these AI application areas. To learn about the perceptions of both target groups, we surveyed policymakers and experts (N = 61) and laypeople (N = 1070) worldwide in late 2023. Combining these two exploratory survey studies allowed us to identify key topics that are relevant to policymakers and citizens to inform policy processes in light of the EU AI Act and beyond. In the context of a larger research project, the results serve as a foundation for designing a citizen deliberation process on FPT and genAI across continents. In this short paper, we motivate and contextualize our research, present our research approach, and describe the first results.
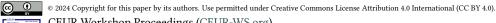
Keywords

facial processing technologies, generative AI, public perception, EU AI Act

1. Introduction

"*There are more questions than answers about how this technology will shape our environments and interactions, and policy is struggling to keep up with developments*" [1]. This assessment of generative AI (genAI) in a recent OECD report [1] highlights the technology's broad impact and the need for research to guide policymaking. Also intensively debated is the impact of facial processing technologies (FPT). Throughout the EU AI Act development phase, initiatives [e.g., 2, 3] pointed out risks such as being biased, or violating basic rights, and campaigned for the ban of facial recognition technologies (FRT). Others argued for safety benefits [e.g., 4].

Strong and partially competing interests by different stakeholders emphasize that at the core of sustainable regulatory approaches should be a continuous review of global societal perceptions of technologies. These can serve EU policymaking as important (warning) signals concerning regulatory blind spots, and explain how and why technology is (not) accepted or adopted. In light of the regulatory legitimization of certain FPT use cases and the rapid roll-out of genAI into various domains, laypeople's perceptions should be steadily studied as they are primarily affected by the adoption of these technologies. As such, spaces for discourse, as well as ongoing alignment checks of the interests of policymakers and citizens' needs, are crucial.

In this short paper, we present the design and initial results of two exploratory survey studies laying the groundwork for a citizen deliberation research project on genAI and FPT by identifying key topics that are relevant to citizens and policymakers. With this project, we aim to provide critical input for policymaking and the implementation of the EU AI Act.

2. Recent Discourse and Related Work

The abrupt rise of genAI has spurred extensive public discourse, prompting the publication of an open letter in March 2023 to pause giant AI experiments to "give society a chance to adapt" [5]. Governments and organizations, such as the OECD, released policy considerations in September 2023 [1], the G7 published the "Hiroshima Process International Guiding Principles for Advanced AI system" [6] in October 2023, and the European Parliament and the Council agreed upon the regulation of genAI through the EU AI Act in December 2023 [7], enforced in August 2024 [8]. Concerning FPT, the EU AI Act bans biometric categorization systems that use sensitive characteristics, untargeted scraping of facial images from the internet or CCTV footage, emotion recognition in work and educational contexts, and facial recognition in public spaces by law enforcement with exceptions [7]. Applications considered high-risk and requiring conformity assessment include facial recognition as a safety component of products, and all unprohibited biometric identification, biometric categorization, and emotion recognition systems [9]. Nevertheless, there has also been continued criticism [10, 11].

Prior work on potential benefits of genAI stated that ChatGPT can enhance customer service, automate repetitive tasks, enable 24/7 accessibility, aid educators and students, and serve as information and accountability tool within institutions [12]. Others highlight opportunities and threats across sectors and for culture and leisure [13]. For text-generating genAI, raised concerns include bias, negative impact on democratic processes, generation of false information and associated privacy concerns, and risk of job loss [12]. For text-to-image systems, further potential risks include discrimination and exclusion, misuse, and mis-/disinformation [14]. Prior work on public perceptions of genAI analyzing Twitter data found a generally positive sentiment across occupations, which correlates with exposure to AI [15]. Likewise, IT practitioners in the public sector indicated a high interest in and are optimistic about genAI, but are also concerned about emerging threats [16]. Raised concerns regarding the unethical use of artworks for training data played a major role in the negative sentiments of illustrators [15].

Prior work on FPT shows that particularly FRT exhibits bias, leading to a performance decline on facial data from children [17], women, and people of color [18], thereby reinforcing "existing racial disparities" [19]. To assess risks, researchers advocate for, e.g., a Human Rights

Impact Assessment and regular audits [20]. Prior work finds that public perceptions vary with the FRT application context [21]. Influencing factors include the control over facial data, the trustworthiness of the organization deploying FRT, the utility of FRT, and the surroundings and location of FRT use [21]. Participants weighed the security, usability, and economic gain of FRT with privacy risks [21]. We use these research findings and recent policy discussions as a foundation for our surveys to cross-nationally compare views on genAI and FPT.

3. Method

Policymakers and experts survey (S1). *Data collection:* In November and December 2023, we first surveyed policymakers and experts by sending out a total of 652 survey invitations. We received 61 responses from across the world. *Survey:* We asked policymakers and experts what topic (referring to genAI and FPT) they would like to hear citizens' opinions on, how relevant genAI and FPT are to their work, how knowledgeable they are, and how they perceive citizen participation. We concluded with closed questions describing their current role. The mean duration was 8.22 minutes. *Participants:* Respondents had 37 different nationalities, max. 3 participants from the same country (9 did not indicate). 20 respondents indicated to be female, 41 to be male. 3 participants stated to work at the European Parliament, 4 at the Council of Europe, 2 at the OECD, 27 at a national government, and 25 in academia, a research center, an NGO, or a private institution related to AI. *Limitations:* Due to the relatively low number of participants, frequency analyses should be interpreted with caution.

Citizen survey (S2). *Data collection:* In December 2023, we surveyed English-speaking laypeople across twelve countries and five continents via Prolific. The final dataset covers 1070 participants. *Survey:* Participants were randomly either drawn into the genAI or FPT technology context, where basic information on either genAI or FPT was provided. We then asked about perceived risks and benefits (open questions), and explored their opinions concerning several touchpoints with and trade-offs for specific use cases (closed questions). The trade-offs and use cases were selected based on prior research and policy discussions [e.g., 14, 22, 23, 24] and provided to participants in an informative manner, enabling them to learn and develop opinions about the technology in different contexts. Aligned with the EU AI Act, the FPT use cases covered high-risk and limited-risk AI systems. Finally, we asked about measures for safe and ethical technology use (open question) and who should introduce such measures or regulate the technology (closed question). The mean survey duration was 22.55 minutes. *Participants:* Participants' nationalities are Nigerian (N=110), South African (N=105), Indian (N=96), Japanese (N=24), South Korean (N=31), German (N=105), British (N=109), American (N=117), Canadian (N=106), Chilean (N=100), Mexican, (N=104), Brazilian (N=63). 526 respondents indicated to be female, 523 to be male, 14 to be non-binary, and 7 preferred not to say. *Limitations:* Participants self-selected via Prolific, potentially impacting generalizability. To counter potential biases, gender balance was ensured. The survey was rolled out in English, possibly attracting more higher-educated individuals.

Data analysis. S1: We applied manual content analysis [25, 26] to the open-text responses. S2: We performed automated topic modeling [27] refined by manual re-clustering on the open-text responses. S1/S2: We applied frequency analysis to multiple-choice and scale questions.

4. First Results and Final Remarks

In the first survey (S1), we find that the majority of policymakers and experts perceive genAI and FPT to be (somewhat) relevant to their fields. They generally perceive citizen participation to be valuable: 84% perceive citizen participation to "strengthen democratic institutions." While 80% think that citizen participation might give voices to those less frequently heard, 70% indicated that they may be able to learn from the public's opinions or judgments. Participants would like to hear citizens' opinions about the areas of regulation (what, who, how), perception of application contexts (individual use, public service, business contexts), awareness and informedness (technical knowledge, awareness, education, trust), and trade-offs and concerns (individual harms, transparency and accuracy, misuse, data security and protection).

In our second survey (S2), for both technologies, participants trust international institutions run by experts the most to establish measures that make the use of genAI and FPT safe. They believe that companies developing the technologies or those using the technologies, and for FPT also the government, should be held accountable if the technology leads to harmful outcomes. Participants indicate to have been most exposed to text-generating genAI use cases and least to audio genAI. The awareness of perceived exposure varies significantly for the presented FPT use cases, with participants being most aware of FPT for unlocking devices or verification at airports, and least with emotion recognition in work or educational contexts. When confronted with specific scenarios, participants across all countries find misinformation through genAI and misuse of facial data for FPT to pose great harm. Opinions are most diverse regarding the benefits and harms of genAI in the arts and the benefits of FPT for societal security. Participants perceive a value of genAI in the field of education and for increased personal efficiency in the workplace. At the end of the survey, participants reported being less excited and more nervous than at the beginning – an effect stronger for participants in the FPT context.

Comparing results from both surveys (S1 and S2), discussed topics with greater shared interest between experts and laypeople emerge: the need for regulation and specific regulatory strategies to ensure data security, privacy, and transparency. Also, assessing risks, benefits, and the overall impact on society, the job market and education are shared topics. Furthermore, ethical concerns stemming from dis-/misinformation and consequences of potential bias, and the need for raising awareness and educating laypeople are mentioned. One topic missing in citizens' responses but being of greater interest to experts is the use of genAI in public administration. Topics that are frequently mentioned by laypeople but rarely by experts are general misuse (aside from dis-/misinformation) and the need for responsibility and accountability for ethical use.

These *initial* results point to topics that are perceived by policymakers and citizens as subject to discussion. Other topics crystallize as being of exclusive interest to policymakers or citizens. We further analyze the data regarding national-specific differences in perception [28]. The results of this research project can help experts and policymakers to direct their attention to the issues that laypeople worldwide perceive to be most central, striving for technology development, deployment, and application in a trustworthy manner. We take these exploratory results as a basis for designing citizen deliberations on genAI and FPT taking place across multiple continents. Methodologically, these deliberations will also be informed by our previous work [29, 30, 31, 24]. We hope this project will fruitfully enhance the public discourse and contribute to a continuous exchange between the policy, academic, and citizen spheres.

Acknowledgments

References

[1] P. Lorenz, K. Perset, J. Berryhill, Initial policy considerations for generative artificial intelligence (2023). URL: https://www.oecd-ilibrary.org/content/paper/fae2d1e6-en. doi:10.1787/fae2d1e6-en.

[2] Reclaim Your Face, 2024. URL: https://reclaimyourface.eu/de/.

[3] Ban Facial Recognition, 2024. URL: https://www.banfacialrecognition.com/.

[4] South Wales Police, Deployments for live facial recognition, 2023. URL: https://www.south-wales.police.uk/police-forces/south-wales-police/areas/about-us/about-us/facial-recognition-technology/deployments-for-live-facial-recognition/.

[5] Future of Life Institute, Pause giant AI experiments: An open letter, 2023. URL: https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

[6] European Commission, Hiroshima process international guiding principles for advanced AI system, 2023. URL: https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system.

[7] European Parliament, Artificial Intelligence Act: Deal on comprehensive rules for trustworthy AI, 2023. URL: https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai.

[8] European Commission, AI Act enters into force, 2024. URL: https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_en.

[9] European Commission, AI Act, 2024. URL: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

[10] G. Volpicelli, EU set to allow draconian use of facial recognition tech, say lawmakers, Politico (2024). URL: https://www.politico.eu/article/eu-ai-facial-recognition-tech-act-late-tweaks-attack-civil-rights-key-lawmaker-hahn-warns/.

[11] S. Wachter, Limitations and loopholes in the EU AI Act and AI Liability Directives: What this means for the European Union, the United States, and beyond, Yale Journal of Law and Technology 26 (2024) 671–718. URL: https://law.yale.edu/sites/default/files/area/center/isp/documents/wachter_26yalejltech671.pdf.

[12] M. T. Baldassarre, D. Caivano, B. Fernandez Nieto, D. Gigante, A. Ragone, The social impact of generative AI: An analysis on ChatGPT, in: Proceedings of the 2023 ACM Conference on Information Technology for Social Good (GoodIT), ACM, 2023, pp. 363–373. doi:10.1145/3582515.3609555.

[13] A. Bahrini, M. Khamoshifar, H. Abbasimehr, R. J. Riggs, M. Esmaeili, R. M. Majdabadkohne, M. Pasehvar, ChatGPT: Applications, opportunities, and threats, in: 2023 Systems and Information Engineering Design Symposium (SIEDS), IEEE, 2023, pp. 274–279. doi:10.1109/SIEDS58326.2023.10137850.

[14] C. Bird, E. Ungless, A. Kasirzadeh, Typology of risks of generative text-to-image models, in: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES), ACM, 2023, pp. 396–410. doi:10.1145/3600211.3604722.

[15] K. Miyazaki, T. Murayama, T. Uchiba, J. An, H. Kwak, Public perception of generative AI on Twitter: An empirical study based on occupation and usage, EPJ Data Science 13 (2024). doi:10.1140/epjds/s13688-023-00445-y.

[16] L. Z. Knutsen, J. David Patón-Romero, J. E. Hannay, S. S. Tanilkan, A survey on the perception of opportunities and limitations of generative AI in the public sector, in: World Conference on Information Systems for Business Management, Springer, 2023, pp. 503–520. doi:10.1007/978-981-99-8349-0_40.

[17] N. Srinivas, K. Ricanek, D. Michalski, D. S. Bolme, M. King, Face recognition algorithm bias: Performance differences on images of children and adults, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 2269–2277. doi:10.1109/CVPRW.2019.00280.

[18] J. Buolamwini, T. Gebru, Gender shades: Intersectional accuracy disparities in commercial gender classification, in: Proceedings of the 1st Conference on Fairness, Accountability and Transparency, volume 81, PMLR, 2018, pp. 77–91. URL: https://proceedings.mlr.press/v81/buolamwini18a.html.

[19] F. Bacchini, L. Lorusso, Race, again: How face recognition technology reinforces racial discrimination, Journal of Information, Communication and Ethics in Society 17 (2019) 321–335. doi:10.1108/JICES-05-2018-0050.

[20] D. Almeida, K. Shmarko, E. Lomas, The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: A comparative analysis of US, EU, and UK regulatory frameworks, AI and Ethics 2 (2022) 377–387. doi:10.1007/s43681-021-00077-w.

[21] S. Seng, M. N. Al-Ameen, M. Wright, A first look into users' perceptions of facial recognition in the physical world, Computers & Security 105 (2021) 102227. doi:10.1016/j.cose.2021.102227.

[22] J. Shi, R. Jain, H. Doh, R. Suzuki, K. Ramani, An HCI-centric survey and taxonomy of Human-Generative-AI interactions, arXiv Working Paper 2310.07127, 2023. doi:10.48550/arXiv.2310.07127.

[23] J. Buolamwini, V. Ordóñez, J. Morgenstern, E. Learned-Miller, Facial recognition technologies: A primer, 2020. URL: https://circls.org/primers/facial-recognition-technologies-a-primer.

[24] C. Ullstein, J. Pfeiffer, M. Hohendanner, J. Grossklags, Mapping the stakeholder debate on facial recognition technologies: Review and stakeholder workshop, in: Companion Publication of the 2024 Conference on Computer Supported Cooperative Work and Social Computing (CSCW), ACM, 2024. doi:10.1145/3678884.3681887.

[25] P. Mayring, Qualitative content analysis: Theoretical foundation, basic procedures and software solution, 1st ed., SSOAR, Klagenfurth, Austria, 2014. URL: https://www.ssoar.

info/ssoar/handle/document/39517.

[26] J. Saldaña, The coding manual for qualitative researchers, 2nd ed., SAGE, Los Angeles, 2013.

[27] M. Grootendorst, BERTopic: Neural topic modeling with a class-based TF-IDF procedure, arXiv Working Paper 2203.05794, 2022. doi:`10.48550/arXiv.2203.05794`.

[28] C. Ullstein, M. Hohendanner, J. Grossklags, What people think about genAI and FPT, and what policymakers want to know laypeople's opinion on: A reality check, Working Paper, Technical University of Munich, 2024.

[29] M. Hohendanner, C. Ullstein, G. Socher, J. Grossklags, "Good and scary at the same time"—Exploring citizens' perceptions of a prospective metaverse, IEEE Pervasive Computing 23 (2024) 27–36. doi:`10.1109/MPRV.2024.3366112`.

[30] M. Hohendanner, C. Ullstein, D. Miyamoto, E. F. Huffman, G. Socher, J. Grossklags, H. Osawa, Metaverse perspectives from Japan: A participatory speculative design case study, Proceedings of the ACM on Human-Computer Interaction 8 (2024). doi:`https://doi.org/10.1145/3686939`.

[31] M. Hohendanner, C. Ullstein, Y. Buchmeier, J. Grossklags, Exploring the reflective space of AI narratives through speculative design in Japan and Germany, in: Proceedings of the 2023 ACM Conference on Information Technology for Social Good (GoodIT), ACM, 2023, pp. 351–362. doi:`10.1145/3582515.3609554`.