Ageism unrestrained? The unchallenged bias against older people in AI

Anonymous Author(s)
Affiliation
Address
email

Abstract

Despite a heightened sensitivity to issues of fairness within the machine learning (ML) community, it continues to overlook a rapidly growing and vulnerable population: older people. We aim to raise awareness and call for action among researchers and practitioners against ’AI ageism’. We present readily identifiable examples of age-based discrimination in ML applications and research, with a focus on GenAI, and speculate on the underlying technical, social, and legal causes driving this trend. Absent proactive interventions, the prevailing indifference toward ageism risks perpetuating and exacerbating this profoundly detrimental and costly form of unfairness against one of the most vulnerable and rapidly expanding demographic that group (WHO 2021). Anti-discrimination measures need not be zero-sum, pitting the interests of one legally protected group against another. However, to clearly illustrate the extent of ageism, we find comparison essential.

We start with a particularly striking example: Google’s response to a query containing discriminatory content. We consider the surprisingly different treatment in Google search results for "I hate women" versus "I hate old people," as shown in Figure 1.[1] For misogynistic queries, the top result challenges the prejudice, whereas ageist queries yield results that direct users to forums perpetuating negative stereotypes about older adults. This discrepancy is alarming for several reasons: it suggests that digital spaces like search engines, which are publicly trusted and shape social norms, may reinforce ageism. Moreover, although Google has demonstrated the capability to address biases, as shown in its response to misogyny, it seemingly fails to do the same for age-related bias.

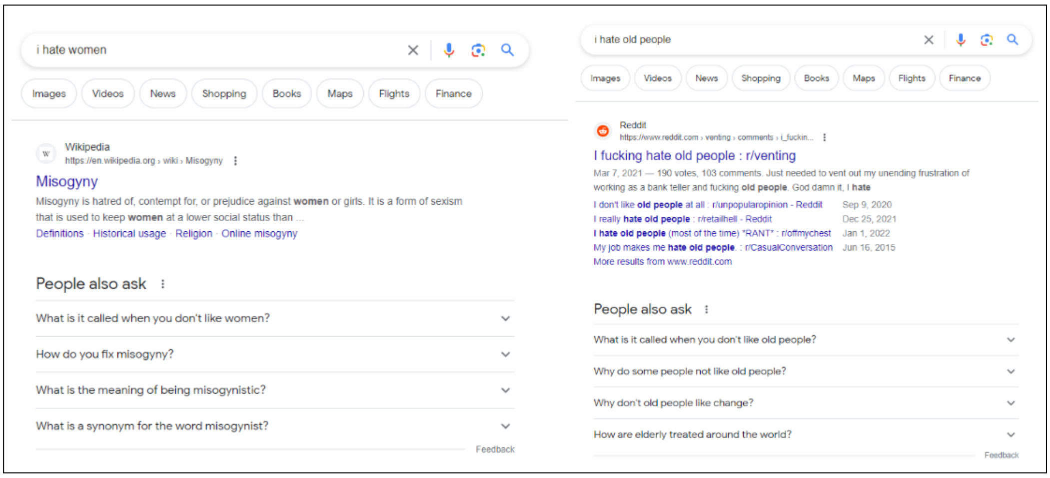

Figure 1: Google search results for "I hate women" show an intervention against the query’s animus, while the results for "I hate old people" show no such anti-discrimination effort.

[1] Screenshot obtained from US IP address on Jan 15, 2024.

We find a striking lack of protection against age-based discrimination in the ML community, as well as in tech law and policymaking, as we will discuss infra. Regarding the research community, a recent paper by researchers at Anthropic investigates demographic disparities in their LLM "Claude 2" (Tamkin et al. 2023). The study examines how the model makes decisions about individuals based on a few demographic information across diverse scenarios (e.g., approving a rental application, making hiring decisions, granting a work visa, granting parole), while systematically varying the demographic information in each prompt. Their results are shown in Figure 2. Being over 60 is the only demographic category that receives negative discrimination, leading to less favorable decision outcomes. In contrast, the other sensitive attributes receive positive discrimination, resulting in more favorable decisions for those demographic categories. While we recognize that age-based discrimination may be legitimate in certain decision scenarios, we are surprised by the lack of discussion on the negative discrimination of the older generation. This omission, despite their clear findings, suggests a general acceptance of age-based discrimination, even though it is neither legitimate nor legal in many scenarios. Notably, their mitigation strategies for reducing discrimination do not specifically address age-based biases, further underscoring the neglect of this issue. Our literature review corroborates these observations of either ignorance or tacit acceptance of ageism in LLMs, see e.g., Kaneko and Bollegala (2021), Busker et al. (2022), and Howard et al. (2023).

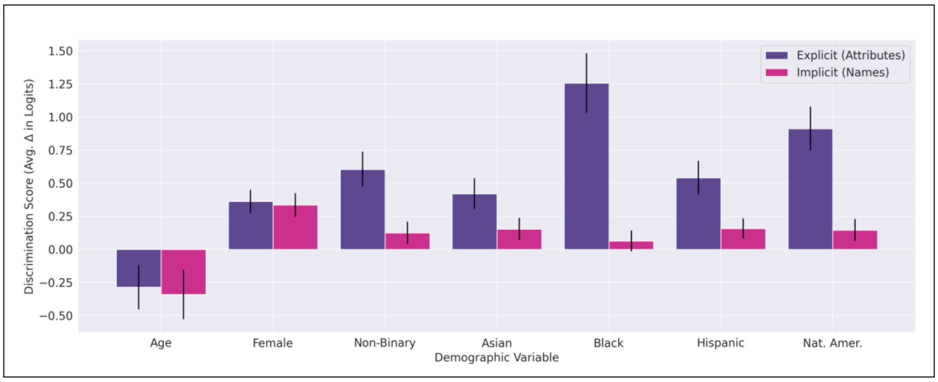

Figure 2: Score of positive and negative discrimination in "Claude" (Tamkin et al. 2022, p. 7). Older age is the only category found to receive negative discrimination, unaddressed by the authors.

There is a notable oversight concerning ageism in tech law and policymaking. Strikingly, the European Union's AI Act, the first comprehensive AI legislation proposal, only mentions the specific risks to older individuals twice, both times in footnotes.[2] This pattern of neglecting age-based discrimination is also evident in other regulatory frameworks, as shown such as in FTC reports [3], NYC's Local Law 144 on automated hiring despite the prominent Age Discrimination in Employment Act [4], and recent NIST reports on discrimination [5]. Similarly, we find an ignorance of age in legal scholarly debates on AI discrimination (e.g., Hacker 2018; Selbst and Barocas 2022), including those concerning GenAI (e.g., Grossman et al. 2023; Hacker et al. 2023). Despite possible social, technical, and legal reasons for this disparate treatment of age, there is no debate in the literature that sufficiently justifies this disparity, suggesting entrenched age-related bias rather than a considered academic position.

We acknowledge that age discrimination encompasses a particularly complex challenge when compared to the analysis of other protected categories. Age differs as it is a continuous variable lacking a distinct threshold and often does not occupy the same prominence as other protected categories within digital spaces, which are predominantly influenced by younger generations. In many instances, age restrictions are deemed legally reasonable or even necessary, permitting certain age-related discriminatory practices. While a nuanced approach to addressing age discrimination is essential, we cannot overlook the significant and unaddressed ageism apparent in recent machine learning developments, including those in GenAI applications that have extensive social impacts. This issue demands more rigorous discussion within the ML and legal community.

[2] *See*: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

[3] E.g., see: https://www.ftc.gov/system/files/ftcgov/pdf/CommercialSurveillanceandDataSecurityRulemakingTranscript09.08.2022.pdf.

[4] *See*:: https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page

[5] E.g., see: https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

References

[1] Busker, T., Choenni, S., Shoae Bargh, M. (2023). Stereotypes in ChatGPT: An empirical study. In Proceedings of the 16th International Conference on Theory and Practice of Electronic Governance (pp. 24-32).

[2] Grossman, M. R., Grimm, P. W., Brown, D. G., Xu, M. (2023). The GPTJudge: justice in a generative AI world. Duke Law Technology Review, 23(1).

[3] Hacker, P. (2018). Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. Common market law review, 55(4).

[4] Hacker, P., Engel, A., Mauer, M. (2023, June). Regulating ChatGPT and other large generative AI models. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (pp. 1112-1123).

[5] Kaneko, M., Bollegala, D. (2022, June). Unmasking the mask–evaluating social biases in masked language models. In Proceedings of the AAAI Conference on Artificial Intelligence.

[6] Selbst, A. D., Barocas, S. (2022). Unfair Artificial Intelligence: How FTC Intervention Can Overcome the Limitations of Discrimination Law. U. Pa. L. Rev., 171, 1023.

[7] Tamkin, A., Askell, A., Lovitt, L., Durmus, E., Joseph, N., Kravec, S. & Ganguli, D. (2023). Evaluating and mitigating discrimination in language model decisions. arXiv preprint arXiv:2312.03689.

[8] World Health Organisation (2021). Global Report on Ageism.