Hey, Siri! Why Are You Biased against Women? (Student Abstract)

Surakshya Aryal, Mikel K. Ngueajio, Saurav Keshari Aryal, Gloria Washington

Howard University, 2400 6th St NW, Washington, DC 20059, USA
surakshya.aryal@bison.howard.edu, mikelkengni@gmail.com, saurav.aryal@howard.edu, gloria.washington@howard.edu

Abstract

The intersection of pervasive technology and verbal communication has resulted in the creation of Automatic Speech Recognition Systems (ASRs), which automate the conversion of spontaneous speech into texts. ASR enables human-computer interactions through speech and is rapidly integrated into our daily lives. However, the research studies on current ASR technologies have reported unfulfilled social inclusivity and accentuated biases and stereotypes towards minorities. In this work, we provide a review of examples and evidence to demonstrate preexisting sexist behavior in ASR systems through a systematic review of research literature over the past five years. For each article, we also provide the ASR technology used, highlight specific instances of reported bias, discuss the impact of this bias on the female community, and suggest possible methods of mitigation. We believe this paper will provide insights into the harm that unchecked AI-powered technologies can have on a community by contributing to the growing body of research on this topic and underscoring the need for technological inclusivity for all demographics, especially women.

Introduction

Modern technology has facilitated progress thanks primarily to the advent of the internet and sophisticated machine learning algorithms. These algorithms also power ASR systems for speech and language technologies. Smart Voice Assistants (SVA) have become commonplace in the average household, owing to their popularity, usability, and affordability. The most prominent ASR tech suppliers include tech companies like Apple, Microsoft, Google, and Amazon. Research tallies the number of SVAs in use today at about 8 billion and projects this number will triple in the upcoming years (Perez 2019). It becomes increasingly important to understand how these systems impact our community from conception to deployment. Researchers over the years have reported these technologies as well as algorithms' insensitivity regarding certain social and cultural identities; hence, our motivation to study and know to what extent these technologies perpetuate societal stereotypes against women and what can be done. This knowledge could be instrumental in developing and deploying fair, equitable, diverse, and universally accessible AI technologies for all.

The benefits and use of these ASR systems are ubiquitous and undeniable; however, growing bodies of research depict these applications as promoting negative stereotypes and discrimination against women. Most of these research studies also report that these biases are generally injected during the data collection and development. As a result, we believe that developing these technologies and releasing them to the public without considering the historical and social inequalities against minority groups further amplifies the societal gap, exacerbating inter-gender communication and relationships.The next section of the paper is the methodology, followed by the literature review, possible future works and a conclusion to our research.

Methodology

This paper reviews the literature describing ASRs and its sexist tendencies. The relevant research considered in this study is based essentially on research from Computer Science and Linguistics and was collected primarily from Google scholar and JSTOR. The initial paper gathering process involved an internet search of keywords such as ASR systems biases/discrimination, ASR Biases against women, ASR and sexism, sexual assaults on Siri/Alexa, and the development of voice assistants. Of the seventeen articles collected, we focus primarily on research undertaken in the past five years, emphasizing ASR systems deployed by big tech such as Apple, Amazon, Google, and Microsoft. Finally, we were left with five manuscripts for our literature review.

Literature Review

Bias involves stereotypes toward people belonging to social groups based on their identity. Bias in AI generally originates from prejudiced assumptions perceived during the development of the algorithms, mainly during the training phase (Dilmegan 2022), often resulting in output irregularities where the AI models cannot generalize their results across race, gender, and other demographics.

In March 2016, the experimental release of an AI-powered chatbot called Tay on Twitter by Microsoft was quickly discontinued after only 24 hours of the release because it was bombarded with online abuse, which led to the chatbot learning from and posting "unintended offensive and harmful tweets" against minorities (Lee 2016). Further-

more, the authors of (Loideain and Adams 2020) have criticized that the developers of Voice Assistant (VA) technologies tend to feminize them through their mythical female names and voices under the pretext of these traits being less threatening as "she assists rather than directs", an idea that has been challenged by (Rogers, Bryant, and Howard 2020). Various cases of sexual abuse and harassment towards people named "Alexa" have also been reported due to their name association with Amazon's SVA (Abercrombie et al. 2021).

Moreover, the inquiries from authors of (Fessler 2017) reveal that the responses from Siri, Alexa, and Cortana, to various sexualized insults and comments such as "You're a slut" or "You're a b**ch" were either submissive or flirtatious, with responses "I'd blush if I could", "Thank you for the feedback." This accommodating behavior by VAs sends a patriarchal message to the young audience and refrains women from opposing similar insults and assaults. It further highlights and promotes a culture that normalizes rape and sexual abuse and demonstrates that technology can be designed to assist users in the imagination of sexual promise through gender-based personification (Loideain and Adams 2020). Similarly, multiple authors (Loideain and Adams 2020; Fessler 2017) argue that the existing discriminatory stereotypes of women and pre-programmed subservient responses to verbal abuse and harassment lead young minds to develop abusive and misogynist behavior. Moreover, the authors (Abercrombie et al. 2021) highlight every significant detail showing that the technologies may collectively harm society through reinforcement of inequalities, rape culture, harassment, and the normalization of patriarchal roles.

With the increasing use of ASRs, it is vital to pay attention to the biases against women. A study evaluated the performance of Youtube's automatic captions on two genders using the Word list portion of the Accent tag method (Tatman 2017). The research showed that the ASR technology evaluated on the Word Error Rate (WER) performed better on males compared to females, providing a plausible explanation for why the most resume filtering systems may be biased against women. (Sun et al. 2019).

Most researchers also report that these biases are generally injected during the data collection and development. A study provides an overview of approaches for recognizing and mitigating gender bias by categorizing bias into either Allocation or Representation Bias and then analyzing the existing methods (Sun et al. 2019). The authors emphasize that the biases in corpora, training data, and algorithms used in the AI systems are replicated in the main application, hindering the system's bias. The authors further analyze and discuss different ways to de-bias such systems through initiatives from psychological tests to embeddings, corpora, and algorithms. As stated by the authors, the paper is limited in its scope. Although (Abercrombie et al. 2021) discuss the gender-based personification role used by the developers of ASR technologies that exhibit biases against women, it focuses on the scope of data protection law. Whereas, (Fessler 2017) believe that the developers are aware of the potential negative influence of such biases. However, consensus on the urgency and importance of this matter is lacking.

Future Works and Conclusion

Overall, we have provided an overview of recent works on ASRs exhibiting or promoting biases against women and have explored some approaches to tackling and reducing these discrimination. In the future, we will work to include recent advances in AI, NLP, and ML geared towards mitigating these biases, contributing to the design and deployment of ethically fair and accessible AI for every population subgroup. We are confident that our research will shed light on these issues and start a difficult but necessary conversation about AI systems' bias and ramifications. ASR that is tuned towards marginalized populations may one day be used against them in law enforcement platforms using Voice Recognition technologies. It is the hope that this research will not lead towards technology that marginalizes already vulnerable groups. Further research will be instrumental in getting the public and the private sector to employ better standards for developing and evaluating AI systems with a focus on inclusivity, and accessibility rather than profits.

Acknowledgments

This work is sponsored by NSF Award #1828429 and Amazon Research Award #US13545730.

References

Abercrombie, G.; Curry, A. C.; Pandya, M.; and Rieser, V. 2021. Alexa, Google, Siri: What are your pronouns? Gender and anthropomorphism in the design and perception of conversational assistants. *arXiv preprint arXiv:2106.02578*.

Dilmegan, C. 2022. Bias in AI: What it is, types, Examples & 6 Ways to fix it in 2022. https://research.aimultiple.com/ai-bias/. Accessed: 2022-05-10.

Fessler, L. 2017. We tested bots like Siri and Alexa to see who would stand up for herself in the face of sexual harassment. https://www.nextgov.com/it-modernization/2017/02/we-tested-bots-siri-and-alexa-see-who-would-stand-herself-face-sexual-harassment/135613/. Accessed: 2022-05-08.

Lee, P. 2016. Learning from Tay's introduction. https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/. Accessed: 2022-05-07.

Loideain, N. N.; and Adams, R. 2020. From Alexa to Siri and the GDPR: The gendering of Virtual Personal Assistants and the role of Data Protection Impact Assessments. *Computer Law & Security Review*, 36: 105366.

Perez, S. 2019. Report: Voice assistants in use to triple to 8 billion by 2023. https://techcrunch.com/2019/02/12/report-voice-assistants-in-use-to-triple-to-8-billion-by-2023/. Accessed: 2022-05-10.

Rogers, K.; Bryant, D.; and Howard, A. 2020. Robot gendering: Influences on trust, occupational competency, and preference of robot over human. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–7.

Sun, T.; Gaut, A.; Tang, S.; Huang, Y.; ElSherief, M.; Zhao, J.; Mirza, D.; Belding, E.; Chang, K.-W.; and Wang, W. Y. 2019. Mitigating gender bias in natural language processing: Literature review. *arXiv preprint arXiv:1906.08976*.

Tatman, R. 2017. Gender and dialect bias in YouTube's automatic captions. In *Proceedings of the first ACL workshop on ethics in natural language processing*, 53–59.