Gianna Pesce

In 2020, *Scientific American* published an article discussing concerns about modern voice assistants like Siri and Alexa: “These systems frequently misunderstand people with non-native accents, regional dialects, and atypical speech patterns” (Lopez Lloreda, 2020). Voice recognition technology is meant to make life easier, yet for millions of users around the world, it does the opposite. From misheard words to constant repetition, these frustrations show a larger issue—bias in artificial intelligence. As voice assistants have become a major part of everyday life, it is important to question who they actually work best for. This paper explores how voice recognition systems like Alexa and Siri show bias against accents, dialects, and non-native English speakers. These biases limit accessibility, worsen communication barriers, and highlight the urgent need for more inclusive technology.

Bias in speech recognition is not just an inconvenience; it also produces social inequalities. The problem begins with who builds and trains these systems. Developers often rely on datasets made up primarily of voices from white, native English speakers. As a result, speech recognition systems learn to understand those accents best, while struggling with others. As Lopez Lloreda (2020) explains, “The technology mirrors social inequalities—it works best for the people it was trained on.” This bias creates barriers for people with different accents or speech patterns, making them feel excluded and frustrated by the very technology designed to assist them.

Scientific evidence confirms that bias in voice recognition is not just anecdotal; it is measurable. In a 2020 study published in the *Proceedings of the National Academy of Sciences*, researchers compared five major speech recognition systems and found “the average word error rate for Black speakers was 0.35, compared to 0.19 for White speakers” (Koenecke et al., 2020). These numbers show a clear disparity. When one group is misunderstood nearly twice as often as another, it points to a serious flaw in design. The same study concluded that “these disparities arise from differences in the acoustic models, which likely stem from underrepresentation of Black voices in the training data” (Koenecke et al., 2020).

While this study focuses on race and dialect within native English speakers, it illustrates a key point: when technology performs worse for certain speech patterns, it reveals systemic bias in the algorithms themselves. If voice recognition struggles with dialects within English, the problem likely worsens for non-native speakers. Everyday users often notice this when Siri or Alexa misinterprets them or fails to respond altogether. These small moments of frustration create a lot of cultural exclusion.

Another study by Hutiri and Ding (2022) supports these findings by explaining that “bias can be introduced at every stage of the speaker recognition pipeline—from data collection to model training and evaluation.” This means the problem does not come from a single mistake, it’s from the entire process altogether. The researchers also found that “women and non-U.S. English speakers are particularly disadvantaged by current speaker recognition systems” (Hutiri & Ding, 2022). These patterns reinforce that bias is systemic, not random, and that it disproportionately affects groups already underrepresented in technology development.

The roots of voice recognition bias begin in the data used to train these systems. Artificial intelligence learns by analyzing massive amounts of audio data and matching speech to text. If those datasets include mostly one type of accent or dialect, the model naturally becomes better at understanding that group. Hutiri and Ding’s (2022) research emphasizes this cycle of bias, explaining that data bias “can be introduced at every stage.” In other words, when the data is not diverse, the technology cannot perform equitably.

Another factor is the dominance of English in technological development. Most major voice systems are designed and tested primarily for American or British English, leaving little room for variation. Even within English-speaking countries, regional dialects like Southern or Boston accents often cause recognition errors. This problem becomes more severe for non-native English speakers who might pronounce words differently. When a system cannot adapt to these differences, users are forced to adapt their own voices instead—something that should not be necessary for inclusive technology.

The lack of diversity among developers and researchers contributes to this issue. Many technology teams are not globally representative, and cultural differences can be neglected during the design process. Developers may not realize how their systems perform for non-native or accented speakers until users report issues after release. At that point nothing can really be done to change that released model.

Beyond the data, the human effects of speech recognition bias are significant. In a 2021 study from *Frontiers in Artificial Intelligence*, African American participants reported that they often had to change their speech to be understood by their devices. As the authors note, "Participants frequently reported that they had to change their speech—slowing down or 'talking white'—to get devices to understand them" (Mengesha et al., 2021). This finding exposes how technology can pressure users to conform linguistically, which causes harmful stereotypes about "acceptable" speech.

These interactions create emotional and cultural consequences. Participants described "frustration and a sense that the technology was not made for people like them" (Mengesha et al., 2021). When users feel upset by the tools they use daily, it affects their relationship with technology as a whole. This frustration is especially harmful for people who rely on voice assistants for accessibility, such as individuals with disabilities or those using them to assist with language learning. Bias in these systems limits the use of technological benefits.

This sense of exclusion also limits adoption of new technologies. If users cannot rely on accurate responses, they are less likely to use or trust these devices. That mistrust widens the digital divide, leaving underrepresented groups even further behind in the growing world of AI-driven innovation. Technology that should serve everyone equally instead ends up reinforcing who gets to benefit most.

Although the problem of bias in voice recognition is serious, researchers are finding ways to fix it. A promising study by Radzikowski et al. (2021) explored using neural style-transfer methods to improve how systems understand non-native accents. Their findings were encouraging: "Our accent-modification framework reduced word error rates for Japanese speakers by over 30% compared with the baseline model" (Radzikowski et al., 2021). This improvement shows that inclusivity in AI is possible when developers prioritize diverse training data and design.

These findings suggest that accent adaptation and inclusive training models can significantly reduce error rates. As Radzikowski et al. (2021) explain, "Accent adaptation can significantly improve ASR accuracy for non-native English speakers." This means companies like Apple, Amazon, and Google have a clear path forward—they simply need to commit to designing with diversity in mind. Expanding the datasets used to train AI models, including voices from different regions and linguistic backgrounds, can make technology truly accessible.

Furthermore, companies should adopt ethical AI practices that emphasize fairness and transparency. Independent audits of training data, community feedback from diverse users, and multilingual testing should all be standard steps in the development of voice recognition systems. Inclusivity should not be an afterthought but a foundational design principle.

Some critics argue that complete fairness in voice recognition may be impossible because language is too diverse. They suggest that systems cannot perfectly recognize every accent or dialect without becoming overly complex or inaccurate. Others claim that users can simply adjust how they speak to the device, viewing it as a minor inconvenience rather than discrimination.

However, these arguments overlook the ethical responsibility of technology companies. Expecting users to change their speech reinforces inequality and sends a message that some voices matter more than others. Moreover, research shows that improvement is achievable. Reducing bias by 30% or more demonstrates that inclusivity can coexist with accuracy. While total perfection may not be attainable, striving toward equity is both realistic and necessary.

Voice recognition systems like Siri and Alexa have transformed how people interact with technology, but they also reveal how bias can quietly shape those interactions. Studies show clear disparities in recognition accuracy across accents, dialects, and languages. These inequalities arise from limited training data, cultural bias in development, and a lack of inclusivity in design. As a result, millions of users experience daily frustration, exclusion, and the pressure to alter their natural speech to be understood.

The solution lies not in asking people to adapt to technology, but in demanding that technology adapt to people. More diverse datasets, inclusive design principles, and accountability in AI development can help ensure that all voices—no matter the accent or language—are recognized equally. As AI becomes more integrated into society, inclusivity is not just an ethical choice; it is a technological necessity. Voice recognition should empower communication, not restrict it. The next generation of AI should be built to listen to everyone.

References

Hutiri, W. T., & Ding, A. Y. (2022). *Bias in automated speaker recognition.* In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)* (pp. 106–116). ACM. https://doi.org/10.1145/3531146.3533089

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). *Racial disparities in automated speech recognition. Proceedings of the National Academy of Sciences, 117*(14), 7684–7689. https://doi.org/10.1073/pnas.1915768117

Lopez Lloreda, C. (2020, July 5). *Speech recognition tech is yet another example of bias. Scientific American.*https://www.scientificamerican.com/article/speech-recognition-tech-is-yet-another-example-of-bias/

Mengesha, Z., Heldreth, C., Lahav, M., Sublewski, J., & Tuennerman, E. (2021). "I don't think these devices are very culturally sensitive": Impact of automated speech recognition errors on African Americans. *Frontiers in Artificial Intelligence, 4.* https://doi.org/10.3389/frai.2021.725911

Radzikowski, K., Wang, L., Yoshie, O., & Nowak, R. (2021). *Accent modification for speech recognition of non-native speakers using neural style transfer. EURASIP Journal on Audio, Speech, and Music Processing, 2021*(1), Article 11. https://doi.org/10.1186/s13636-021-00199-3