



# Preventing Bias in Speech Technologies

September 2021

# Introduction

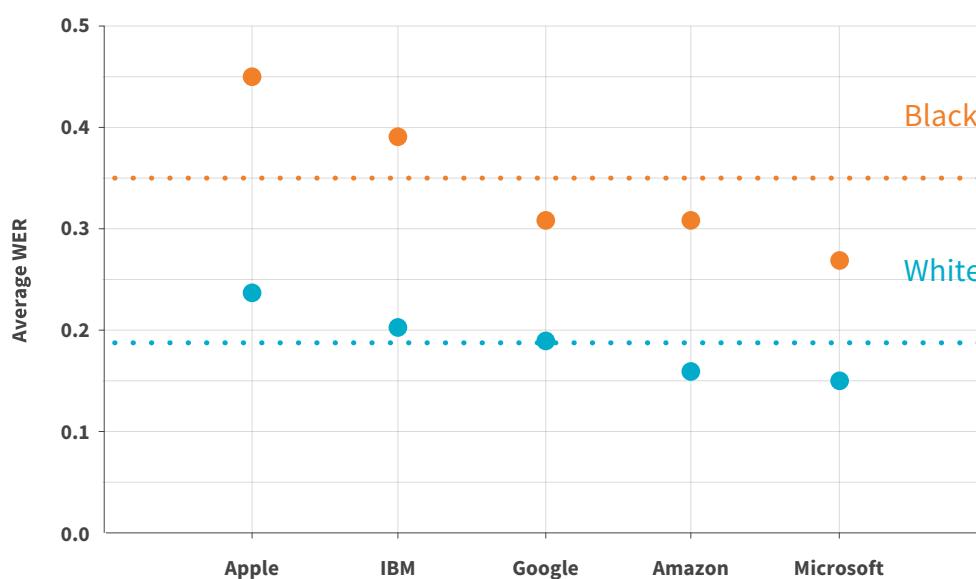
## Siri, Alexa, Google Assistant, and other Speech Recognition Systems can have trouble understanding the accents and speech patterns of people from minority groups.

As speech recognition technology has gained traction over the last few decades, one prevalent problem has emerged: linguistic bias, which in its most pernicious forms constitute a type of racial bias. A [recent study by PNAS](#) demonstrates “large racial disparities in the performance of five popular commercial ASR systems,” meaning that speech recognition technologies are far less useful to some people than they are to others.

Of course, speech recognition isn’t alone in this growing issue of algorithmic bias, but it is notable given the proliferation of technologies that rely on parsing human language in order to provide services.

Bias becomes of particular concern as companies, spurred on by both the adoption of cutting edge tech and the realities of doing business [amidst an ongoing global pandemic](#), are increasingly turning to and implementing automated and AI-assisted services. It therefore doesn’t take much imagination [to envision scenarios in which a newly automated service can further entrench gender, race, and socioeconomic divides](#) by leaving whole subsections of the population without service, or even recourse, should these automated systems consistently fail to recognize them.

With the growth of AI adoption throughout industry, concerns over fairness and how it is ensured throughout these systems arise. Understanding how to detect and avoid bias in AI models is thus a crucial research topic, growing ever more in urgency as AI steadily and inexorably expands to new sectors, becoming an ubiquitous presence in day-to-day life.



**Figure 1:**  
Average word error rate (WER) across five popular automated speech recognition (ASR) services. Higher WER indicates a higher rate of error. (Source: Racial Disparities in Automated Speech Recognition, PNAS; <https://www.pnas.org/content/117/14/7684>)

# The Accent Gap



As the PNAS study implies, Black English speakers in the US – a population that contains even further subtleties in regional dialects and is by no means monolithic – are not the only demographic that speech recognition technologies have had difficulty serving. The algorithms also struggle to understand non-native English speakers, which account for 350 million people in the country (the majority – 60% – being Spanish speakers).

With the [demographic shifts toward a more ethnically diverse United States](#), businesses employing speech technologies need to speak to and understand numerous demographics if they hope to remain relevant.

A recent feature on the [“accent gap,”](#) published by the Washington Post, further highlighted that Spanish and Chinese accents are the most difficult for popular services like Amazon’s Alexa and Google Home to understand, despite the fact that they are the [top two spoken languages in the world.](#)

As many multilingual users will attest, these shortcomings leave them feeling less able to use these AI-powered services “just because [they] have an accent”. By underserving these populations, companies do more than lose out on their business – they further entrench an antipathy toward AI technologies and the perception that the technology space is a privileged one that does not include people like them.

If anything, these negative experiences also betray the historically problematic biases inherent in technology in general, and how long we’ve tried to eliminate them.

How could it have gotten to this point? Simply put, speech recognition technologies reflect the structural biases and priorities of the people who, historically, are largely responsible for building them. Like most technology in the digital age, speech recognition models were predominantly designed and built by engineers of a white, cisgender, upper-middle-class, male background. A [lack of diverse teams thus translated into a lack of diverse benchmarks and training data,](#) systematically baking bias into the end product, whether consciously intended or not.

Speaking for companies looking to grow into ever larger markets, this is clearly bad for the long-term success of their business.

**“For companies with AI solutions to compete in the large non-native English-speaking market in the US, speech models need to be able to understand a wide range of different Spanish accents, originating from all the Americas.”**

— [Christopher Shulby](#), Director of Machine Learning Engineering at DefinedCrowd

# The Accent Gap



While the tech industry has made promising efforts to diversify engineering teams over the last decade, [change has been slow to realize](#), as most technology that relies on voice recognition models continue to have issues with inclusivity.

So, how can speech and voice technologies keep up? How can we create technologies that not only capture the eclectic vibrance of language, but also ensure that **everyone** is heard and understood? The answer is to consciously collect and employ diverse, representative training data.

Despite how simple it sounds, the secret to inclusivity really is to intentionally be inclusive. How this looks in the speech recognition space is to train models with specially collected and curated speech training data that is broad in scope and contains representative distributions of the accents of diverse groups of people. By using this more representative speech corpora (both in the words that are used, as well as how they are said), speech recognition systems will become more and more robust, understanding even more accents and ways of speaking than the current standard.

**AI Systems are only good  
as the data we put into them.”**

IBM Research

Naturally, the more people your model can understand and serve, the more likely your business can acquire and retain customers. However, this is no trivial task, as many companies do not have the resources to train or test their systems with data specific to different accents. Thankfully however, that's where DefinedCrowd can help.

# Closing the Accent Gap



As we've learned over the history of speech recognition, there's no "right" way to speak English, but we can all agree there's a right way to build AI: ethically and inclusively for all people.

To enable AI developers to test for the accent gap in their technologies, DefinedCrowd is giving away nine hours of Spanish-accented English speech data from the Americas, worth \$1350.

[Simply register on the marketplace here to download the free dataset and test how effective and inclusive your speech recognition systems are.](#)

[Get the Free Dataset now!](#)



Scripted Monologue  
Speech

Free

**English, Americas**  
(Spanish Accented)

Generic

Hours Word Error Rate  
9:26:30.2 4.1%

DefinedCrowd's first-of-its kind platform combines human intelligence and machine-learning backed quality assurance to deliver the quality-guaranteed, project specific data necessary to successful AI initiatives.

**Want to know how a quality-focused training data partner will improve your products?**

We'd love to talk:  
[sales@definedcrowd.com](mailto:sales@definedcrowd.com)

