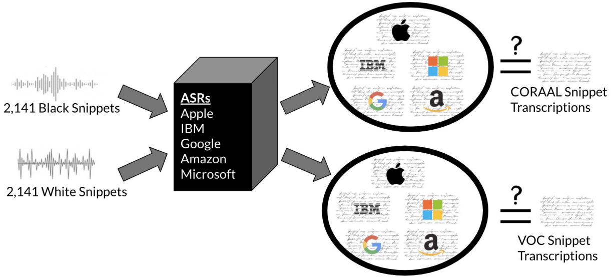
Racial Disparities in Automated Speech Recognition

Allison Koenecke

{koenecke@stanford.edu}

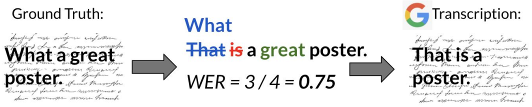
Motivation

Methodology



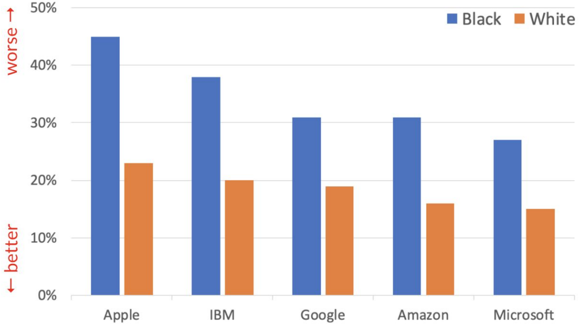
Metric

$$\text{Word Error Rate} = \frac{\text{Substitutions} + \text{Deletions} + \text{Insertions}}{\text{\# Ground Truth Words}}$$

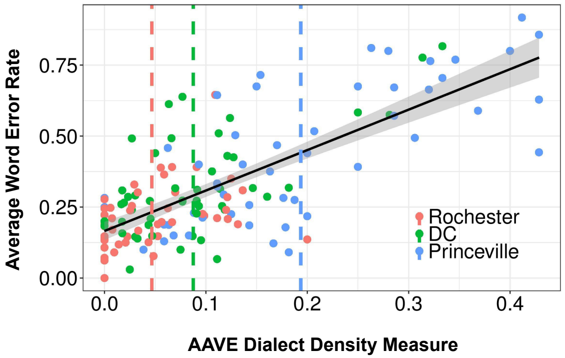


Results

- Speech-to-text transcription error rates are twice as high for Black speakers versus white speakers

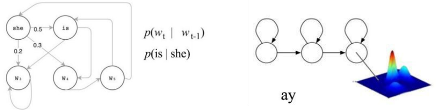


- High word error rates are correlated with high usage of African American Vernacular English linguistic features



Models

- ASR = Language Model + Acoustic Model



- Controlling for for Black and white speakers uttering the same words, we find the acoustic model alone yields racial disparities
- Surprising result: language model is *not* the driver of disparity, despite GPT-2 analysis:

	AAVE Perplexity	**SE Perplexity**
He's a pastor.	305	67
We're going to the ark.	190	88
We're able to fight for the cause.	54	51
Where are they from?	570	20
Have you decided what you're going to sing?	106	25

Call to Action

- Invest in resources to ensure inclusivity in ASR systems and institutions building them
- Collect more diverse training data: for AAVE and other non-standard varieties of English
- Regularly assess and publicly report ASR progress in fairness over time
- Study technical and regulatory progress made in other domains (computer vision)