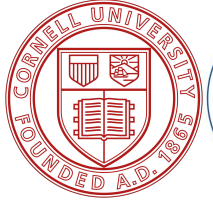
AAAI / ACM conference on
ARTIFICIAL INTELLIGENCE,
ETHICS, AND SOCIETY



Fairness of Automatic Speech Recognition: Looking Through a Philosophical Lens

AIES 2025

Anna Seo Gyeong Choi, Hoon Choi

Three Users, Same Task, Different Burdens

SAE Speaker:
I've been trying to refill my mother's blood pressure medication since Monday.

🗣→✓ (30 sec)

Three Users, Same Task, Different Burdens

SAE Speaker:
I've been trying to refill my mother's blood pressure medication since Monday.

AAE Speaker:
I been trying to refill my mama blood pressure medicine since Monday.

Disfluent Speaker:
I've, I've been trying to re- refill my m- mother's blood, uh, blood pressure medication since, um, since Monday.

Three Users, Same Task, Different Burdens

SAE Speaker:
I've been trying to refill my mother's pressure m… since Monday.

AAE Speaker:
I been trying to refill my… medicine since Monday.

Disfluent Speaker:
I've, I've been trying to re-refill my m- mother's …d pressure medication since, um, since Monday.

All users **"succeeded."** So what's the problem?

The Gap in Current ASR Fairness Research

What we know: Performance gaps across dialects, accents, speech patterns

Racial disparities in automated speech recognition

Allison Koenecke[a], Andrew Nam[b], Emily Lake[c], Joe Nudell[d], Minnie Quartey[e], Zion Mengesha[c], Connor Toups[c], John R. Rickford[c], Dan Jurafsky[c,f], and Sharad Goel[d,1]

Machine Listening: Making Speech Recognition Systems More Inclusive

APRIL 30, 2024
JASA EXPRESS LETTERS | NEWS

SHARE:

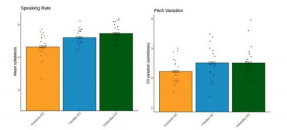
From the Journal: JASA Express Letters

WASHINGTON, April 30, 2024 – Interactions with voice technology, such as Amazon's Alexa, Apple's Siri, and Google Assistant, can make life easier by increasing efficiency and productivity. However, errors in generating and understanding speech during interactions are common. When using these devices, speakers often style-shift their speech from their normal patterns into a louder and slower register, called technology-directed speech.

Research on technology-directed speech typically focuses on mainstream varieties of U.S. English without considering speaker groups that are more consistently misunderstood by technology. In JASA Express Letters, published on behalf of the Acoustical Society of America by AIP Publishing, researchers from Google Research, the University of California, Davis, and Stanford University wanted to address this gap.

One group commonly misunderstood by voice technology are individuals who speak African American English, or AAE. Since the rate of automatic speech recognition errors can be higher for AAE speakers, downstream effects of linguistic discrimination in technology may result.

"Across all automatic speech recognition systems, four out of every ten words spoken by Black men were being transcribed incorrectly," said co-author Zion Mengesha. "This affects fairness for African American English speakers in every institution using voice technology, including health care and employment."



African American English speakers adjust rate and pitch based on audience. Credit: Michelle Cohn, Zion Mengesha, Michal Lahav, and Courtney Heldreth

The Laryngoscope

Quantification of Automatic Speech Recognition System Performance on d/Deaf and Hard of Hearing Speech

Robin Zhao, BS; Anna S.G. Choi, MS; Allison Koenecke, PhD; Anaïs Rameau, MD, MPhil, MS

The Gap in Current ASR Fairness Research

What we know: Performance gaps across dialects, accents, speech patterns

What we're missing:

- Why ethically problematic?
- Different from acceptable statistical variation?
- What kind of harm?

The Gap in Current ASR Fairness Research

What we know: Performance gaps across dialects, accents, speech patterns

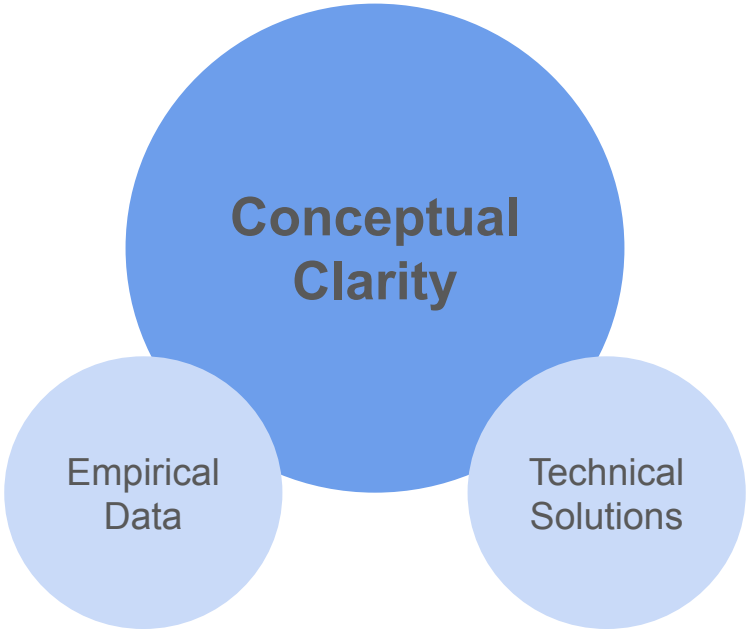
What we're missing:

- Why ethically problematic?
- Different from acceptable statistical variation?
- What kind of harm?

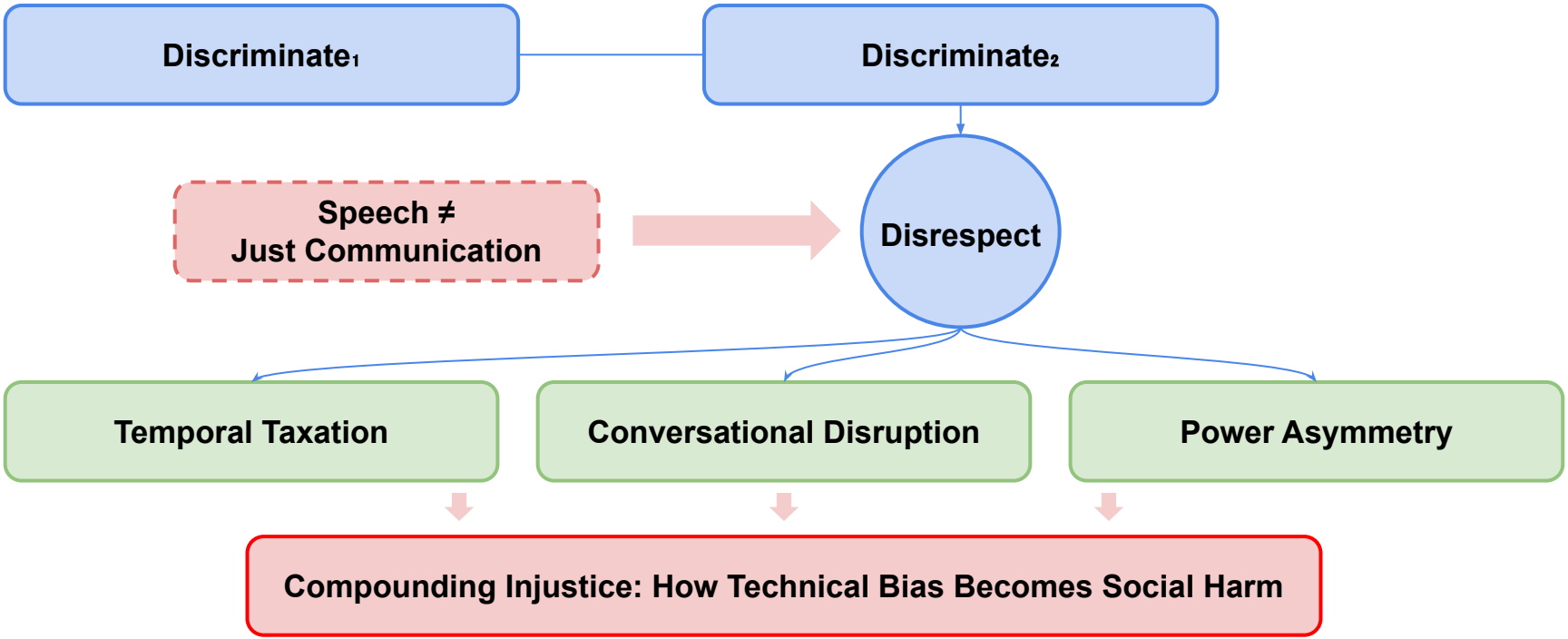
This paper: Philosophical framework for understanding ASR bias as a form of disrespect

A Note on Method

This work provides conceptual analysis, not empirical evidence



Philosophical Framework: Overview



When Does Classification Become Discrimination?

Inductive Reasoning and Statistical Discrimination

ML systems learn through induction:

- Analyze patterns in historical data
- Extrapolate to new inputs
- Inherit biases embedded in training data

When Does Classification Become Discrimination?

Inductive Reasoning and Statistical Discrimination

Discrimination has two meanings:

Discriminate$_1$:
Morally neutral classification (sorting, categorizing)

Discriminate$_2$:
Harmful discrimination (reinforcing social hierarchies)

When Does Classification Become Discrimination?

Inductive Reasoning and Statistical Discrimination

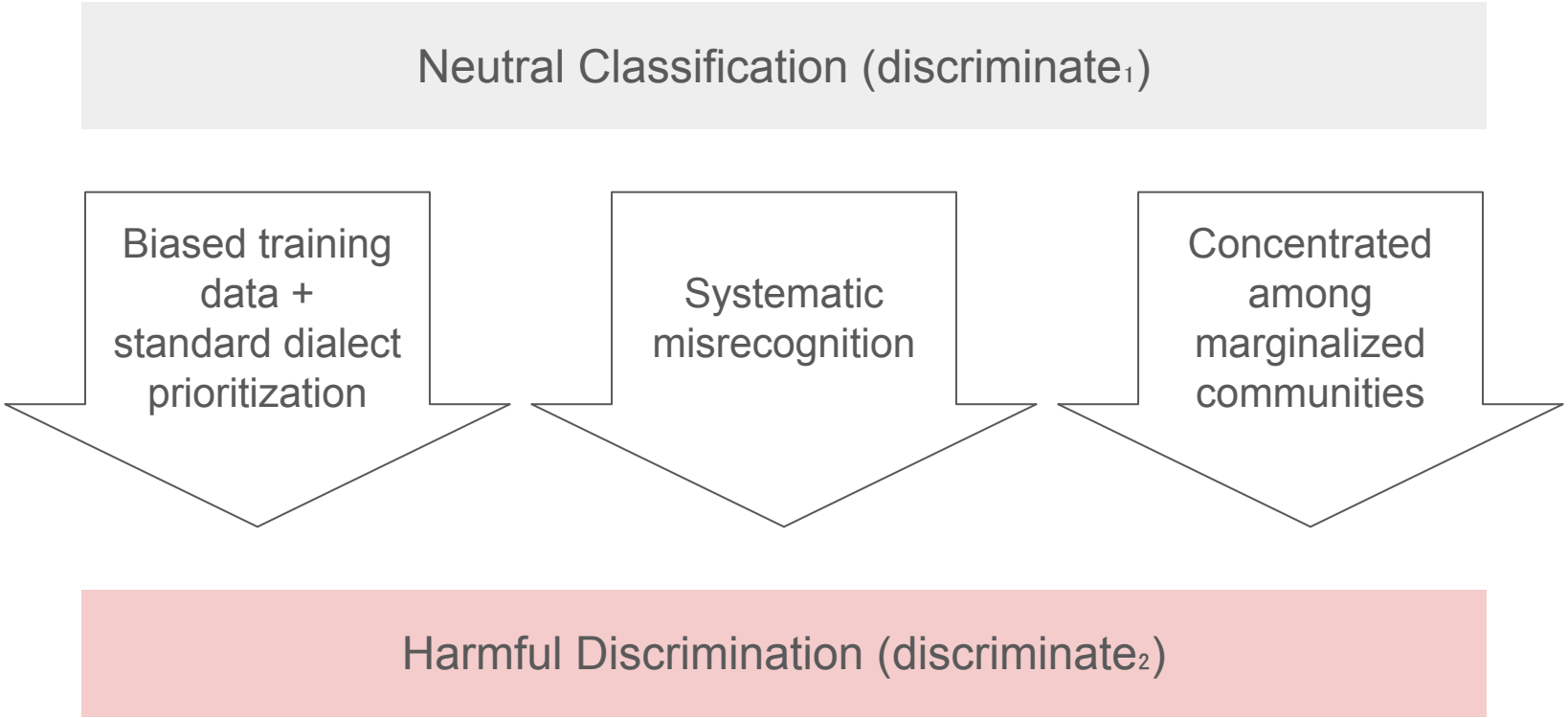
Discrimination has two meanings:

Discriminate$_1$:
Morally neutral classification (sorting, categorizing)

Discriminate$_2$:
Harmful discrimination (reinforcing social hierarchies)

The question: When does discriminate$_1$ become discriminate$_2$?

When Does Classification Become Discrimination?



Why This Matters: Compounding Injustice

Compounding Injustice

Hellman’s concept: Algorithmic harm accumulates when decisions build on existing inequalities

Why This Matters: Compounding Injustice

Compounding Injustice

Hellman's concept: Algorithmic harm accumulates when decisions build on existing inequalities

In ASR:
AAE historically stigmatized → Social discrimination
→ ASR adds tech marginalization

Why This Matters: Compounding Injustice

Compounding Injustice

Hellman's concept: Algorithmic harm accumulates when decisions build on existing inequalities

In ASR:
AAE historically stigmatized → Social discrimination
→ ASR adds tech marginalization

Result: Not just technical errors, but reinforcement of social hierarchies

What Makes ASR Bias Disrespect?

Speech traits straddle moral categories:

Category 1:
Visible & unchangeable (race, congenital speech patterns)

Category 2:
Visible & changeable with difficulty (dialect, accent, socioeconomic markers)

Category 3:
Not visible & readily changeable (temporary conditions)

What Makes ASR Bias Disrespect?

Speech traits straddle moral categories:

Category 1:
Visible & unchangeable (race, congenital speech patterns)

Category 2:
Visible & changeable with difficulty (dialect, accent, socioeconomic markers)

Category 3:
Not visible & readily changeable (temporary conditions)

Speech patterns fall here

Immediately audible, deeply tied to cultural identity, not readily changeable without significant personal cost

What Makes ASR Bias Disrespect?

Speech traits straddle moral categories:

Category 1:
Visible & unchangeable (race, congenital speech patterns)

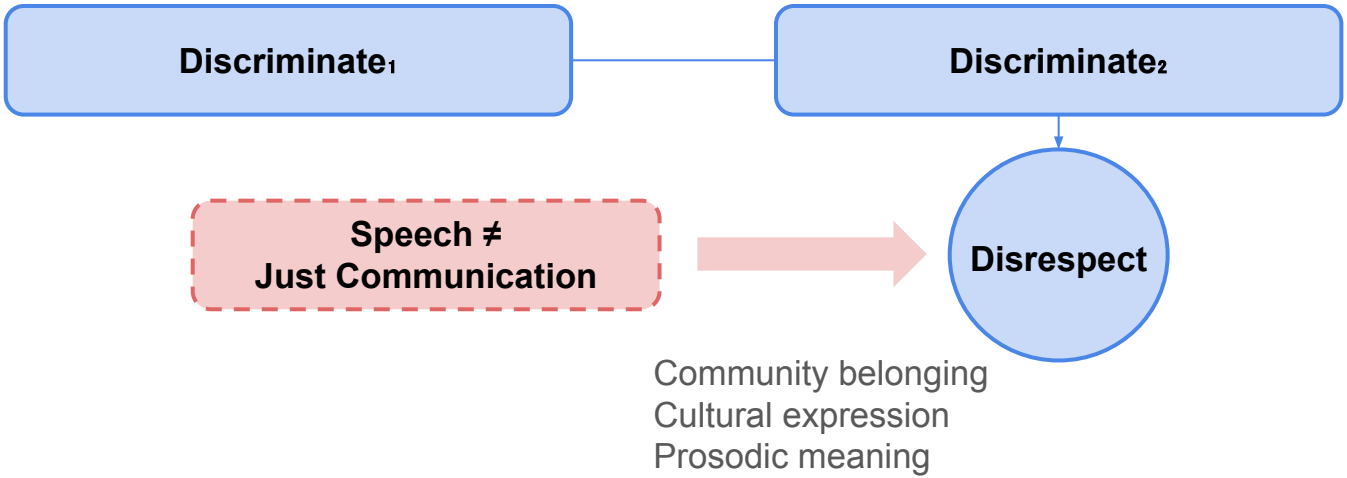
Category 2:
Visible & changeable with difficulty (dialect, accent, socioeconomic markers)

Category 3:
Not visible & readily changeable (temporary conditions)

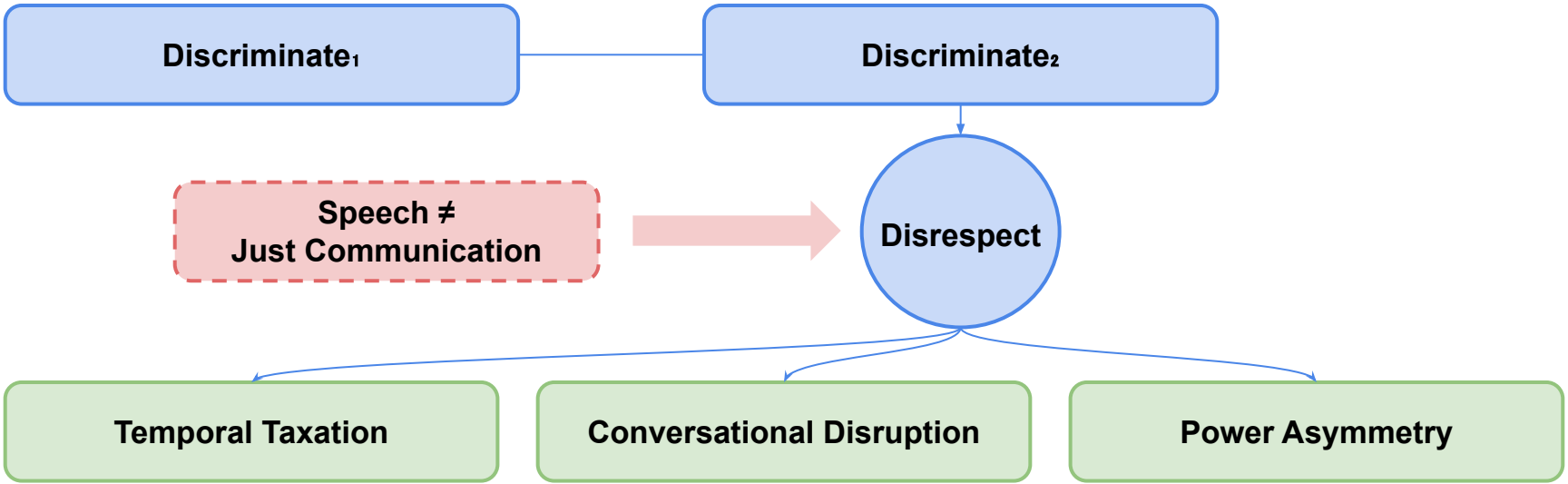
When ASR consistently fails certain speech

→ treats speakers as less worthy of accommodation
→ constitutes disrespect

Philosophical Framework: Overview



Philosophical Framework: Overview



Dimension 1: Temporal Taxation

Unequal distribution of time costs

SAE Speaker:
I've been trying to refill my mother's blood pressure medication since Monday.

AAE Speaker:
I been trying to refill my mama blood pressure medicine since Monday.

Dimension 1: Temporal Taxation

Unequal distribution of time costs

SAE Speaker:
I've been trying to refill my mother's blood pressure medication since Monday.

AAE Speaker:
I been trying to refill my mama blood pressure medicine since Monday.

6x longer for the same task

20 voice interactions

~200 hours/year

(30 sec)

(3 min)

Dimension 1: Temporal Taxation

Unequal distribution of time costs

SAE Speaker:
I've been trying to refill my mother's blood pressure medication since Monday.

AAE Speaker:
I been trying to refill my mama blood pressure medicine since Monday.

Beyond time

- 🧠 Cognitive monitoring
- 🔄 Linguistic labor
- 💼 Economic loss

🗣️→✓ (30 sec)

🗣️→✗→🗣️→✗🗣️→✗→🗣️→✓ (3 min)

Dimension 2: Conversational Disruption

Speech requires temporal flow for meaning

I been trying to refill my mama blood pressure medicine since Monday

Dimension 2: Conversational Disruption

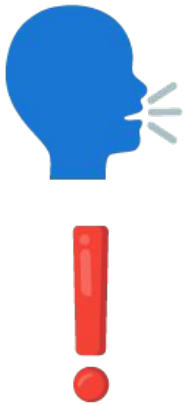
Speech requires temporal flow for meaning

I been trying to refill my mama blood pressure medicine since Monday

Refill, Mama medicine, Blood Pressure, Monday

Dimension 2: Conversational Disruption

Speech requires temporal flow for meaning



I been trying to refill my mama blood pressure medicine since Monday

Refill, Mama medicine, Blood Pressure, Monday

MEDICINE. REFILL. MAMA. BLOOD. PRESSURE. MONDAY.

Dimension 3: Power Asymmetry

System controls pace:

- Can interrupt at will
- Demands infinite repetitions
- Forces conformity to its expectations

Speaker without reciprocal power:

- Cannot pause the system
- Cannot request alternative modes
- Cannot negotiate interaction terms

Dimension 3: Power Asymmetry

System controls pace:

- Can interrupt at will
- Demands infinite repetitions
- Forces conformity to its expectations

Speaker without reciprocal power:

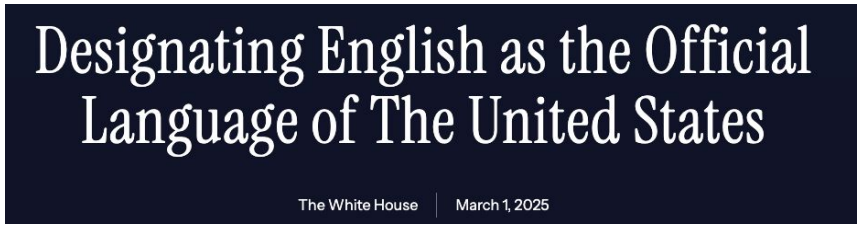
- Cannot pause the system
- Cannot request alternative modes
- Cannot negotiate interaction terms

In high-stakes contexts:

- Clinical errors
- Interview disadvantages
- Emergency delays

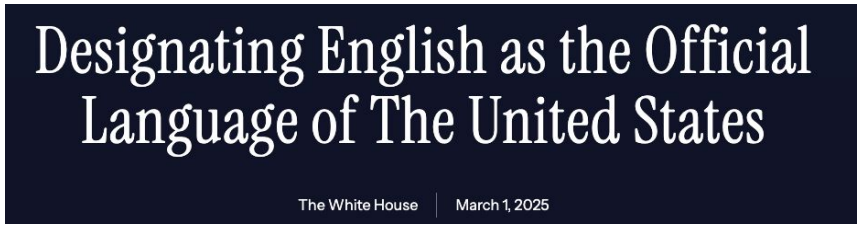
Recent Policy Context

Executive Order 14224 (March 1, 2025): English official language of US



Recent Policy Context

Executive Order 14224 (March 1, 2025): English official language of US



Creates “cascade of legitimation”
→ ASR bias can claim policy alignment

What This Framework Reveals

Standard fairness metrics miss critical harms:

✓ Measure: Final accuracy

✗ Miss: Time-to-completion equity

✓ Measure: Aggregate performance

✗ Miss: Identity erasure & Disrespect

What This Framework Reveals

Standard fairness metrics miss critical harms:

✓ Measure: Final accuracy

✗ Miss: Time-to-completion equity

✓ Measure: Aggregate performance

✗ Miss: Identity erasure & Disrespect

Two speakers might both “succeed”
while experiencing vastly different burdens

Implications: Beyond Technical Fixes

This philosophical analysis shows:

ASR bias is disrespect that compounds historical injustice

Implications: Beyond Technical Fixes

This philosophical analysis shows:
ASR bias is disrespect that compounds historical injustice

Changes the questions:

- “Improve accuracy” → “Whose autonomy matters?”
- “Reduce errors” → “Whose time matters?”
- “Balance datasets” → “Whose speech is legitimate?”

Implications: Beyond Technical Fixes

This framework suggest new directions:

- Evaluation: Temporal burden metrics
- Accommodation: Proactive not reactive
- Governance: Linguistic data sovereignty
- Legal: Address compounding injustice

Implications: Beyond Technical Fixes

This framework suggest new directions:

- Evaluation: Temporal burden metrics
- Accommodation: Proactive not reactive
- Governance: Linguistic data sovereignty
- Legal: Address compounding injustice

How do we build ASR systems that respect linguistic diversity rather than enforce standardization?

Thank You!

Link to paper

📧 sc2359@cornell.edu