

Perceptions of Language Technology Failures from South Asian English Speakers

South NLP Symposium 2024

April 5, 2024 | Emory University, Atlanta GA

WILL HELD, willheld3@gatech.edu

FAYE HOLT, mholt9@gatech.edu

DIYI YANG, diyiy@stanford.edu



INTRODUCTION

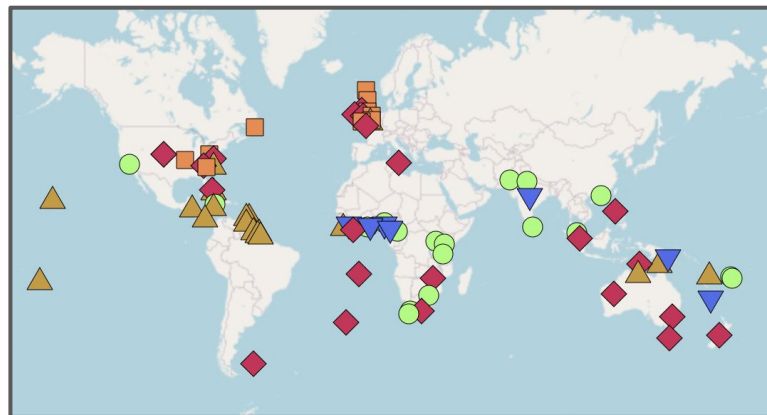
Motivation: *User Perceptions* of Dialect Disparity

- Pervasive dialect disparity in NLP ([Ziems et al., 2022](#))
- Language (Technology) is Power ([Blodgett et al., 2020](#))
- *Perspectives* of AAVE speakers on ASR ([Mengesha et al., 2021](#))

“I think technologies should be designed in a way that they are able to understand ever[y] dialect.” - Participant 18 (P18)

BACKGROUND

- English is spoken worldwide, with diverse regional variations and dialects
- NLP systems are designed primarily with the Standard American English (SAmE) variety in mind
- Performance disparities exist for non-SAmE varieties of English across several NLP tasks



eWave: World Varieties of English

PROBLEM

How should these discrepancies be addressed? **What do users want?**

- The next step in designing more inclusive NLP systems is to understand user **preferences** and **needs**. Systems should meet user preferences and **avoid reinforced harms** to global English speakers.

▶ BACKGROUND

- Attitudes toward dialectal variation
- Surveying as a tool for understanding user perceptions ([Mengesha et al., 2021](#))
- South Asian Englishes (SAsE): widespread usage and prior empirical exploration in NLP

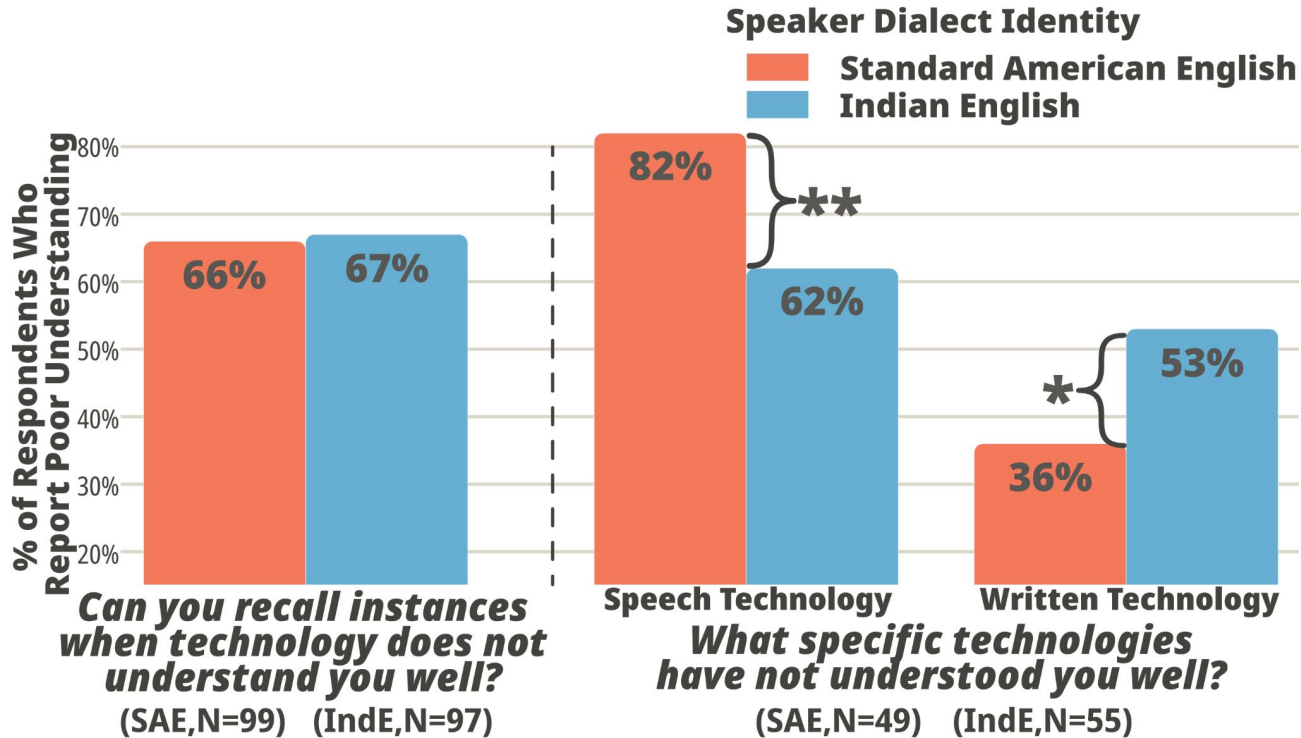
RESEARCH AIM

To understand SAsE user preferences and challenges in relation to language technology and assess how these perceived challenges manifest in current Large Language Models (LLMs).

METHOD

- User-centric diagnostic study of failures
 - 78 SAsE and 97 SAmE speakers surveyed on Prolific
- Intrinsic benchmark of SAsE knowledge
 - lexical understanding assessment from Wiktionary
 - minimal pair syntactic language modeling evaluation ([Demszky et al., 2021](#))
- Extensive evaluation of LLMs
 - 8 open-source models and 3 industrial LLMs

SURVEY RESULTS



SURVEY RESULTS - PERCEIVED CHALLENGES

#1 Failures with stand-alone dialect words (Occurrence: 43%)

“[I avoid using] some slang words.

‘Buggy’ instead of ‘shopping cart’ for example.” - P2

SURVEY RESULTS - PERCEIVED CHALLENGES

#2 Codeswitching (Occurrence: 18%)

“I want to be able to speak bilingually with technology.” - P7

#3 Register and Syntax (Occurrence: 20%)

“Language in for technology is so much more formal than spoken.” - P19

BENCHMARKING LLMS

Existing benchmarks do not cover **all** of the reported challenge categories and notably **omit stand-alone lexical variation**.

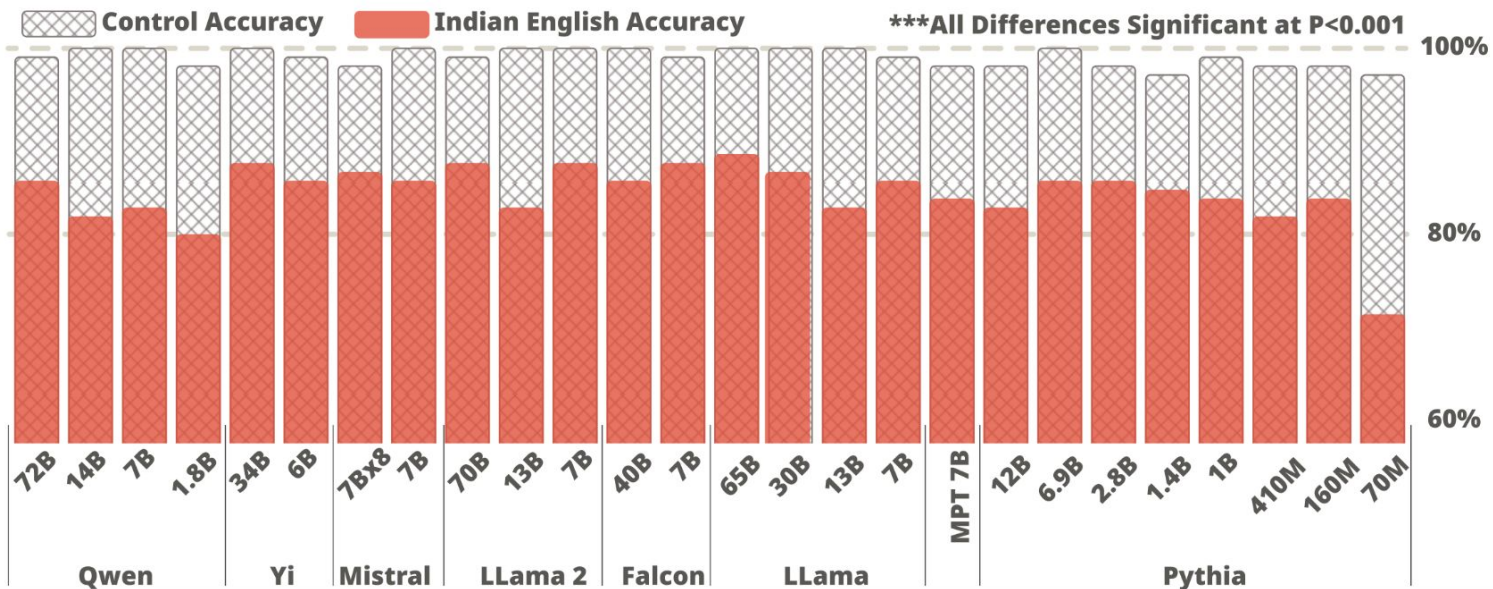
```
Which of the following could \"{TERM}\" mean in
Indian English when used as a {
PART_OF_SPEECH}?
{OPTIONS A THROUGH D}
Answer:
```

Syntactic variation

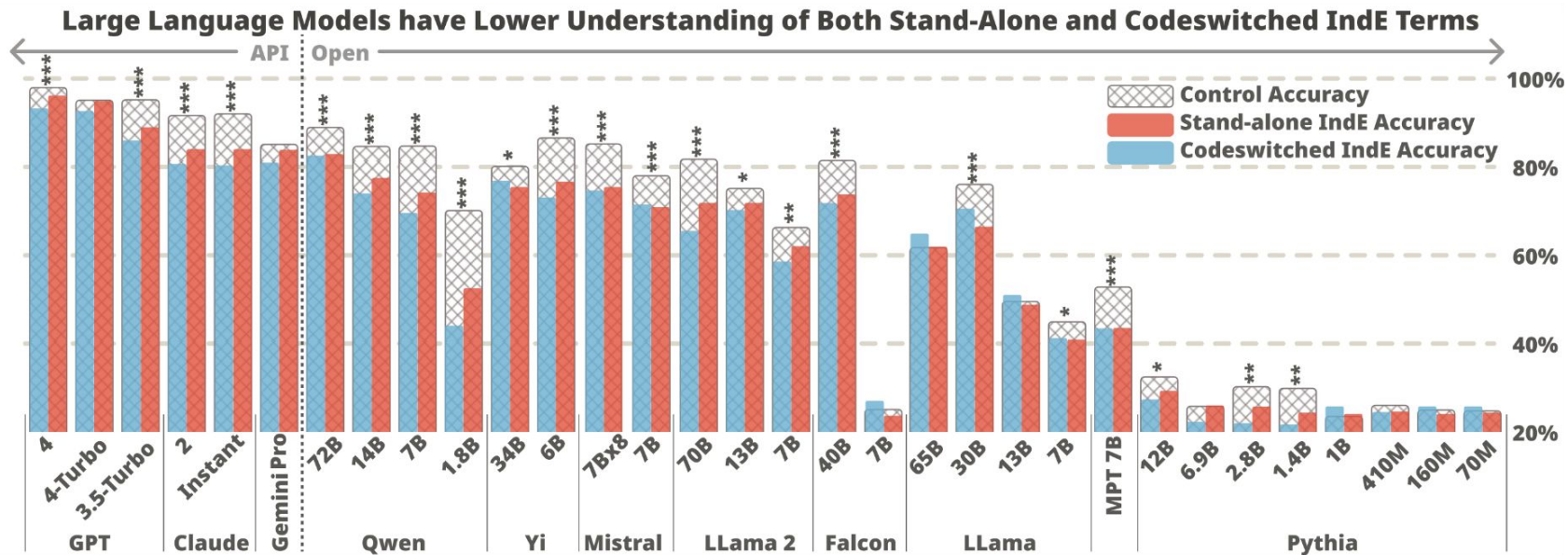
The following is an example of acceptable Indian English: \"{SENTENCE}\"

BENCHMARKING LLMS - RESULTS

Large Language Models Are Less Able to Identify Attested Indian English Syntax



BENCHMARKING LLMS - RESULTS



CONCLUSIONS AND IMPLICATIONS

- More SAsE speakers recall language technology failures than SAmE speakers
- Challenges extend beyond accent differences
- Users modify SAsE features to improve technology performance
- Benchmark results confirm user experiences, **highlighting the need for language technologies to accommodate dialectal variations even for monolingual systems**

Perceptions of Language Technology Failures from South Asian English Speakers

WILL HELD, FAYE HOLT, DIYI YANG

