# Contextual Real-time Intelligent Typo Identifier & Corrector (CRITIC)

Final Report - 12/12

Trevor La
CS Major
ML Engineer

Kevin Francis
CS Major
ML Engineer

Nicholas Miklaucic
CS Ph.D. Student
ML Research Scientist

# Problem Statement

If you type

"Judge Thomas Ambro, Majotity Opinion"

You probably wanted to say

"Judge Thomas Ambro, Majority Opinion"

Maybe a computer could help you out!

# Motivation

- Typos can be embarrassing, but hard to catch

- You've definitely used spell checkers: phone, computer, Google Docs, Outlook, etc.

> Do you want me to order a pizza for dinner and I'll pick it up on my way home?
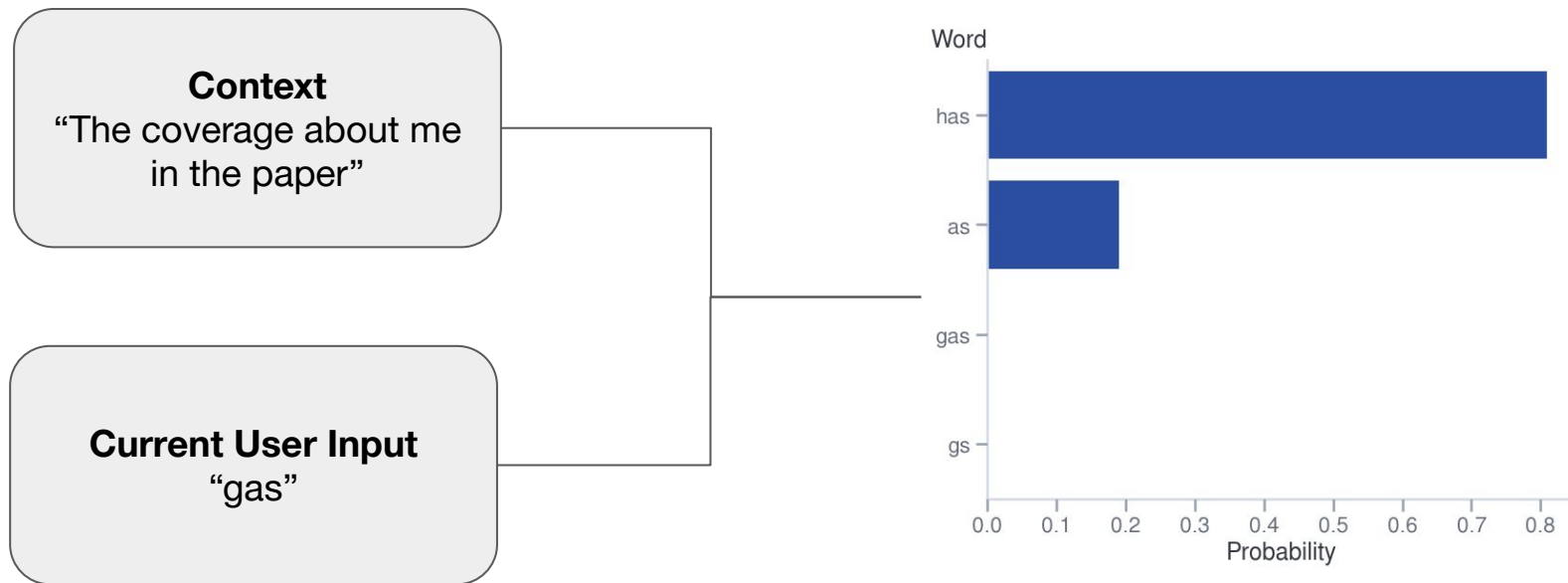
> I am GOD

> ...good

> I am good

> Uh

Read 8:28 PM

# Project I/O

From output probabilities of what the user meant to say, correction can be modulated depending on user preferences



K

# Technical Challenges

- **Real-time**—as fast as you type

- **Local**—keystrokes are private user data

- **Consumer hardware**—myriad different configurations

T

# Related Work

- Grammarly is most similar spell checker to CRITIC but lacks offline usage and real-time corrections

- JetBrains' AI autocomplete features is not a spell checker but, like CRITIC, features context awareness and real-time feedback

T

# Overview of Solution

"The coverage about me in the paper **gas**" →
"The coverage about me in the paper **has**"

Two components that work together:

- **Language Model:** How likely is **has** as a completion given context?
- **Keyboard Model:** If the user meant to type **has**, what's the chance they typed **gas**? (**h** → **g**)

Combining these sources of information results in a more robust corrector.

N

# Datasets and Metrics

- Our keyboard model was trained on an Aalto University/University of Cambridge study that collected over 136 million keystrokes: we evaluated on a subset of 2503 unique errors representing many thousands of typed words
- Each participant was prompted, so we can compare CRITIC outputs to the ground truth of what they meant to type
- Classification metrics:
  - Average probability for true word
  - Average rank of true word in corrections
  - Top-1 accuracy
- Time per word

N

# Results

- **91.3%** top-1 accuracy
- Average rank: **1.263**
- Average probability: **0.905**
- **0.345** seconds/word—real-time completion for even **173 wpm**

When filtering to just completions that exist in the base dictionary CRITIC uses:

- **94.1%** top-1 accuracy
- Average rank: **1.263**
- Average probability: **0.933**

N

# Results

Interpretation is difficult: not all errors are "reasonable" corrections

"Will you and **KB** be around this afternoon?" →
"Will you and **B** be around this afternoon?"

The model weakly prefers B, but in a real-world deployment this is not likely to be corrected either way

Depending on specific text and user, accuracy could be significantly higher

N

# Future Improvements

While CRITIC is functional and able to make real-time typo corrections, there is room for improvement:

- Integration into operating systems/text editors
- Personal dictionaries
- User studies to refine user experience

T

# Limitations

- Inevitably, the use of an AI Spell Checker will have more CPU cost than a built-in typo correction algorithm, such as the one we used while making this Google Slides presentation.

- The project is limited to languages Llama 3.2 supports, and can only support alphabets: e.g., Japanese is completely different

- The keyboard model is optimized for QWERTY, requiring technical adjustments for other layouts.

T

# References

- Cambridge Study- https://userinterfaces.aalto.fi/136Mkeystrokes/

- Llama-https://huggingface.co/Solshine/Llama-3.2-1B-Q4_K_M-GGUF

- Our Github Repo-https://github.com/csce585-mlsystems/critic-llm