Measuring and Modifying Readability of English Texts with GPT-4

Sean Trott & Pamela Rivière
Department of Cognitive Science
UC San Diego

RQ: Can Large Language Models (LLMs) reliably assess and manipulate the readability of English text?

Methods

Corpus: CommonLit Ease of Readability (CLEAR) Corpus → 4,724 text excerpts with readability ratings derived from human teachers (Crossley et al. 2023)

LLMs: GPT-4 Turbo & GPT-4o mini

Approach: "zero-shot" (no examples provided) LLM prompting to produce:

readability estimates for each excerpt

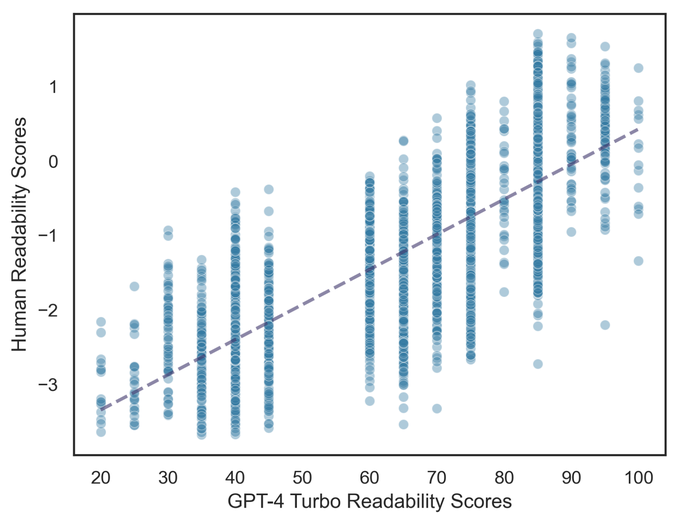
```
"Read the text below. Then, indicate the readability of the text, on a scale from 1 (extremely challenging to understand) to 100 (very easy to read and understand). In your assessment, consider factors such as sentence structure, vocabulary complexity, and overall clarity."
```

modified excerpts according to target difficulty

```
"Read the passage below. Then, rewrite the passage so that it is easier to read. When making the passage more readable, consider factors such as sentence structure, vocabulary complexity, and overall clarity. However, make sure that the passage conveys the same content.Finally, try to make the new version approximately the same length as the original version."
```

Validation: collected new human readability judgments (1 = very challenging; 5 = very easy) to assess LLM-modified excerpts (n = 59 ppts)

GPT-4 Turbo readability estimates predict CLEAR Corpus human teacher judgments

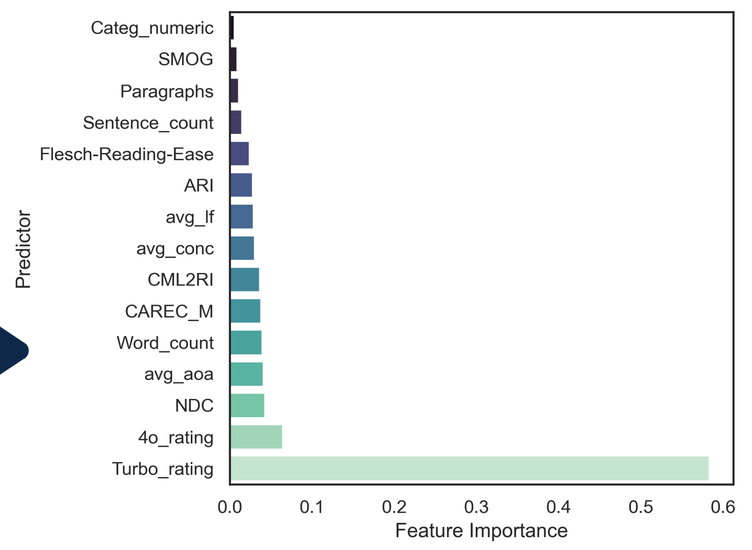


LLM readability estimates are positively related to human readability ratings from the CLEAR Corpus

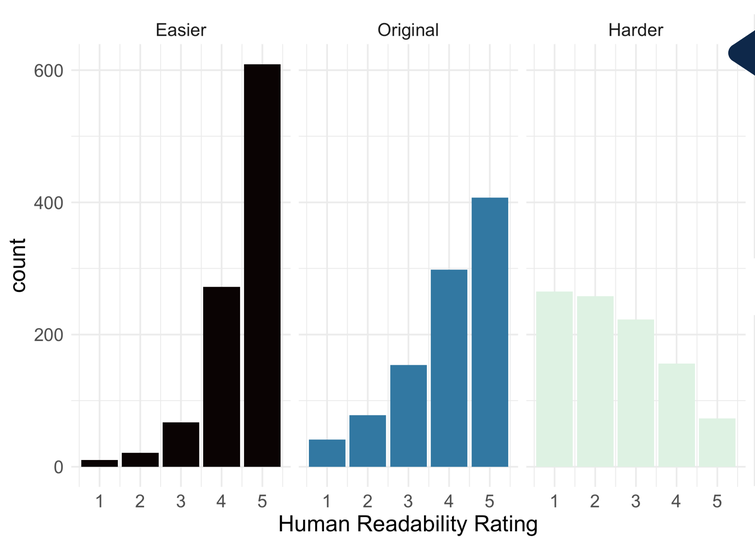
GPT-4 Turbo (r = 0.76)

GPT-4o mini (r = 0.74)

GPT-4 Turbo's ratings are better predictors of human teacher ratings than other metrics

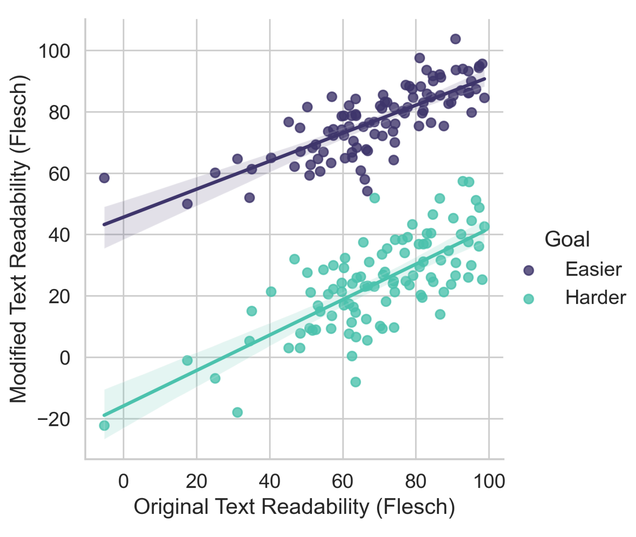


GPT-4 Turbo successfully modified texts to target difficulty, according to human judges

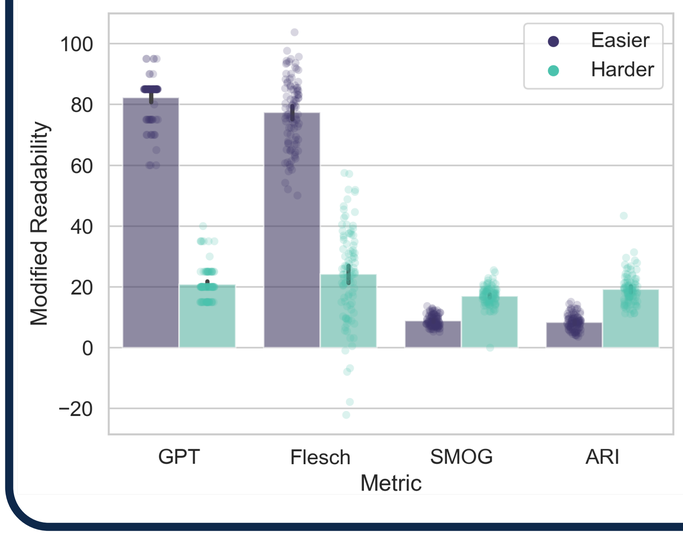


Human judges were more likely to rate texts as challenging when target difficulty was "harder"; and more likely to rate texts as easy to read when target difficulty was "easier"

Readability of modified text exhibits residual correlation with original text's readability



GPT-4 Turbo successfully modified texts to target difficulty, according to existing readability metrics



Scores from existing readability metrics reflect the target difficulty for LLM-modified text excerpts

Discussion

GPT-4 Turbo & 4o mini can estimate the readability of English text

GPT-4 Turbo can successfully modify English text to a target readability

Acknowledgements

University of California San Diego Chancellor's Postdoctoral Fellowship Program

Benjamin Bergen, Cameron Jones