Text Simplification For Children Application

Presenters: Surya Appana, Ved Karan Reddy Peesu, Wil Rothman

Advisor: Anastasia V. Smirnova

ECOLE Lab at San Francisco State University and UC Berkeley CDSS

Background

Text Simplification (TS): the process of reducing structural and linguistic complexity of a text, while retaining that text's original meaning [2].

Previous Work

- This poster is an extension of our lab's previous work on large language model (LLM) text-simplification for children. By comparing human-written simplifications from the Newsela corpus with parallel LLM-generated output
- This work shows an emergent ability for GPT-4o to simplify texts for fourth-grade children in the news-genre by comparing human-written simplifications with parallel LLM-generated output [1]; we use these results as the basis for our application.

Motivation

- LLMs hold great potential for education, as they can remove barriers to information access [3], address lack of high-quality educational materials [4], and to support teachers [5].
- Previous HCI researchers and ethicists have called for an application that can automatically simplify texts [6][7][8]. So, our goal is to develop a prototype application which can take a complex text, simplify it, and return the result to the user through LLMs.

Methods

Data and Analysis

- We use the **Newsela** corpus [9] to compare LLM-simplified text with human-simplified text. In addition to generated questions, we compared **Automated Readability Index** (ARI) [10] between the texts. We used `textstat` to compute ARI scores.

Application Structure

- Our website application uses iterative prompting to simplify text at the user's reading level. Three kinds of users are considered either native, international speakers, and international learners of English.
- The user can input text and request a simplified version at a particular grade level (1-12).
- The simplified text will be accompanied by a generated multiple-choice comprehension question to test the user's understanding of the text, after which further simplification can be requested.

Prompt Engineering

- We gave GPT-4o the following prompts, inspired by [7].
  - For initial simplification: `"Please simplify the following text so that it may be understood by native speakers at the grade {grade} level: {text}"`
  - For question writing: `"Please form a multiple_choice comprehension question with choices (A), (B), (C), and (D) for the following text that would be appropriate at the grade {grade} level. Please write 'Correct Answer' before writing the letter corresponding to the correct answer: {text}"`

Results

Findings

We assigned ARI scores for 139 sampled grade 4 Newsela articles (see *Figure 2*). We observe a discrepancy between ARI scores and Newsela's subjective grade level designations. According to [10], non-simplified articles have a score typically assigned for sixth grade, Newsela has one for first grade, and LLMs have ones for fourth grade (the correct reading levels). This indicates a potential flaw with the Newsela corpus and [9]. The high variance for LLM-generated questions suggest GPT-4o might struggle to design grade-appropriate questions.

Conclusions

The complete application simplification text based on user proficiency, grade level, and user responses to questions. The LLM-simplified text and LLM-generated questions accurately matched user reading levels and background through prompt engineering.

Future Work

We will look into detailed simplification metrics for LLM-generated text on the Newsela corpus. Additionally, we will compare our results with different LLM models like GPT-o3 Gemini 2.5. Also, improvements can be made if we analyze LLM-generated multiple choice similarity using different text embedding models (TEIs). Lastly, we can utilize the Hugging Face text embeddings to compare the similarity in the generated multiple choice samples and change accordingly.

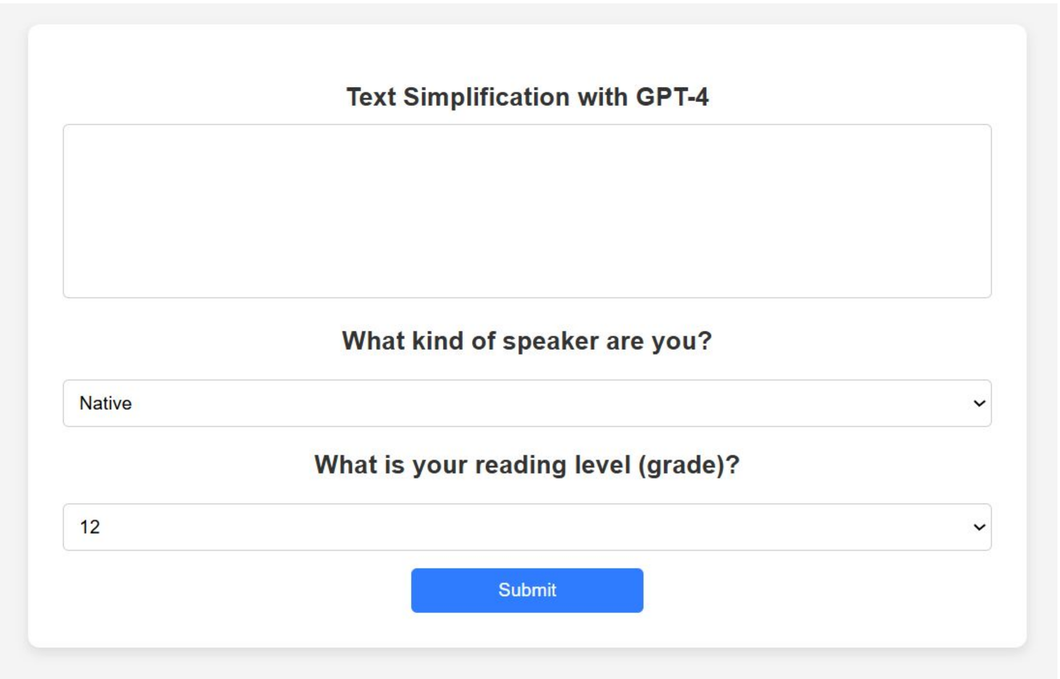

Figure 1: graphical user interface (GUI) for the application.

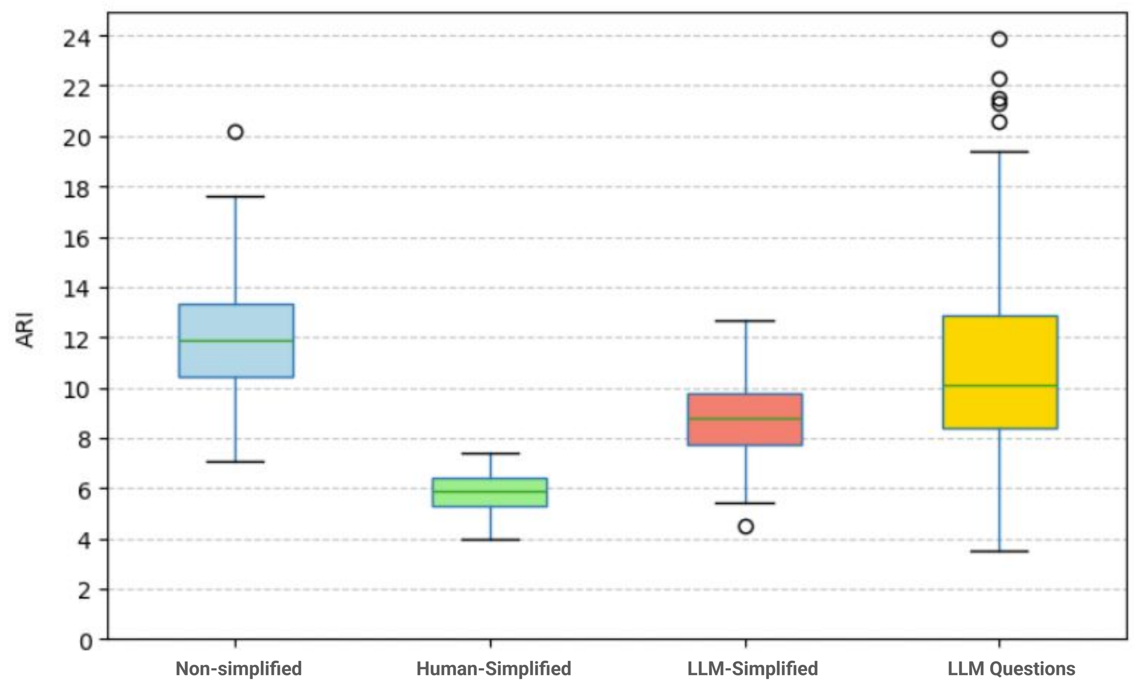

Figure 2: ARI scores for original text, human-simplified text, and LLM-simplified text, and LLM-simplified comprehension questions.

References and Acknowledgments

[1] Smirnova, Anastasia, et al. "Text Simplification for Children: Evaluating LLMs vis-à-vis Human Experts." *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (ACM CHI '25)*, 2025.
[2] Siddharthan, Advaith. "A survey of research on text simplification." *ITL-International Journal of Applied Linguistics* 165.2 (2014): 259-298.
[3] Schicchi, Daniele, and Davide Taibi. "AI-driven inclusion: Exploring automatic text simplification and complexity evaluation for enhanced educational accessibility." *International Conference on Higher Education Learning Methodologies and Technologies Online*. Cham: Springer Nature Switzerland, 2023.
[4] Freyer, Nils, et al. "Easy-read and large language models: on the ethical dimensions of LLM-based text simplification." *Ethics and Information Technology 26.3 (2024):* 50.
[5] Chiu, Thomas KF, et al. "Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education." *Computers and Education: Artificial Intelligence* 4 *(2023)*: 100118.
[6] Kim, Jenia, Stefan Leijnen, and Lisa Beinborn. "Considering Human Interaction and Variability in Automatic Text Simplification." *Proceedings of the Third Workshop on Text Simplification, Accessibility and Readability (TSAR 2024)*. 2024.
[7] Scaria, Nicy, et al. "How Good are Modern LLMs in Generating Relevant and High-Quality Questions at Different Bloom's Skill Levels for Indian High School Social Science Curriculum?." *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024).* 2024.
[8] Gooding, Sian, and Manuel Tragut. "One size does not fit all: The case for personalised word complexity models." arXiv preprint arXiv:2205.02564 (2022).
[9] Xu, Wei, et al. "Problems in current text simplification research: New data can help." *Transactions of the Association for Computational Linguistics 3* (2015): 283-297.
[10] Smith, Edgar A., and R. J. Senter. "Automated readability index". Vol. 66. No. 220. Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, 1967.

Part of the San Francisco State University Experimental and Computational Linguistics Ensemble Laboratory and the University of California, Berkeley College of Computing, Data Science, and Society. Project mentor Anastasia V. Smirnova, ECOLE Lab Director. Presented at CDSS Discovery Symposium Fall 2024, UC Davis Symposium on Language Research, and Stanford Research Conference.