Persona-Based Prompting: Enhancing Readability and Understanding in AI Responses for children

Jordano B. de Castro[1]

Supervisor(s): Sole Pera, Hrishita Chakrabarti[1]

[1]EEMCS, Delft University of Technology, The Netherlands

Introduction

children are increasingly interacting with large language models to explore topics, seek help with schoolwork, or engage in creative inquiry. However, these systems frequently generate responses that exceed young users' reading or comprehension levels.

Previous approaches have attempted to improve readability by modifying prompts, such as appending terms like "for kids" or explicitly stating the target demographic. However, these strategies have produced limited and inconsistent improvements in the readability of the generated content. This project investigates an alternative method: persona-based prompting. Instead of telling the model to generate content for a child, we instruct it to take on the role of a familiar and trusted figure, such as a teacher or caregiver. We hypothesize that this role-based strategy may result in responses that are better suited to the comprehension abilities of children. We use a dataset of real search queries written by children and evaluate the responses generated by several leading open-source language models, both with and without persona instructions. These responses will be analyzed using standard readability metrics and word-level analysis based on Age of Acquisition ratings [1] to assess how well they match the reading capabilities of a young audience

Methodology

We test our hypothesis using current leading state-of-the-art open-source language models to evaluate how well they generate readable content for children. The models included in our study are:

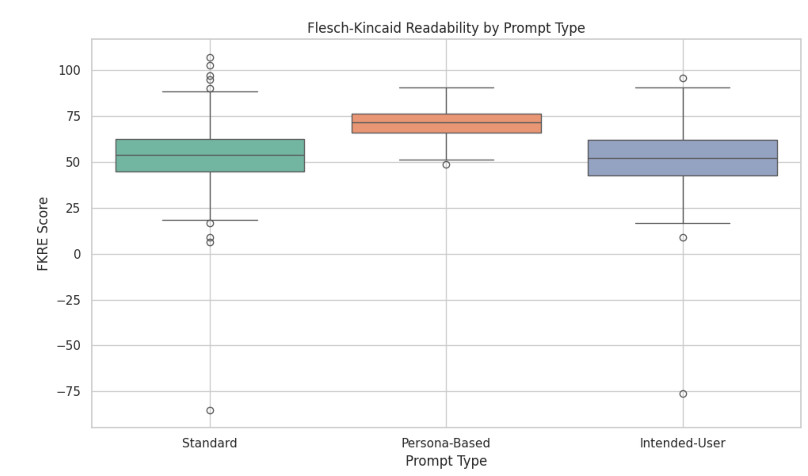
- Mistral 7B
- DeepSeek-R1
- Qwen2.5

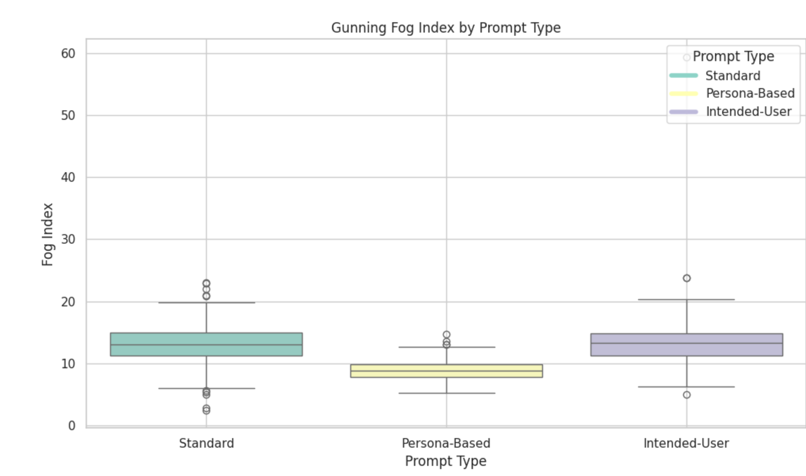
Strategy

- **Persona-Based Prompting**: The LLM is instructed to respond from the perspective of a role suited for children (e.g., a parent or teacher), helping it generate more age-appropriate content.
- **Intended-User Prompting**: The prompt specifies that the response is for children 6 - 13 but does not ask the model to take on a specific role. This replicates the approach from Rooein et al. [2], updated with newer models.
- **Standard Prompting**: A neutral prompt with no audience information, used to assess how the models perform without any special instruction.

Mistral 7B results

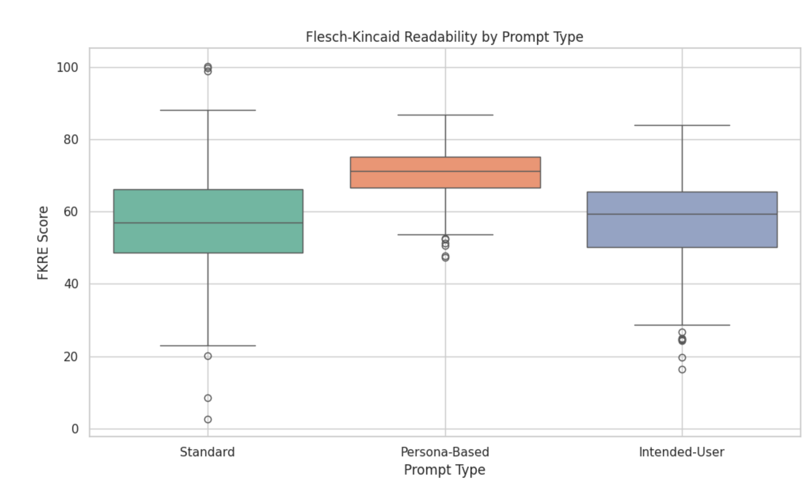
Across all metrics, Persona-Based Prompting significantly outperformed both Standard and Intended-User Prompting. Notably, for the Flesch-Kincaid Reading Ease and Gunning Fog Index, Intended-User Prompting did not show significant improvement over Standard Prompting

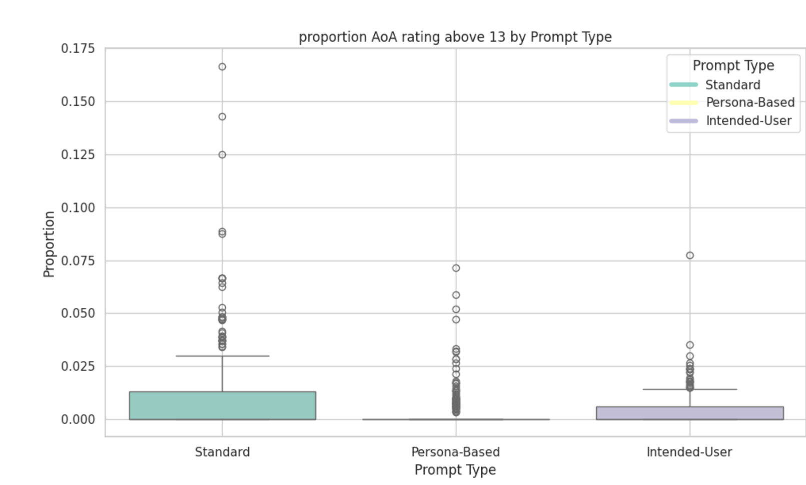




DeepSeek-R1 results

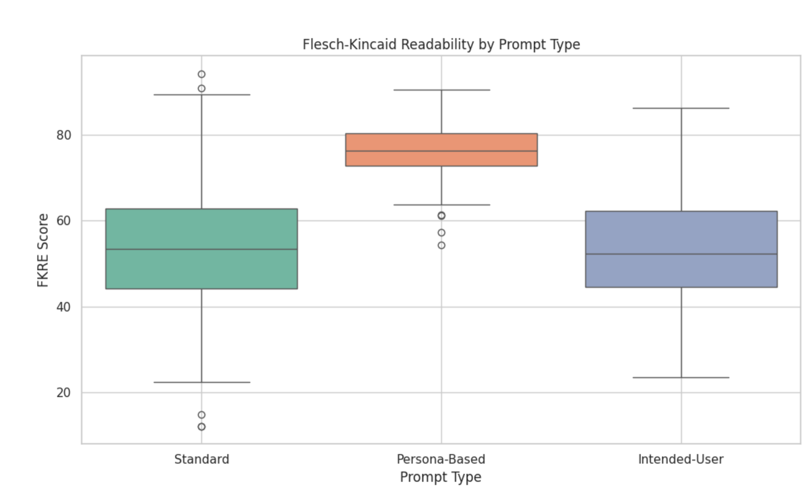
DeepSeek was the only model where Persona-Based Prompting did not significantly outperform Intended-User Prompting for the proportion of words with an AoA above 13. However, for all other metrics, Persona-Based Prompting showed clear and significant improvements.

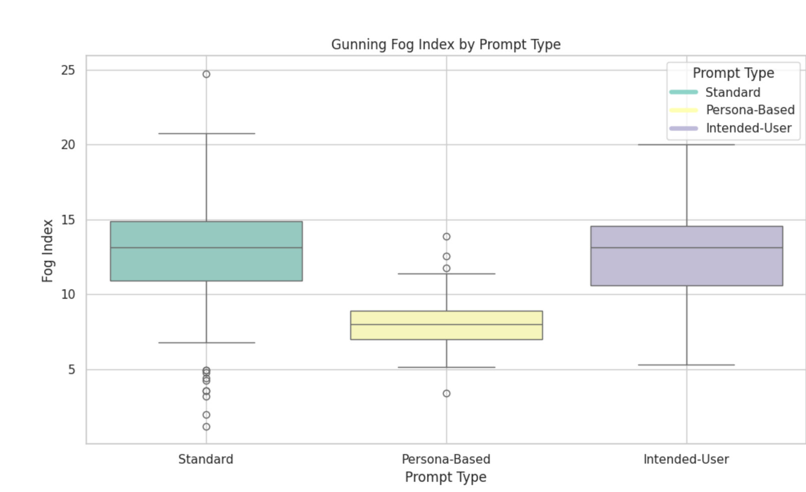




Qwen2.5 Results

Across all metrics, Persona-Based Prompting significantly outperformed both Standard and Intended-User Prompting. Notably, for the Flesch-Kincaid Reading Ease and Gunning Fog Index, Intended-User Prompting did not show significant improvement over Standard Prompting





Metrics

The following metrics were used to evaluate the readability and accessibility of LLM responses:

Flesch Reading Ease (FRE)

The FRE score measures how easy a text is to read. It combines sentence length and syllable count, with higher scores indicating simpler, more accessible text. The formula is:

$$FRE = 206.835 - 1.015\left(\frac{words}{sentences}\right) - 84.6\left(\frac{syllables}{words}\right)$$

Gunning Fog Index

This index estimates the years of education required to understand a text on first reading. It factors in sentence length and the percentage of complex words (words with three or more syllables). Lower Fog scores indicate easier text. The formula is:

$$Fog = 0.4\left[\left(\frac{words}{sentences}\right) + 100\left(\frac{complex\ words}{words}\right)\right]$$

Age of Acquisition (AoA) – Average

We compute the average AoA for content words (excluding stop words) using ratings from Brysbaert et al. [1]. Lower AoA scores suggest vocabulary learned at a younger age and thus higher accessibility for children.

Age of Acquisition (AoA) – Threshold

This metric calculates the proportion of words with AoA ratings above 13 years old, excluding stop words, to identify potentially advanced vocabulary beyond the target child age range.

Conclusion

Persona-Based Prompting consistently improved readability and accessibility across all models and metrics, outperforming both Standard and Intended-User strategies. Although DeepSeek was an exception, the results show that telling the LLM to take on a persona helps create content that is better suited for children. We see this as a step toward developing practical guidelines for directing LLM behavior and equipping designers with tools to fine-tune model outputs in child-facing applications.

References

[1] Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior Research Methods*, 46(3):904–911, 2014. doi: 10.3758/s13428-012-0210-4.

[2] Donya Rooein, Amanda Cercas Curry, and Dirk Hovy. Know your audience: Do llms adapt to different age and education levels? *arXiv preprint arXiv:2312.02065*, 2023. URL https://arxiv.org/abs/2312.02065.