ReadEasy: Bridging Reading Accessibility Gaps using Responsible Multimodal Simplification with Generative AI*

Sharv Murgai
Monta Vista High School
ReadEasy.org
murgai.sharv@gmail.com

Shivatmica Murgai
Stanford University
ReadEasy.org
smurgai@stanford.edu

Abstract

We present ReadEasy.org, a multimodal, retrieval-augmented system that simplifies text and images for improved accessibility in education, healthcare, and technical domains. The system integrates Age-of-Acquisition guidance, word-sense disambiguation, graph-based retrieval-augmented generation (RAG), image captioning, and a human-in-the-loop feedback loop. Across 14,000 items, it improves readability over GPT-4 baselines (+22.21% SARI, +14.11% Flesch), increases domain retrieval precision (+11%), and yields further gains with user feedback (+8% content relevance, +15% satisfaction). In classroom use with 200 K–12 students and additional professional cohorts, users rated outputs as easier to understand and more useful. This Creative Demo highlights how responsible AI design can support accessibility while maintaining semantic integrity.

1 Motivation

Complex content in education, healthcare, and technical fields often creates accessibility barriers. Existing LLM simplifiers lack adaptability, struggle with multimodal input, and fail to personalize outputs. Our demo system addresses these gaps with knowledge-grounded retrieval and interactive feedback.

2 System Overview

- **AoA + WSD:** Flags difficult terms and resolves ambiguities. Users can specify a **target age group**, and AoA ratings ensure that simplified text matches developmental reading levels.
- **Graph-based RAG:** Retrieves domain-specific definitions from medical/technical glossaries, improving precision by 11%.
- **Synonym Selection:** LLM + cosine similarity ensures semantic fidelity.
- **Image Captioning:** Simplifies medical diagrams/technical schematics for non-experts.
- **Feedback Loop:** Iterative refinement; +8% SARI, +15% satisfaction.

As illustrated in Figure 1, the system processes input text and images through AoA/WSD, graph-based RAG, synonym selection, and image captioning. A human-in-the-loop feedback loop continuously refines outputs for different target audiences.

*Project site: ReadEasy.org

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Generative and Protective AI for Content Creation (GenProCC).

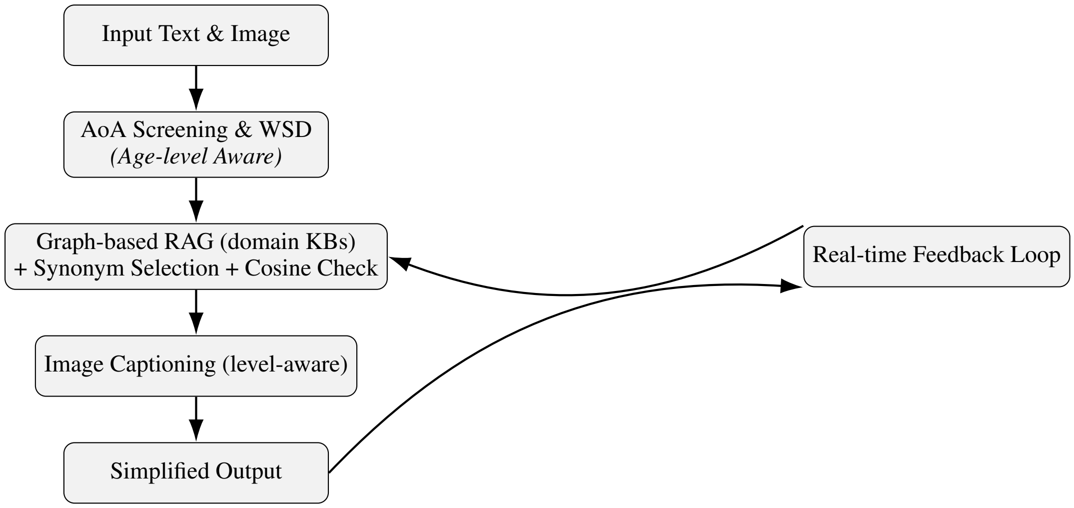

Figure 1: System flowchart: multimodal pipeline with AoA/WSD (*age-level aware*), graph-based RAG, synonym selection, image captioning, and a human-in-the-loop feedback loop.

Table 1: Illustrative simplifications across education (age-targeted) and domain-specific contexts.

Context / Model	**Simplified Text**
Education (Photosynthesis)	
Original	Photosynthesis is how green plants and some other organisms use light to create energy.
ChatGPT-4 (age 9)	Photosynthesis is when green plants and some other living things use sunlight to make food.
Proposed System (age 9)	Photosynthesis is like a magic trick that plants use to make their own food from sunlight.
Proposed System (age 7)	Photosynthesis is like a magic trick that green plants use to turn sunlight into food.

3 Demo Features

Users can:

1. Upload complex text and images.
2. Specify a **target age group** (e.g., 7, 9, 11 years old) to receive outputs tuned to that reading level.
3. Receive simplified outputs tailored by AoA and graph-based RAG.
4. View retrieved definitions alongside the original text.
5. Provide feedback to refine outputs in real time.

4 Results Snapshot

Human evaluations **K–12:** N=200 students, ages 5/7/9/11 with group sizes 45/52/48/55. **Professionals:** N=50 total (25 medical, 25 technical) via short preference surveys. We report 95% confidence intervals using nonparametric bootstrap over items (10,000 resamples), and treat SARI/Flesch improvements as *absolute* deltas over the GPT-4 baseline.

- **Education:** +22.21% SARI, +14.11% Flesch (200 K–12 students).
- **Healthcare/Technical:** Professionals reported improved clarity and 18% higher satisfaction with retrieved explanations.

5 Creative Impact

The demo shows how multimodal, knowledge-grounded, and feedback-driven AI can transform education, healthcare and technical fields by making complex content age-appropriate and accessible.

6 Availability

URL: https://readeasy.org
Video: https://www.youtube.com/watch?v=Uwfy2D4Tq5I

Ethics Statement

K–12 classroom activities were introduced and facilitated by teachers within normal instruction, following school policies. Teachers aggregated student ratings at the *class level* and shared only de-identified summaries (e.g., counts by age group) with the authors. No personal identifiers, contact information, detailed demographics, audio/video, or device data were collected or stored, and no student accounts were created by the study team. Adult professional participants provided consent before completing brief surveys; only de-identified responses were retained. The system is for informational use and not a substitute for professional advice (e.g., medical or clinical). To reduce potential harms from oversimplification, the interface preserves the original text, provides retrieved definitions, and allows user feedback to adjust outputs.

References

[1] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443, 2019.

[2] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

[3] Raman Chandrasekar, Christine Doran, and Srinivas Bangalore. Motivations and methods for text simplification. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, pages 1041–1044, 1996.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 4171–4186, 2019.

[5] Rudolf Flesch. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221–233, 1948.

[6] Braden Hancock, Antoine Bordes, Pierre-Emmanuel Mazare, and Jason Weston. Learning from dialogue after deployment: Feed yourself, chatbot! In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 3667–3684, 2019.

[7] Gautier Izacard and Edouard Grave. Leveraging passage retrieval with generative models for open domain question answering. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2021.

[8] Douwe Kiela, Edouard Grave, Armand Joulin, and Tomas Mikolov. Efficient large-scale multimodal classification. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2018.

[9] Victor Kuperman, Hans Stadthagen-Gonzalez, and Marc Brysbaert. Age-of-acquisition ratings for 30,000 english words. *Behavior Research Methods*, 44(4):978–990, 2012.

[10] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Naman Goyal, Vladimir Karpukhin, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[11] Shuang Long, Jun Ruan, Wenqi Zhang, Xiangjian He, Wenhao Wu, and Cong Yao. Scene text detection and recognition: The deep learning era. *International Journal of Computer Vision*, 129:161–184, 2021.

[12] Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[13] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 311–318, 2002.

[14] Kristian Woodsend and Mirella Lapata. Learning to simplify sentences with quasi-synchronous grammar and integer programming. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2011.

[15] Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. Optimizing statistical machine translation for text simplification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016.

[16] Daniel M Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. Fine-tuning language models from human preferences. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.