



SARA: Smart AI Reading Assistant for Reading Comprehension

Enkeleda Thaqi
enkeleda.thaqi@tum.de
Technical University of Munich
Munich, Germany

Mohamed Mantawy
mohamed.mantawy@tum.de
Technical University of Munich
Munich, Germany

Enkelejda Kasneci
enkelejda.kasneci@tum.de
Technical University of Munich
Munich, Germany

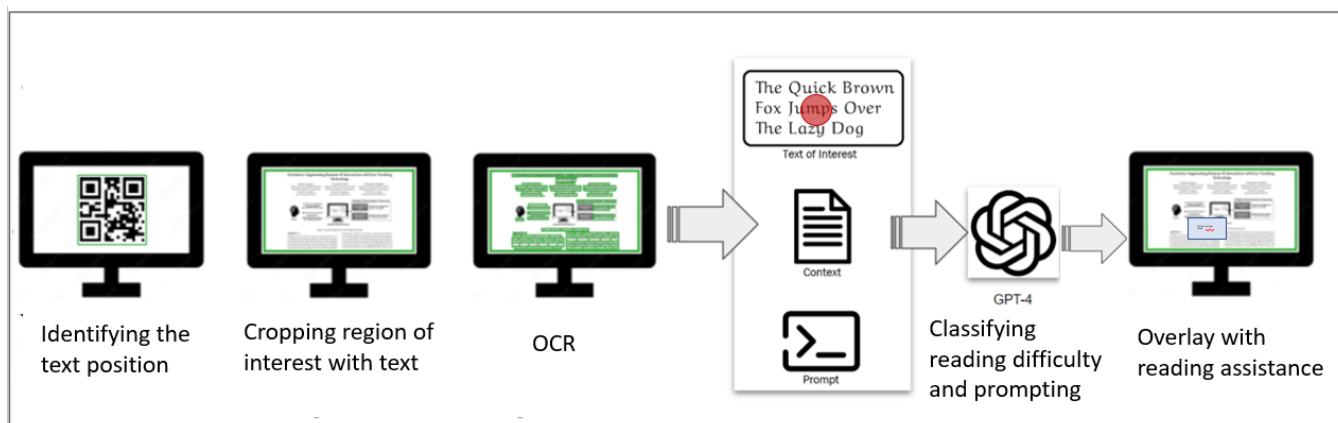


Figure 1: Illustration of key components SARA uses: text position identification, OCR, reading difficulty classification, prompt initiation for necessary help, and displaying assistance within the user's field of view. The workflow of the components is exemplified using the paper by [Langner et al. 2023] as reading material.

ABSTRACT

SARA integrates Eye Tracking and state-of-the-art large language models in a mixed reality framework to enhance the reading experience by providing personalized assistance in real-time. By tracking eye movements, SARA identifies the text segments that attract the user's attention the most and potentially indicate uncertain areas and comprehension issues. The process involves these key steps: text detection and extraction, gaze tracking and alignment, and assessment of detected reading difficulty. The results are customized solutions presented directly within the user's field of view as virtual overlays on identified difficult text areas. This support enables users to overcome challenges like unfamiliar vocabulary and complex sentences by offering additional context, rephrased solutions, and multilingual help. SARA's innovative approach demonstrates it has the potential to transform the reading experience and improve reading proficiency.

CCS CONCEPTS

• **Human-centered computing** → **HCI theory, concepts and models; Interactive systems and tools;**

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ETRA '24, June 04–07, 2024, Glasgow, United Kingdom

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0607-3/24/06

<https://doi.org/10.1145/3649902.3655661>

KEYWORDS

Eye tracking, Augmented Reality, Large Language Models, Reading Comprehension

ACM Reference Format:

Enkeleda Thaqi, Mohamed Mantawy, and Enkelejda Kasneci. 2024. SARA: Smart AI Reading Assistant for Reading Comprehension. In *2024 Symposium on Eye Tracking Research and Applications (ETRA '24)*, June 04–07, 2024, Glasgow, United Kingdom. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3649902.3655661>

1 INTRODUCTION

Technological innovation is continuously redefining our interaction with our environment. Mixed reality (MR) stands out within these advances as a particularly promising framework in which physical and virtual dimensions blend seamlessly to create deeply immersive experiences [Gardony et al. 2020]. Within this context, the development of the Smart AI Reading Assistant (SARA) represents a significant breakthrough that aims to improve the reading experience. SARA addresses common reading challenges by leveraging state-of-the-art head-mounted display (HMD) devices and advanced large language models (LLMs). This work explores the SARA application, demonstrating its technological framework and highlighting its ability to detect reading problems, extract text from the user's field of view, and provide personalized real-time support tailored to individual needs. Through a comprehensive exploration of SARA's functionalities, this research aims to show the potential of eye tracking, LLMs, and MR technology in improving reading experiences and increasing educational efficiency.

2 RELATED WORK

2.1 Eye Tracking in Reading Studies.

Eye-tracking technology is used in research studies to explore reading processes. It is commonly applied to investigate predictive abilities in reading comprehension [Mézière et al. 2023], the cognitive load of reading and processing of read text [Delgado and Salmerón 2022], and for the diagnosis of reading difficulties such as dyslexia [Nilsson Benfatto et al. 2016]. Additionally, eye tracking serves as a suitable technology for the treatment and improvement of reading skills [Caldani et al. 2020].

2.2 LLMs and AI-based tools in Education.

[Kasneci et al. 2023] explores the integration of LLMs into educational settings and highlights the potential advantages and challenges. The paper emphasizes the significance of critical thinking skills and advocates for clear strategies and pedagogical approaches to fully exploit the educational potential of LLMs while managing their complexity and limitations. [Van Wyk 2024] concludes that using AI-based tools presents an opportunity for innovative pedagogical approaches for students.

2.3 AR/MR in Education.

In medicine, eye tracking and MR frameworks are used to train medical students and professionals on surgical skills [Lu et al. 2020]. In science, developing such frameworks enables the evaluation of novice and expert problem-solving skills during experiments and training novices based on the results [Sonntag and Bodensiek 2021].

3 METHODS

SARA was implemented and deployed for HoloLens2 [hol 2024]. This MR-HMD device was chosen for its embedded features, such as integrated eye tracking, a built-in scene camera, and publicly available toolkits for Unity. These features provide a robust set of tools and resources that make it possible to provide a comprehensive solution for reading assistance in the MR environment. The application operates through the following steps that are illustrated in Figure 1:

- (1) **Text position identification.** A QR code is displayed, and when detected and decoded, augmented virtual markers are placed in their respective positions in the real world. This approach guarantees the accurate identification of the text position containing the reading material the user is currently engaged with.
- (2) **Text extraction.** In a multi-step approach, frames are taken by the HoloLens2 scene camera to capture visual data from the user's field of view for text processing. The focus is then narrowed down to the region of interest surrounding the virtual screen placed based on the QR code. The captured frames are then cropped to isolate the area containing the text being read.
- (3) **Optical Character Recognition (OCR).** OCR is used to analyze the cropped images and extract text content. This process not only identifies text in the captured images but also generates bounding boxes around each recognized text

element, providing precise information on the location of the text.

- (4) **Gaze tracking for focus detection.** The user's gaze is captured in the MR environment, providing ray information indicating the location of the user's focus on the text region, including the direction of gaze and subsequent eye movements.
- (5) **Gaze alignment.** In this step, a series of coordinate transformations are performed. First, the point coordinates at which the user's gaze intersects with the screen are logged in relation to the world coordinates. Then, the coordinates of this point relative to the center of the virtual screen are calculated by subtracting the position of the virtual screen from the gaze intersection point. This step helps to create a local coordinate system centered around the virtual screen. Next, the intersection point, which is now in the coordinate system of the virtual screen, is transformed into the coordinate system of the cropped image by scaling the point based on the ratio between the width and height of the virtual screen in the world coordinate system and the width and height of the cropped camera image in the pixel coordinate system. This scaling ensures that the transformed coordinates accurately represent the position of the user's gaze on the region of interest in pixel units. Finally, the transformed coordinates are further adjusted to align with the top-left axis of the screen. To do this, the x-coordinate is shifted by half the width of the virtual screen in pixels, and the y-coordinate is mirrored to match the pixel orientation of the cropped image. The resulting transformed coordinates represent the exact pixel position of the user's gaze on the screen and allow the exact mapping between the user's eye gaze and the position of the words in the text material.
- (6) **Classifying reading difficulty.** SARA addresses two main challenges in reading: encountering unfamiliar words and difficulties comprehending paragraphs. Based on gaze dwell time, SARA identifies instances where users spend more time on unfamiliar words compared to the surrounding text, indicating potential difficulties. In addition, SARA detects regressions in reading patterns where users exhibit eye movements that do not follow the standard left-to-right and top-to-bottom progression. By using these detection techniques, SARA can identify areas of reading difficulty and offer immediate assistance.
- (7) **Reading Support.** SARA uses the advanced features of state-of-the-art GPT-4 to overcome user's reading difficulties. For an unfamiliar word, SARA prompts a definition in context or a translation according to the user's needs to ensure accurate comprehension. Offering contextual definitions or translations improves comprehension and enables a smoother reading experience. When users have difficulty understanding paragraphs, SARA prompts for simplifications and paraphrasing. This approach aims to enhance comprehension by presenting complex text in a more accessible way, which ultimately helps users to understand the content more efficiently. Finally, SARA seamlessly integrates these results into virtual components within the user's augmented reality environment.

4 SUMMARY AND FUTURE WORK

SARA represents a significant advancement in mixed reality technology for educational support in reading. Using tools such as the HoloLens2 and GPT-4, it provides personalized, real-time assistance tailored to individual reading needs. Through its methodology, including identifying text position, detecting difficulty, and providing targeted support, SARA enhances reading experiences and overcomes comprehension difficulties. Ongoing research and development work aims to improve SARA's capabilities, address limitations, and assess its impact on user engagement and reading comprehension through extensive case studies. SARA demonstrates the potential of mixed reality and advanced language models in redefining reading assistance and advancing inclusive learning environments.

ACKNOWLEDGMENTS

We gratefully acknowledge IT-Stiftung Esslingen for their generous support of our hardware lab.

REFERENCES

2024. Microsoft HoloLens. <https://www.microsoft.com/de-de/hololens>. Accessed: March 18, 2024.
- Sarah Caldani, Claire L Gerard, Hugo Peyre, and Maria Pia Bucci. 2020. Visual Attentional Training Improves Reading Capabilities in Children with Dyslexia: An Eye Tracker Study During a Reading Task. *Brain Sciences* 10, 8 (2020), 558. <https://doi.org/10.3390/brainsci10080558>
- Pablo Delgado and Ladislao Salmerón. 2022. Cognitive Effort in Text Processing and Reading Comprehension in Print and on Tablet: An Eye-Tracking Study. *Discourse Processes* 59, 4 (2022), 237–274. <https://doi.org/10.1080/0163853X.2022.2030157> arXiv:<https://doi.org/10.1080/0163853X.2022.2030157>
- Aaron L. Gardony, Robert W. Lindeman, and Tad T. Brunyé. 2020. Eye-tracking for human-centered mixed reality: promises and challenges. In *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR)*, Bernard C. Kress and Christophe Peroz (Eds.), Vol. 11310. International Society for Optics and Photonics, SPIE, 113100T. <https://doi.org/10.1117/12.2542699>
- Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* 103 (2023), 102274.
- Moritz Langner, Peyman Toreini, and Alexander Maedche. 2023. A Situation Awareness Perspective on Intelligent Writing Assistants: Augmenting Human-AI Interaction with Eye Tracking Technology. (2023).
- Shang Lu, Yerly Perdomo, Xianta Jiang, and Bin Zheng. 2020. Integrating Eye-Tracking to Augmented Reality System for Surgical Training. *Journal of medical systems* 44 (09 2020), 192. <https://doi.org/10.1007/s10916-020-01656-w>
- Diane C. Mézière, Lili Yu, Erik D. Reichle, Titus von der Malsburg, and Genevieve McArthur. 2023. Using Eye-Tracking Measures to Predict Reading Comprehension. *Reading Research Quarterly* 58, 3 (2023), 425–449. <https://doi.org/10.1002/rrq.498> arXiv:<https://doi.org/10.1002/rrq.498>
- Mattias Nilsson Benfatto, Gustaf Öqvist Seimyr, Jan Ygge, Tony Pansell, Agneta Rydberg, and Christer Jacobson. 2016. Screening for Dyslexia Using Eye Tracking during Reading. *PLOS ONE* 11, 12 (12 2016), 1–16. <https://doi.org/10.1371/journal.pone.0165508>
- D. Sonntag and O. Bodensiek. 2021. EYE-TRACKING-BASED DESIGN OF MIXED REALITY LEARNING ENVIRONMENTS IN STEM. In *INTED2021 Proceedings (Online Conference) (15th International Technology, Education and Development Conference)*. IATED, 9519–9524. <https://doi.org/10.21125/inted.2021.1990>
- Micheal M Van Wyk. 2024. Is ChatGPT an Opportunity or a Threat? Preventive Strategies Employed by Academics Related to a GenAI-based LLM at a Faculty of Education. *Vol. 7 No. 1* (2024). <https://doi.org/10.37074/jalt.2024.7.1.15>