Student name: Michael Rice

Student number: 20347541

Supervisor contact name: Waqar Shahid Qureshi

Supervisor contact email: waqarshahid.qureshi@universityofgalway.ie

From Sight to Insight: Harnessing Gaze and LLMs to provide Real-Time Educational Assistance

Introduction

Advancements in wearable technology and artificial intelligence have opened new possibilities for real-time educational assistance. This project leverages Meta's Project Aria [1] glasses, object detection, and large language models (LLMs) to create a gaze-based system that provides users with contextual information about the objects they focus on. By integrating egocentric gaze tracking with AI-driven content generation, this system aims to enhance learning in various domains.

Background Research or Context

A wide variety of regions will require extensive background research in order to complete this project. Firstly, I will need to understand the hardware, the Project Aria Glasses, that I will be using, their usefulness to my project and the proprietary file system that comes with them, VRS. Following this, I will research object detection options, along with each of their strengths and weaknesses. Finally, research into Large Language Models (LLMs) will be required. This will involve understanding the various open-source and viable LLM options available and how they can be refined for my specific use case.

Proposed Project

This project will develop a real-time educational aid that identifies objects in a user's field of view and provides relevant information using LLMs. The system will rely on Project Aria's eye tracking cameras and gaze estimation to determine where the user is looking. An object detection algorithm (e.g., YOLO [2] or Detectron2) will then classify the object, and an LLM will generate concise, context-aware descriptions or explanations.

The key components of the system include: Gaze Tracking, Object Detection and Recognition, and Contextual Information Retrieval.

This tool has applications in education, professional training, and even specialized fields such as agriculture, where it can provide detailed insights into plant species or farming equipment based on gaze detection.

Timeline

February - March 2025: Literature Research and Getting Project Aria Glasses up and running.

April - May 2025: Creating base implementations of each algorithmic aspect.

June - July 2025: Move from base implementations to final models and integrating them all together.

August 2025: Tidying up final implementation and writing thesis.

References

[1] J. Engel, K. Somasundaram, M. Goesele, A. Sun, A. Gamino, A. Turner, A. Talattof, A. Yuan, B. Souti, B. Meredith, C. Peng, C. Sweeney, C. Wilson, D. Barnes, D. DeTone, D. Caruso, D. Valleroy, D. Ginjupalli, D. Frost, E. Miller, E. Mueggler, E. Oleinik, F. Zhang, G. Somasundaram, G. Solaira, H. Lanaras, H. Howard-Jenkins, H. Tang, H. J. Kim, J. Rivera, J. Luo, J. Dong, J. Straub, K. Bailey, K. Eckenhoff, L. Ma, L. Pesqueira, M. Schwesinger, M. Monge, N. Yang, N. Charron, N. Raina, O. Parkhi, P. Borschowa, P. Moulon, P. Gupta, R. Mur-Artal, R. Pennington, S. Kulkarni, S. Miglani, S. Gondi, S. Solanki, S. Diener, S. Cheng, S. Green, S. Saarinen, S. Patra, T. Mourikis, T. Whelan, T. Singh, V. Balntas, V. Baiyya, W. Dreewes, X. Pan, Y. Lou, Y. Zhao, Y. Mansour, Y. Zou, Z. Lv, Z. Wang, M. Yan, C. Ren, R. D. Nardi, and R. Newcombe, "Project Aria: A New Tool for Egocentric Multi-Modal AI Research," Oct. 2023.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," May 2016.