EnAct: A Safety-Aware Actionable Guidance System via Multi-Agent Vision-Language Models for People With Vision Impairments

Zhang, Amy (School: Lakewood High School)

Independent living for vision-impaired individuals requires them to perform daily object-reaching tasks, such as grasping an apple to eat. But without precise spatial awareness, even simple actions like reaching for something can lead to accidentally knocking items over, creating messes, or even posing safety risks. While current assistive technologies can describe scenes, they often do not provide enough information when it comes to guiding physical interaction, so most vision-impaired individuals still depend on other people's help for these tasks, limiting their independence. In this project, I developed a safety-aware actionable guidance system via multi-agent vision-language models. This system builds upon the capabilities of VLMs by adding 3D action intelligence, providing real-time action guidance/correction to guide vision-impaired individuals for object-reaching tasks effectively. For pre-motion high-level path planning, a multi-agent VLMs architecture enables individual agents to perform user query reasoning, spatial reasoning, safety assessment, and path planning, respectively. For in-motion low-level action guidance, the system enables real-time 3D reach status checking, collision avoidance and direction guidance all the way until reaching target using YOLO8-World and Depth Pro. This research bridges the knowledge gap in the VLM-based generic scene description and 3D spatial action intelligence. It allows vision-impaired users to follow its executable guidance to interact with the environment, enabling activities such as cooking, cleaning, and more, and thus improves independence and quality of life for vision-impaired individuals in a way existing assistive technologies are not able to.

Awards Won: