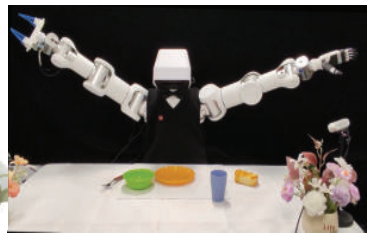
HONG KONG
CENTRE FOR LOGISTICS ROBOTICS
香港物流機械人研究中心

VLM-Powered Embodied AI System

The VLM-Powered Embodied AI System represents a next-generation robotic platform that leverages large-scale vision-language models (VLMs) for semantic perception, reasoning, and embodied decision-making. By unifying multimodal sensing with VLM-powered intelligence, the system achieves human-level understanding of tasks and environments, enabling robust operation in logistics, healthcare, smart homes, and advanced manufacturing.



Innovations and Technologies

- Multi-Modal Foundation Model Integration: Embeds vision-language models into the robotic perception–action loop, allowing semantic grounding of tasks, high-level reasoning, and natural language–based command execution
- Multi-Modal 3D Sensor: IIntegrates Stripe and Speckle structured light to enable precise 3D imaging, employs a monocular AI model to address challenging materials and complex scenes
- Closed-loop Embodied Intelligence System: Fuses sensor feedback, motion control, and task planning to support autonomous learning and adaptation in unpredictable environments

Target Applications

- Domestic Assistance: Supports household tasks, kitchen operations, and interactive functions in smart home environments
- Logistics & Warehousing: VLM-based semantic sorting, collaborative handling, and natural language coordination among multiple robots
- Industrial Collaboration: Facilitates flexible loading/unloading, precision assembly, and seamless human-robot cooperation in manufacturing settings