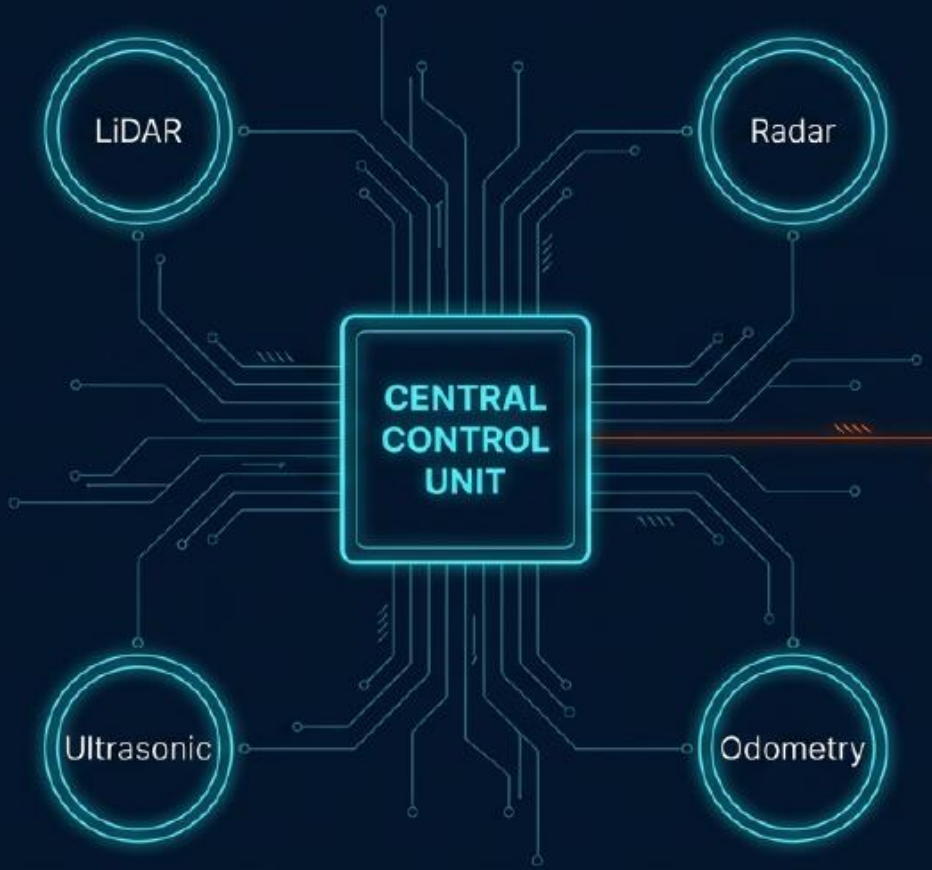
Present and future of AGVs:

How do **visual and language understanding** fit in the landscape?

Scope, risks and promises

Team 3

CURRENT ARCHITECTURE: THE SENSOR FUSION BASELINE



- **Multi-Modal Dependency:** Rigid fusion of LiDAR, Radar, and Camera.
- **Control Logic:** Centralized master governance.
- **Critical Constraint:** Millisecond-level processing required for safety.

LATENCY SENSITIVITY

(Lynch et al., 2018; VDA 5050, 2025)

The Energy Price of Being Smart

The Cycle Never Ends

Cycles of work are increasing. AGVs are now expected to work 24 hours a day to meet delivery speeds.

(IndustryARC, 2024)

Eyes Cost Power

Sensors like cameras and Lidar act as “eyes.” Keeping these eyes open constantly to avoid crashes uses a lot of electricity.

(Zhang, Chen, & Guo, 2022)

Thinking Burns Energy

The onboard AI models used for navigation and decision-making need heavy computing power. This “thinking” process reduces the driving range.

(Lin et al., 2018)

AGV current and future challenges

1. Immediate: Energy & Thermal Limits

2. Immediate: Sensor fusion and Real-Time Latency

3. Future: Flexibility & Human-Robot Interaction

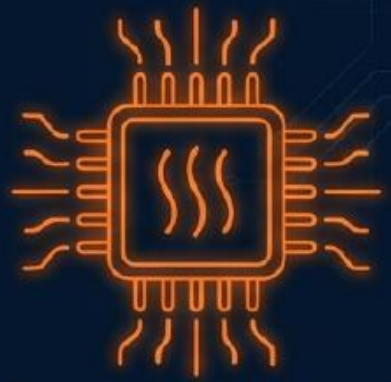
CRITICAL RISKS & COMPLEXITY

Latency & Reaction

VLM reasoning is computationally heavy. Processing lag (**>100ms**) risks collision in dynamic environments. **Target >10 FPS** is difficult.

Hallucinations

AI model misinterpretation. Identifying shadows as obstacles or failing to see transparent surfaces (glass).

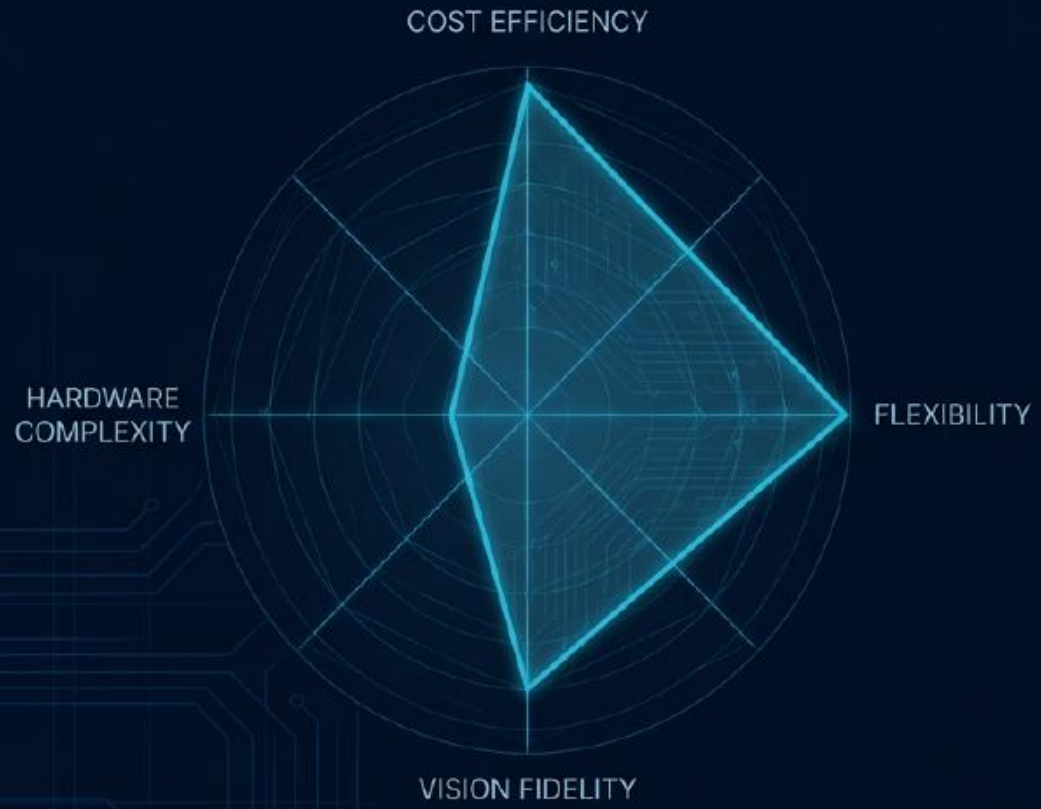


Processing Load

Thermal bottlenecks. Massive compute requirements drain battery and require complex cooling solutions.

(Real-time open-vocabulary perception, 2025; Challenges Facing Warehouse AGV Manufacturers, 2025)

STRATEGIC TARGET: VISION-FIRST ARCHITECTURE



CORE CAPABILITIES

- **Camera-First Design**: Removing expensive LiDAR for high-fidelity image integration.

- Zero-Shot Task Execution: Instruction-based navigation without pre-mapping (e.g., "Find the exit").

- Human-Robot Interaction: Natural language understanding via Vision-Language Models (VLMs).

(Analog Devices, 2025; History-Augmented Vision-Language Models, 2025)

Final remarks

1. **Prototype for the future (seminal stage):**

2. **Sharpen the current challenges and add some news:**

   Battery, Latency, Hallucinations

3. **Simulation environment to address issues out of scope:**

   Battery, Latency, damages, AGV communication.

REFERENCES

Analog Devices. (2025, Sept 24). Fine-Tuning Vision Language Models (VLM) for Agile Robotics.

(Author Unknown). (2025). Autonomous Forklifts: State of the Art—Exploring Perception... MDPI.

(Author Unknown). (2025). Real-time open-vocabulary perception for mobile robots... PMC.

IndustryARC. (2024). AGV AMR Market (2024-2030).

Interact Analysis. (2024). Interoperable Robots to Take Over by 2035.

MarketsandMarkets. (2025). Autonomous Mobile Robots (AMR) Market Size, Share and Trends.

NMB Technologies Corporation. (2025). Challenges Facing Warehouse AGV Manufacturers.

TrendMinds. (2026). Cloud Robotics: Use Cases & Architecture Explained 2026.

Verband der Automobilindustrie (VDA). (2025). VDA 5050 Version 2.1.0.