The industry is transitioning from fixed-path Automated Guided Vehicles (AGVs) to Autonomous Mobile Robots (AMRs) with cognitive capabilities. AMRs will gain market share in the next 10 years.

(Interact Analysis, 2026; Fraifer et al., 2025).

VLM-driven High-Level Navigation from AGV to AMR

Pi-Wei Chen
Donato Cerciello
Lingyu Qiu
Juan Camilo España
Tomasz Skowron

Agenda

- Background and motivations
- The path to the solution
- Architecture
- Prototype
- Status of research
- Future opportunities

Background and motivations

Precision & Multi-Modal Awareness

2025–2027

Millimeter positioning and 3D mapping enable navigation in complex, crowded environments.

2026–2029

Operational Efficiency & Sustainability

Self-correcting navigation and edge intelligence optimize energy use and reaction times.

Fleet Orchestration & Collective Intelligence

2028–2031

Robots learn and share shortcuts across the fleet without central control.

2030–2033

Human-Centric Social Navigation

Predictive safety models allow robots to anticipate human paths and yield naturally.

Fluent Cognitive Interaction

2032–2036+

Real-time reasoning enables verbal task execution and two-way human-machine feedback.

Key Transition: Mobile robotics from tactical tools to socially aware, cognitive partners

The path to the solution:

Initial Phase: Camera-Only Foundation

Focused on minimizing sensor costs and avoiding the complexities of data fusion.

The VLA Bottleneck

Image-to-action models require excessive training time and are often optimized for 7D arms rather than mobile bases.

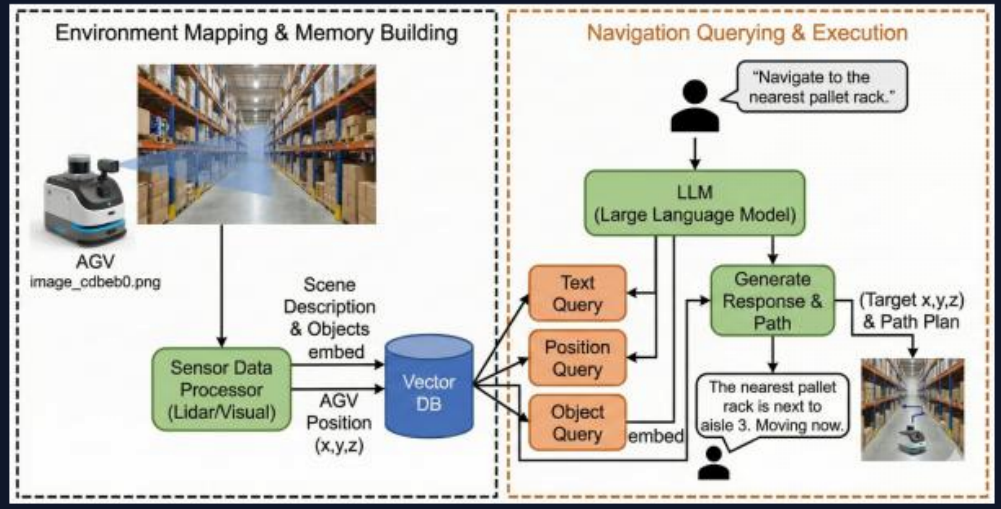
YOLO as Safety Guardrail

Integrated to detect obstacles, yet currently lacks the standalone accuracy required for primary navigation.

Final State: VLM + LiDAR Fusion

Combines semantic reasoning with precise 3D environmental awareness for maximum reliability.

Architecture of solution



Multi-Layer Pipeline

Raw sensor data is abstracted into semantic text before being processed by the VLM

Cognitive Decoupling

Separates high-level strategic reasoning from low-level geometric calculations to maximize safety.

LEVEL 1: Perception & Safety

- Sensor Input: 2D LiDAR + Odometry
- Zone-based Semantic Abstraction
- Safety Watchdog (High Priority)

Semantic Bridge (Text Prompt)

LEVEL 2: VLM Strategic Decision

- LLM Reasoning (Gemini API)
- Strategic Output: [Dir, Offset, Dist]

Cognitive Reasoning Phase

Trajectory Parameters

LEVEL 3: Execution & Control

- Parametric Waypoint (Bezier) Gen
- Pure Pursuit & PID Stabilization
- **Final Output: Wheel Velocities**

60Hz EMERGENCY OVERRIDE

Level 1: Perception and Semantic Abstraction

Nature of LiDAR Raw Data

LiDAR outputs a Point Cloud list containing N points of (x, y, z) coordinates. It uses the AGV local coordinate system (Robot as origin). For example, (0.0, -1.0) represents a point 1 meter directly in front (assuming -Y is forward)

Raw Data Challenges

Thousands of geometric numbers are unintelligible noise to LLMs and cause Token Explosion. LLMs cannot perform direct high-dimensional spatial reasoning, necessitating a "Semantic Abstraction" bridge.

Semantic Bridge

Through zone segmentation and feature extraction, coordinates are transformed into logical descriptions. LLM performs Logical Judgement instead of geometry math: Front Blocked + Right Clear = Turn Right.

Level 1: Perception and Semantic Abstraction

Semantic Bridge Strategy

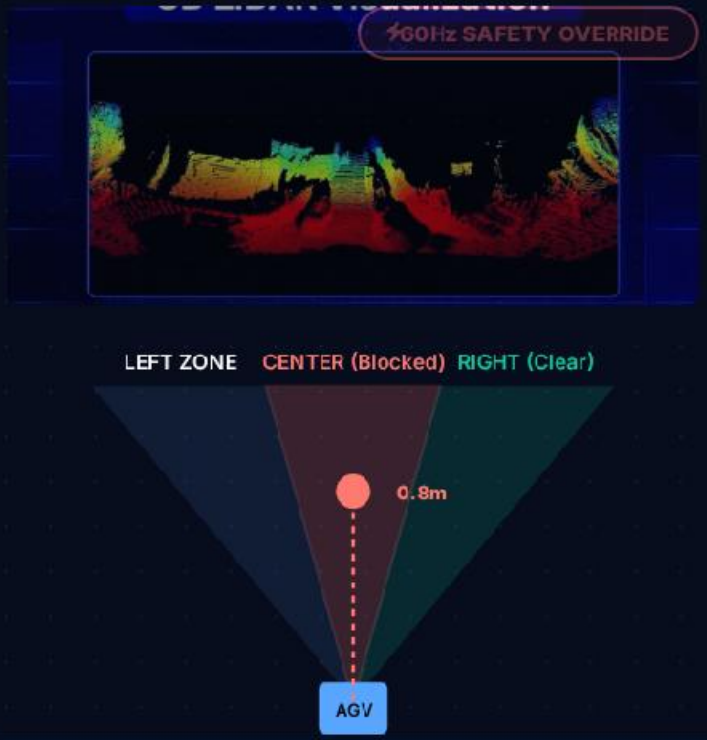
Instead of raw point clouds, we perform **Zone-based Abstraction**. High-dimensional data is summarized into key metrics (Proximity & Traversable Width) to prevent LLM hallucinations and token overflow.

60Hz Safety Watchdog

Independent Safety Monitor detects imminent risks at high frequency. If the Emergency Stop Distance is breached, it overrides all cognitive commands to halt the robot immediately.

Level 1 Output (Semantic Prompt):

```
"Obstacle detected 0.8m ahead in Center Zone; path blocked.
wide clearance available on the right (Traversable: 2.5m)."
```



Level 2: LLM-Driven Decision Making

Strategic Reasoning via Gemini

Leverages the Gemini API to resolve complex avoidance scenarios that paralyze traditional rule-based planners. The LLM acts as the high-level brain for spatial context evaluation.

Strategy vs. Computation

Unlike RL agents outputting discrete motor steps, we decouple strategic intent from geometric math. The LLM decides "what to do," leaving "how to track" to Level 3.

Level 2: LLM-Driven Decision Making

✕ Why not output coordinate Waypoints?

LLM is essentially a language model rather than a mathematical engine and is insensitive to absolute numerical values. If it is required to output path points, it is extremely easy to produce geometric illusions, resulting in the collapse of the navigation system.

√ Structured Output (Parametric)

We require LLM to output "decision parameters" rather than a specific path. LLM is responsible for qualitatively determining the direction (left/right) and intuitively understanding the safety margin, leaving the precise trajectory calculation to downstream algorithms.

```
{
   "reasoning": "Obstructed ahead and with ample space to
    the right, therefore detour to the right."
    "avoidance_direction": "right", // Strategic direction
    "lateral_offset": 1.5, // Offset (meters)
    "pass_distance": 3.0 // Detour length (meters)
}
```

L3: TRAJECTORY GENERATION & EXECUTION

`L2 INPUT: {dir, offset, dist}`

Parametric Path Generation

The Parametric Waypoint Generator acts as the mathematical bridge, converting abstract strategic intent into local target points.

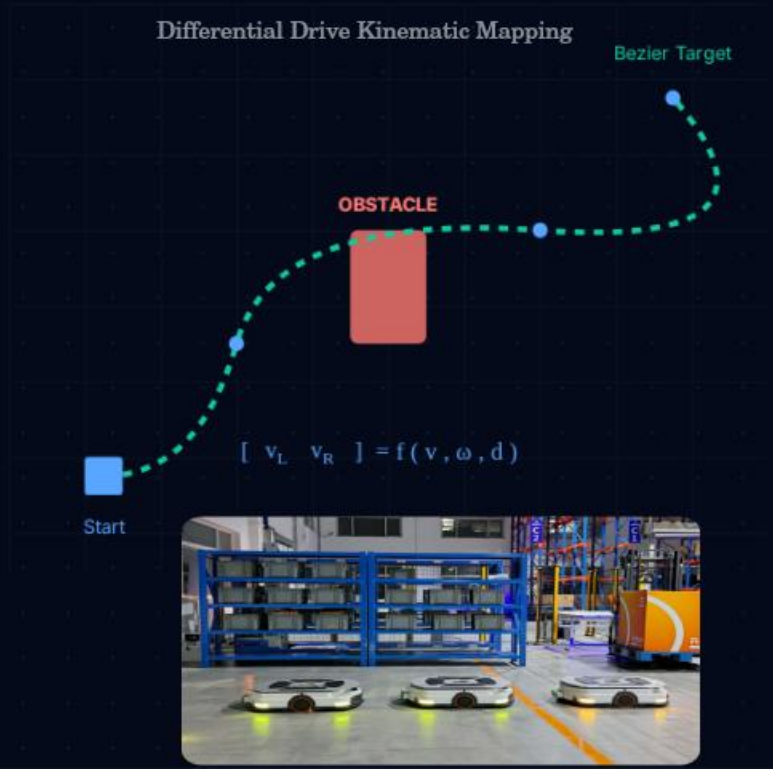
- **Bezier Trajectory:** Generates a smooth, C^2 continuous curve to bypass obstacles.
- **Local Integration:** Seamlessly merges local avoidance waypoints into the global navigation module.

Motion Control & Actuation

Ensuring deterministic tracking within the physics simulation environment through robust control loops.

Pure Pursuit: Look-ahead algorithm for geometric path tracking.

PID Controller: Corrects linear and angular velocity errors in real-time.



NVIDIA Isaac Sim: Advancing Robotics Development

NVIDIA Isaac Sim, built on the Omniverse platform, provides a powerful and extensible simulation environment for robotics. It accelerates development, testing, and deployment of AI-powered robots by offering physically accurate environments and realistic sensor data.

Realistic Simulation

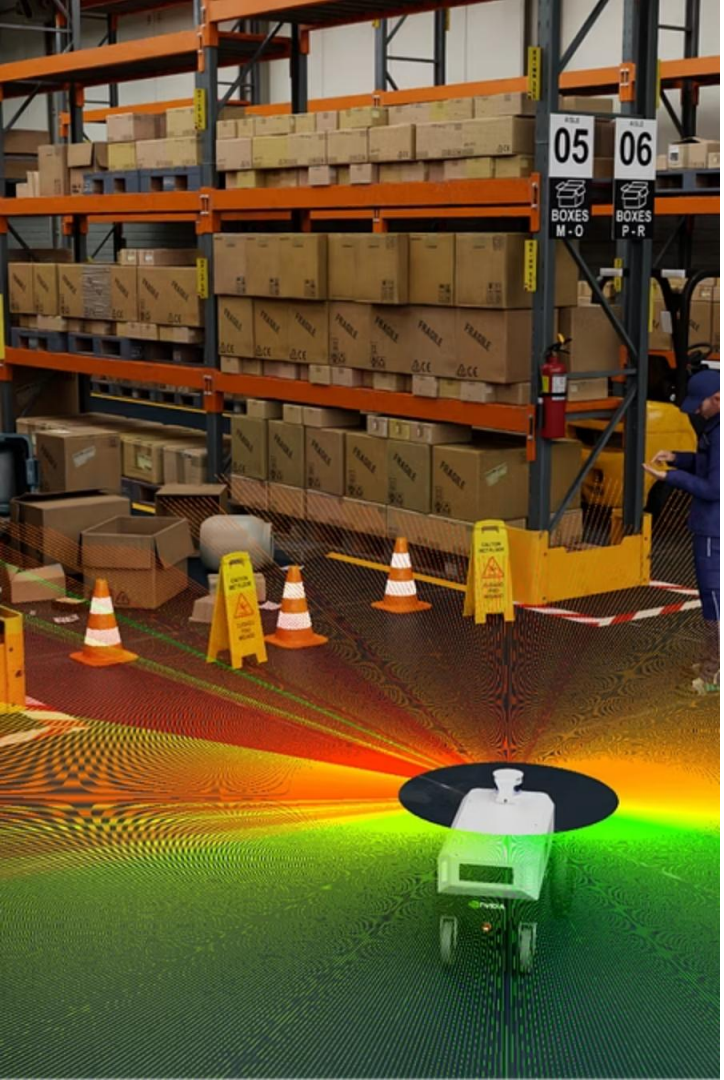
High-fidelity physics and rendering create accurate digital twins of robots and environments.

Synthetic Data Generation

Generate vast amounts of diverse, labeled data for training robust AI models.

Scalable Development

Utilize GPU-accelerated computing for rapid iteration and testing of robotic applications.



Dynamic AGV Scenario Generation

Our approach leverages Isaac Sim's capabilities to dynamically generate unique AGV scenarios. This ensures a rich and varied dataset for robust AI training.

Randomized Environments

Each simulation run creates a new, distinct environment with varying obstacles and layouts, mimicking real-world unpredictability.

- Random obstacle placement
- Diverse path configurations
- Varied lighting conditions

Accelerated AI Training

By continuously generating novel synthetic data, we can train AI algorithms for AGVs more efficiently and safely, reducing the need for extensive physical testing.

- High-volume data generation
- Reduced real-world training risks
- Improved model generalization

AGV Simulation in Action

Tested on: NVIDIA GeForce RTX 5060 Laptop GPU

Witness our AGV navigating challenging, dynamically generated environments within Isaac Sim. These simulations demonstrate the effectiveness of our approach in creating robust and adaptable robotic systems.

Camera-Based Perception

The camera provides visual awareness for obstacle detection and navigation.

LiDAR-Based Sensing

See the AGV adapt its route in real-time to maintain efficiency and avoid collisions.

Output example


```
{
  "reasoning": "Right side has more space (99.9m vs Left 0.0m)",
  "avoidance_direction": "right",
  "lateral_offset": 1.5,
  "pass_distance": 3.0,
  "confidence": 0.85
}
```


Decision Parameter	Value	Description
Direction	right	Avoid to the right side
Lateral Offset	1.5m	How far to move sideways
Pass Distance	3.0m	How far to travel before clearing obstacle
Confidence	85%	Model's confidence in this decision

```
Generated Waypoints (World Coordinates):
  WP1: (-12.52, 8.25)  ← Begin lateral shift
  WP2: (-13.11, 9.26)  ← Continue avoidance maneuver
  WP3: (-14.24, 9.67)  ← Passing obstacle zone
  WP4: (-16.16, 8.78)  ← Return to centerline
  WP5: (-21.17, 17.84) ← Resume toward target
```

Current Research Milestone: Experimental Validation

- **Controlled Simulation:** Initial testing within high-fidelity virtual environments.
- **Low-Complexity Scenarios:** Navigation in structured environments with minimal dynamic variables.
- **Latency-Constrained Decision Making:** Processing speeds require optimization for real-time response.
- **Cognitive Reasoning:** Leveraging Vision-Language Models (VLM) for scene interpretation.
- **Prompt-Based Fusion:** Integrating LiDAR and visual data via semantic prompting.

Future Work & Optimization

- **System Efficiency:** Model distillation and latency reduction for optimized battery performance.
- **Global Multi-Camera Reasoning:** Integrating external camera feeds for holistic environmental awareness.
- **Embodied AI:** Utilizing Reinforcement Learning to evolve from VLM to Vision-Language-Action (VLA).
- **Stochastic Testing:** Deployment and validation in unstructured, high-complexity industrial environments.
- **Hardware Integration:** Transitioning from standalone prototypes to robust industrial architectures.

Strategic Inquiry: *How do we effectively incorporate VLM into the current AGV architecture?*

Future opportunities

STRATEGIC RESEARCH TRACK

Dedicated pipeline for advanced cognitive development.

INTELLIGENT OPERATIONAL LAYERS

Reasoning-based recovery & Natural Language HRI

SENSOR DECOUPLING

Leverage SOTA Vision, reduce specialized hardware dependency.

LMM INFORMATION FUSION

Synthesize multi-modal data, optimize total sensor costs.

DATA-DRIVEN EVOLUTION

Capture telemetry to train adaptive VLA models.

ARE WE READY TO LEAD THE COGNITIVE ERA OF AMR?

THANK YOU!

ANY QUESTIONS?