Agentic UAVs: LLM-Driven Autonomy with Integrated Tool-Calling and Cognitive Reasoning

SmartTech 2025 – AI and Autonomous Systems Track

Anis Koubaa[1] Khaled Gabr[2]

[1]AlfaisalX Center of Excellence, Alfaisal University, Saudi Arabia
[2]RIOTU Lab, Prince Sultan University, Saudi Arabia

akoubaa@alfaisal.edu

January 14–16, 2025 — Tunis, Tunisia

AlfaisalX: Center of Excellence

- **Institution**: Alfaisal University, Riyadh, KSA
- **Focus Areas**:
  - Cognitive Robotics & Autonomous Agents
  - UAV/Drone Technologies
  - AI & Large Language Models
- **Collaboration**: RIOTU Lab, Prince Sultan University
- **Mission**: Advancing intelligent aerial systems for defense, disaster response, and logistics

Problem Statement

The Challenge

UAVs are increasingly deployed in **defense, surveillance, and disaster response**, yet most remain limited to **SAE Level 2–3 autonomy**.

- **Reliance on rule-based control** restricts adaptability
- **Narrow AI** limits response in dynamic environments
- **No context-aware reasoning** for complex missions
- **No ecosystem integration** with external services

Research Question

How can we design a UAV architecture that fuses LLM-driven reasoning with real-time perception and ecosystem integration for general-purpose autonomy?

Gap Analysis: Current UAV Limitations

Existing approaches fall short:

Approach	Strength	Limitation
Classical Control (A*, RRT*)	Efficient, deterministic	Fails under uncertainty
Reinforcement Learning	Adaptive policies	Low-dimensional only
Swarm Intelligence	Distributed coordination	No cognitive reasoning
LLM Integration (REAL, UAV-VLN)	Semantic parsing	Isolated planner only

Critical Gap

No UAV system leverages LLM agents with tool-calling for:

- Real-time knowledge access and API interaction
- Ecosystem-level integration (databases, alerts)
- Peer-to-peer reasoning in multi-agent swarms

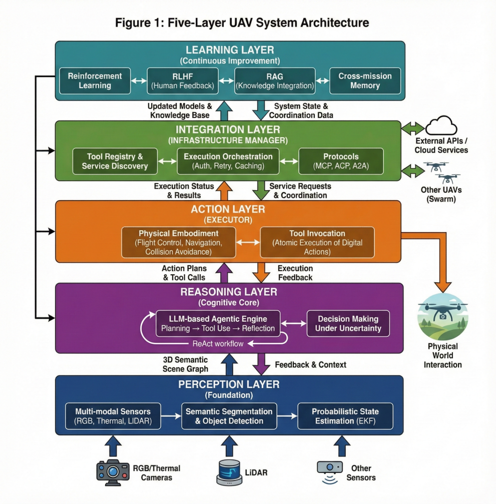
Key Contributions

1. **Architectural Novelty**
   - First **five-layer architecture** fusing LLM reasoning with continuous perception and flight control
2. **Ecosystem Integration**
   - Novel Integration Layer enabling tool-calling, API interaction, and multi-agent protocols (MCP, ACP, A2A)
3. **Collaborative Swarm Cognition**
   - First demonstration of distributed LLM-based reasoning for task negotiation among UAV swarms

Novelty

Transforms UAVs from **isolated planners** to **networked cognitive agents**

Methodology: Five-Layer Architecture



Reasoning Layer: LLM-Driven Cognition

ReAct Workflow

The LLM operates as a stateful agentic engine via LangGraph:

- **Planning**: Decomposes high-level goals into action sequences
- **Tool Use**: Invokes APIs, databases, weather services
- **Reflection**: Monitors outcomes, triggers replanning on failure

Example Output:


```
{"goal": "Inspect anomaly", "steps": [{"action": "call_tool", "tool":
"weather.get_forecast"}, {"action": "fly_to", "target": "Anomaly-01"}]}
```


Key Innovation

LLM decides *what* to do; Action Layer executes; Integration Layer manages *how*

Integration Layer: Ecosystem Connectivity

Transforms UAVs into networked digital actors:

- **Tool Registry**: Dynamic catalog of APIs, databases, services
- **Execution Orchestration**: Authentication, retry logic, caching
- **Protocol-Governed Communication**:
  - *MCP*: Model Context Protocol for secure LLM tool use
  - *ACP*: Agent Communication Protocol for UAV-to-cloud
  - *A2A*: Agent-to-Agent for peer negotiation
- **Security**: mTLS, JWT tokens, prompt injection detection

Multi-Agent Coordination

Task allocation, spatial deconfliction, shared situational awareness

Experimental Validation: Setup

Simulation Environment:

- **Platform**: Gazebo Harmonic + PX4 SITL v1.14.3
- **Middleware**: ROS 2 Humble
- **Perception**: YOLOv11 @ 30 Hz (Intel RealSense D455)
- **Reasoning**: GPT-4 API + Local Gemma-3 (4B parameters)
- **Hardware**: Intel i7-13900K, 32GB RAM, RTX 4070

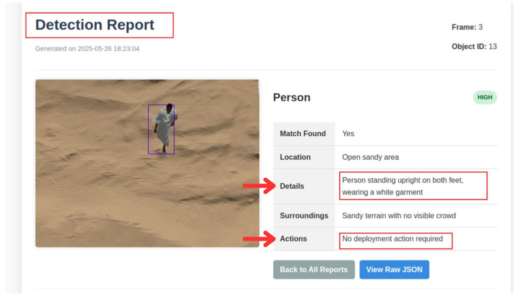
Scenario: Hajj Pilgrimage Search & Rescue

- **Scenario 1**: Normal activity monitoring (crowd surveillance)
- **Scenario 2**: Emergency medical intervention (collapsed person)

Dataset: $n = 44$ samples per system, $N = 132$ total detections

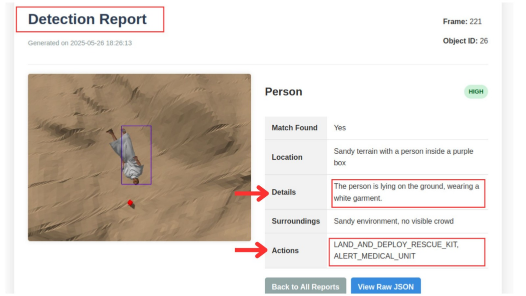
SAR Scenario Demonstrations

Scenario 1: Normal Monitoring



- Normal pilgrim activity detected
- Contextual understanding provided
- No intervention required

Scenario 2: Emergency Response



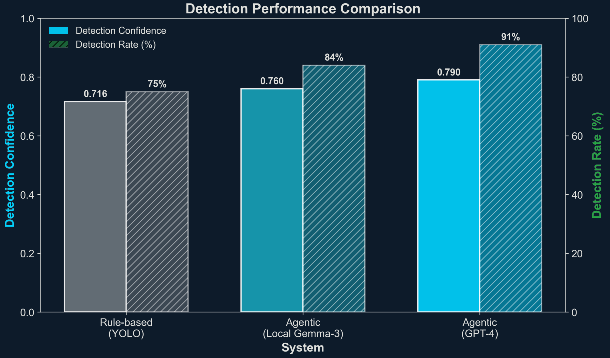
- Collapsed person detected (conf: 0.89)
- Emergency classified in ¡ 3 sec
- Alert + rescue kit deployed

Results: Detection Performance

Metric	YOLO	Local	GPT-4
Confidence	0.716	0.760	**0.790**
Detection	75%	84%	**91%**

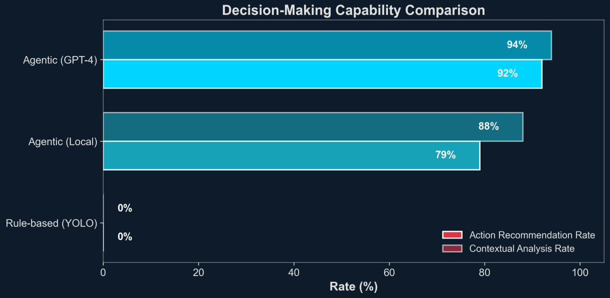
Key Results

+16% improvement in detection rate
ANOVA: $F(2, 129) \approx 3.96$, $p = 0.021$



Results: Decision-Making Capability

Metric	YOLO	Local	GPT-4
ARR	0%	79%	92%
CAR	0%	88%	94%



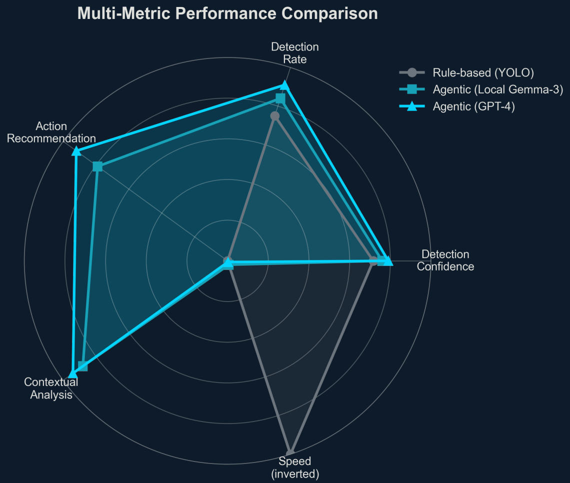
Critical Finding

Rule-based: **0% autonomy**
Agentic: **92% action rate**
$\chi^2 = 82.92,\ p \ll 0.001$

Results: Multi-Metric Comparison

Trade-off Analysis:

- YOLO excels at **speed only**
- Agentic systems dominate in:
  - Contextual understanding
  - Action recommendation
  - Ecosystem integration
- Local Gemma-3: **70% faster** than GPT-4 with acceptable quality



Recommendation

Hybrid architecture: YOLO for screening, LLM for critical cases

Main Findings

Finding 1: Superior Detection

Agentic UAVs achieve **0.79 confidence** vs 0.72 baseline, **91% detection rate** vs 75%

Finding 2: Autonomous Decision-Making

92% action recommendation rate (rule-based: 0%) – qualitatively new capability

Finding 3: Computational Trade-off Justified

Modest overhead ($\sim$5s) enables **contextual reasoning, ecosystem integration, and reduced operator dependence**

Overall: Computational cost is an *investment* that elevates UAVs from passive sensors to **intelligent agents**.

Future Work

- **Short-term**:
  - Field trials in controlled outdoor environments
  - Hybrid local-cloud deployment optimization
- **Long-term**:
  - Scalable swarm cognition for multi-UAV collaboration
  - Safety assurance for LLM-driven decision-making
  - Regulatory compliance and UTM integration
- **Limitations**:
  - Current validation is simulation-based (sim-to-real gap)
  - Real-world factors (wind, GPS degradation) need testing

Thank You!

Questions?

Prof. Anis Koubaa
AlfaisalX Center of Excellence
Alfaisal University, Saudi Arabia
akoubaa@alfaisal.edu
aniskoubaa.org

SmartTech 2025
January 14–16, 2025 — Tunis, Tunisia