SAFE EXPERIMENTATION WITH LLM-CONTROLLED UAVS
AN AGILE SYSTEMS ENGINEERING APPROACH TO REQUIREMENTS DEVELOPMENT FOR AUTONOMOUS SYSTEMS

MATTHEW HARRIS, CO-FOUNDER & CTO
SAIF AUTONOMY
matt@SAIFautonomy.ai

AI4SE & SE4AI Workshop 2025, Sept. 17, 2025
Washington, DC

BACKGROUND

PROBLEM

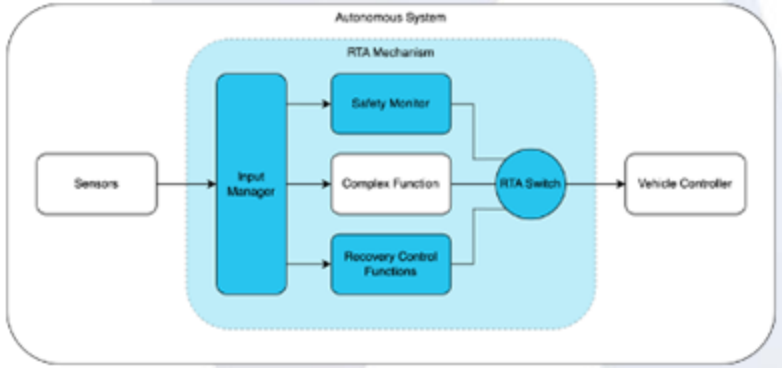
Need to experiment with advanced AI-based control systems

but, also

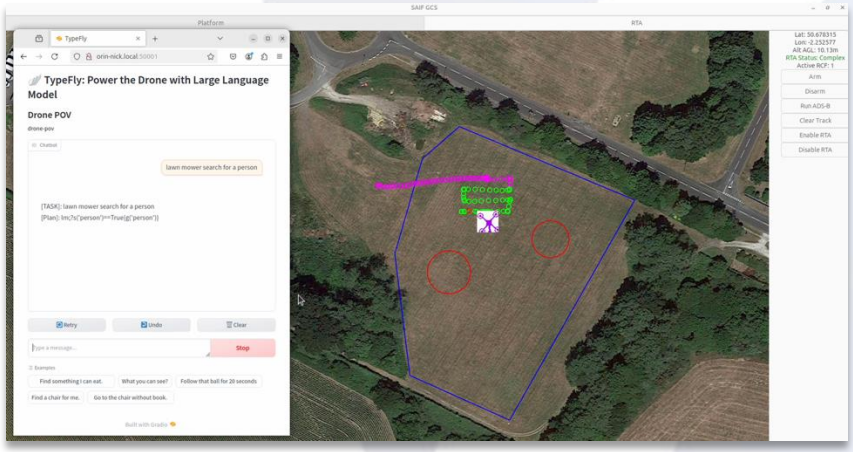
Need to ensure safety and security

APPROACH

- Employ a Runtime Assurance (RTA) architecture
- Independent of the autonomy or ‘AI’ controller
- Safety is assured by the RTA mechanism to enable experimentation with novel controllers



TEST UAV SYSTEM

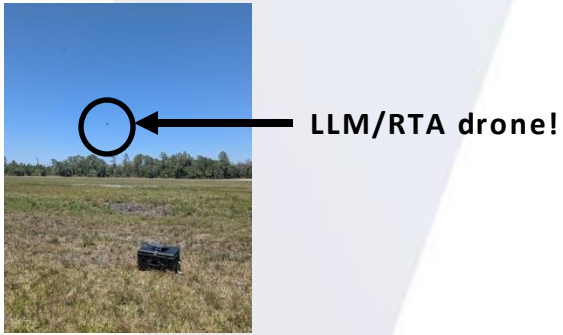


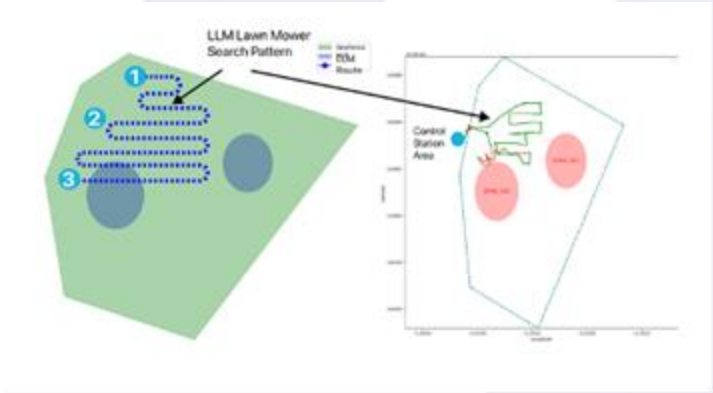
METHODOLOGY

- Goal: experiment with an LLM-based controller for a small UAV (with an RTA mechanism to ensure safety) to elicit requirements

RESULTS

Scenario	Results Summary
Scenario 1: Time Based ROZ	The SAIF RTA Module respected the time-based restrictions, whilst ensuring no violations for both polygonal and circular based ROZ and Geofences.
Scenario 2: ADS-B Infringement	The SAIF RTA Module predicted a violation with the non-cooperative air traffic, executing avoiding action to the east before successfully completing its assigned mission.
Scenario 3: LLM Controller Based Search	The RTA Module throughout the course of the LLM search ensures no violations with the defined restrictions.
Scenario 4: Air Corridor	The UAV platform successfully navigates through a confined air corridor, with the RTA Module making many micro-corrections in the platform's trajectory to ensure no violations.





ELICITED REQUIREMENTS: LLM CONTROLLER

ID	Title	LLM Controller Requirement
1	Mission Constraints Awareness	The LLM Controller shall incorporate known operational constraints (no-fly zones, altitude limits, corridors) in its planning logic to avoid obviously infeasible or unsafe actions.
2	RTA Feedback	The LLM Controller shall handle cases where its command is denied by the RTA.
3	Command Formalism and Bounded Output	The LLM Controller shall output commands in a formal language/format.
4	State Awareness and Timely Goal Execution	The LLM Controller shall be aware of mission progress.
5	Safety in Language Understanding	The LLM Controller's natural language understanding should be constrained to prevent dangerous misinterpretations.

ELICITED REQUIREMENTS: RTA SYSTEM

ID	Title	RTA Module Requirements
1	Recovery Action Effectiveness	Each Recovery Control Function used by the RTA shall be proven to bring the UAV to a safe state for the specified violation type.
2	Minimal Mission Interference	The RTA should aim to preserve mission objectives while assuring safety.
3	Switching Stability	The RTA system shall avoid frequent toggling that could destabilise control or become a nuisance to human supervisors/operators.
4	Performance and Latency	The RTA system decisions (monitoring + switching) shall occur within a bounded latency.
5	Transparency and Logging	The RTA system shall log all interventions and the reasons (which constraint triggered) and provide an interface for status monitoring (so an operator or a safety auditor can understand what the RTA is doing).

CONCLUSIONS

1. Demonstrated an RTA architecture in a novel UAV application with an LLM controller
2. Proposed and refined specific requirements for LLM controllers and RTA/safeguarding systems
3. Demonstrated an agile requirements engineering approach for AI-based systems
4. Bridges the gap between traditional certification approaches and ‘AI’ assurance

FUTURE WORK

1. Expanded scenario testing
2. Verification of RTA components
3. LLM controller improvement
4. Human-On the Loop interfaces
5. Applying to other domains
6. Certification pathways

matt@SAIFautonomy.ai