Unit Testing KPIs for Topic 3 — TIAGo Cooking Assistant

Introduction to Unit Testing KPIs

In order to ensure the reliability, robustness, and responsiveness of the developed cognitive architecture for assisting elderly individuals in meal preparation, a set of targeted Key Performance Indicators (KPIs) has been defined for unit testing.

These KPIs aim to systematically verify the correct behavior of each mandatory component, validate their interaction with the system environment, and assess their ability to handle real-world scenarios such as command conflicts, state transitions, and dynamic planning.

The KPIs have been designed to cover both **functional correctness** and **performance requirements**, guaranteeing that the TIAGo robot can support elderly users safely, efficiently, and intuitively throughout the cooking process.

Each KPI is linked to specific testing approaches and measurable success criteria, allowing for objective validation of the system's core functionalities.

This section defines the **Key Performance Indicators (KPIs)** for the **unit testing** of the mandatory components:

Recipe Tracking

Action Planning Based on Cooking State and Task History

Human Command Monitoring and Conflict Resolution

1. Recipe Tracking

KPI	Description	Success Criteria
Recipe Parsing Reliability	Evaluates robustness in parsing JSON recipes (valid/invalid inputs).	100% success rate for valid JSON; graceful rejection of invalid inputs without crashes.
Correct Step Recording	After executing a cooking step, the system correctly updates the execution history.	100% of completed n actions are logged correctly.
Step Transition Accuracy	Measures correct step transitions after IDLE period and stability during EXECUTING Urgent /FAILURE states.	100% correct step increments after 3s IDLE; no transitions during EXECUTING Urgent /FAILURE.
Step Order Integrity	The sequence of performed actions matches the predefined recipe sequence.	100% sequence consistency after N steps.
Missing Step Detection	The system detects if a step is skipped or missing in the sequence.	95%+ detection accuracy on test cases with induced missing steps.
Recipe Representation	Given a Representation, the system retrieves the correct latest performed step and	100% correct retrieval over all test cases.

KPI	Description	Success Criteria
	upcoming step.	

2. Action Planning

KPI	Description	Success Criteria
Action Sequence Mapping Accuracy	Verifies correct generation of action sequences from predefined step mappings.	100% match between generated sequences and predefined mappings.
Command Insertion Correctness	Tests if new commands are inserted into sequences according to urgency rules.	≥95% of insertions follow urgency policy (non-urgent ratio ≥50%).
Planning Responsiveness	Measures time taken to generate action sequences after receiving a step.	Total planning time <200ms from step reception to sequence generation.
Interruption Handling	Validates system behavior when new commands interrupt ongoing sequences.	100% correct sequence restructuring without errors or data loss.
State Synchronization	Ensures internal action index matches controller feedback.	0% desync between current_action_index and /action_index updates.
Service	Tests robustness when	Node retries gracefully

KPI	Description	Success Criteria
Availability Handling	required services are temporarily unavailable.	without crashes until service becomes available.
Next Action Correctness	The planner chooses the correct next action based on current state and history.	≥95% correct decisions across 20+ diverse scenarios.
Dynamic Adjustment Capability	System adjusts plans correctly when unexpected events occur.	Correct adjustment in ≥90% of disruption cases.
Plan Consistency	Verifies logical coherence of new plans after adjustments.	100% logical plan verification post-adjustment.
Response Time	Time taken to generate next action after state update.	≤500ms per decision in normal operation.
Next Action Correctness	The planner chooses the correct next action based on the current cooking state and history.	95%+ correct next-action decision across 20+ diverse scenarios.
Dynamic Adjustment	The system adjusts the plan correctly when	Correct adjustment in

KPI	Description	Success Criteria
Capability	unexpected events occur (e.g., ingredient missing).	90%+ of induced disruption cases.
Plan Consistency	After replanning, the new plan remains logically coherent with the original goal (completing the recipe).	100% logical plan verification post adjustment.
Response Time	Time taken to generate the next action after state update.	≤ 500 ms per decision in normal cases.

3. Human Command Manager

KPI	Description	Success Criteria
Command Recognition Accuracy	The system correctly identifies and understands verbal commands related to ingredients/tools.	≥ 90% recognition accuracy on a 50-command test set.
Conflict Detection Rate	Correctly identifies when a command conflicts with the current recipe execution plan.	≥ 95% conflict detection across all tested scenarios.
Correct Decision on Conflicts	The system appropriately accepts valid commands and rejects invalid ones.	≥ 95% correct decision rate.

KPI	Description	Success Criteria
Conflict Handling Response Time	Time taken to resolve a conflict (accept or reject a verbal command).	≤ 800 ms average per conflict scenario.
Non-Urgent Ratio Calculation	Accuracy in calculating urgency ratios for action sequences	100% accurate calculations across test cases
Response Latency	Time taken to process and respond to verbal commands	<100ms from command receipt to publishing decision
Sequence Update Handling	System's robustness during concurrent command and sequence updates	0% corruption or data loss during concurrent updates
Command Recognition Accuracy	Ability to correctly identify spoken ingredients/tools commands	≥90% accuracy on 50-command test set
Conflict Detection Rate	System's ability to identify command-recipe plan conflicts	≥95% detection across all test scenarios
Conflict Handling Response Time	Speed of conflict resolution process	≤800ms average resolution time
Correct Decision on Conflicts	Appropriateness of accept/reject decisions for conflicting commands	≥95% correct decision rate

KPI	Description	Success Criteria

Conclusion on Unit Testing KPIs

The defined unit testing KPIs provide a structured methodology to assess the performance and correctness of the cognitive architecture at a granular level.

Through systematic testing aligned with these KPIs, we ensure that each component not only meets its functional specifications but also behaves reliably under varying operational conditions.

Successful validation against these KPIs establishes a solid foundation for subsequent integration testing and real-world deployment of the TIAGo Cooking Assistant system.

---

4. Cross-Component KPIs (Integration Level)

KPI	Metric	Testing Approach	Success Criteria
End-to-End Decision Latency	< 500 ms from verbal command to action	Full system testing	Response chain within latency constraint

KPI	Metric	Testing Approach	Success Criteria
	sequence update		
Error Recovery	100% recovery from malformed commands	Inject invalid or conflicting commands	System maintains safe operation and recovers
Concurrency Performance	< 5% performance drop under 10x load	Stress test with rapid-fire state changes and commands	System maintains high responsiveness
Memory Leak Resistance	< 1% memory increase after 1000 command cycles	Long-term stress testing	Memory remains stable without leaks

Conclusion on Unit Testing KPIs

The defined unit testing KPIs provide a structured methodology to assess the performance and correctness of the cognitive architecture at a granular level.

Through systematic testing aligned with these KPIs, we ensure that each component not only meets its functional specifications but also behaves reliably under varying operational conditions.

Successful validation against these KPIs establishes a solid foundation for subsequent integration testing and real-world deployment of the TIAGo Cooking Assistant system.