ENHANCING 6G NETWORK RESILIENCE THROUGH HORSE'S LLM AGENT-BASED SECURITY MECHANISMS

Using LLM agents, to automate, refine and execute threat mitigation actions in 6G networks.

Authors: M. Danousis, A. Piemonti, F. Granelli, X. Masip-Bruin, E. Rodriguez, A. Carrega, C. Skianis, E. Kafetzakis, I. Giannoulakis

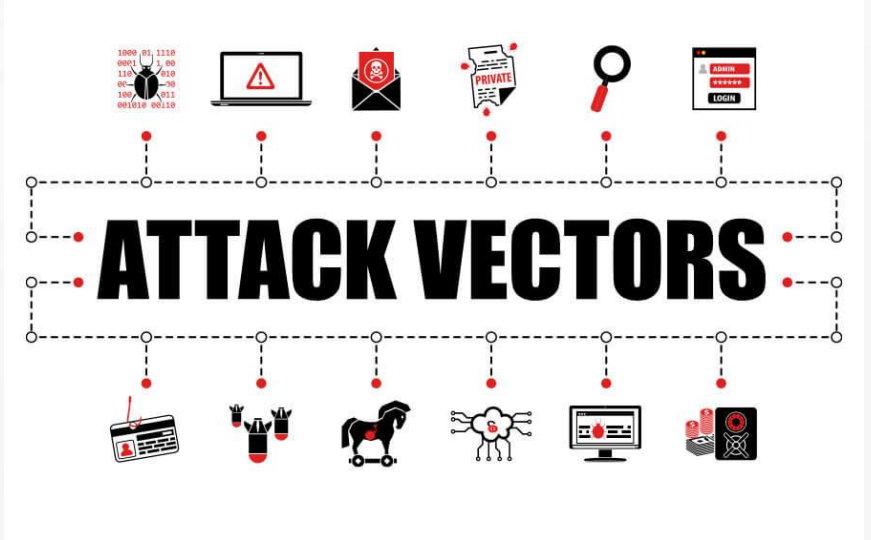
IEEE Globecom 2025

PRESENTATION CONTENT

- The 6G Security Imperative: Complexity & Vulnerability
- Introducing HORSE: A Holistic 6G Security Framework
- HORSE Architecture: AI-Driven Orchestration
- The Common Knowledge Base (CKB) Workflow
- Contribution : LLM Agent-Based Mitigation Workflow
- The Iterative Feedback Loop for Refinement
- Experiment Setup and LLMs Evaluated
- Key Result: Execution Success Rate, Efficiency & Adaptability, Correct Intention Translation
- Summary of Contributions and Future Work
- Q&A

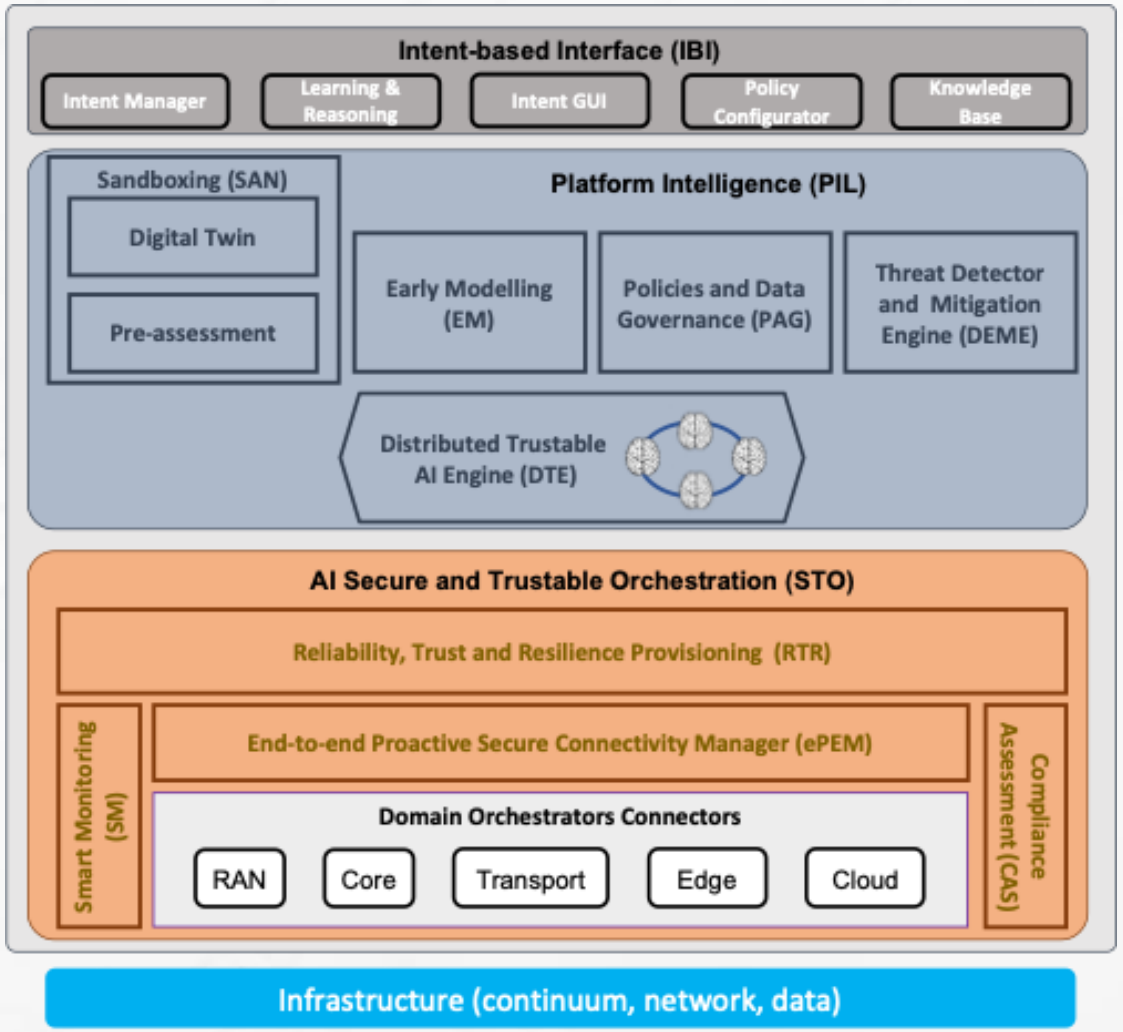
THE 6G SECURITY IMPERATIVE: COMPLEXITY & VULNERABILITY

- **6G Networks:** Expected to be highly complex, scalable, and heterogeneous, introducing new attack vectors and vulnerabilities.
- **Traditional Security Gap:** Signature-based and rule-based systems struggle to keep pace with rapidly evolving threats.
- **The Need for Automation:** Securing these systems requires robust, end-to-end, and *proactive* security frameworks, like HORSE.



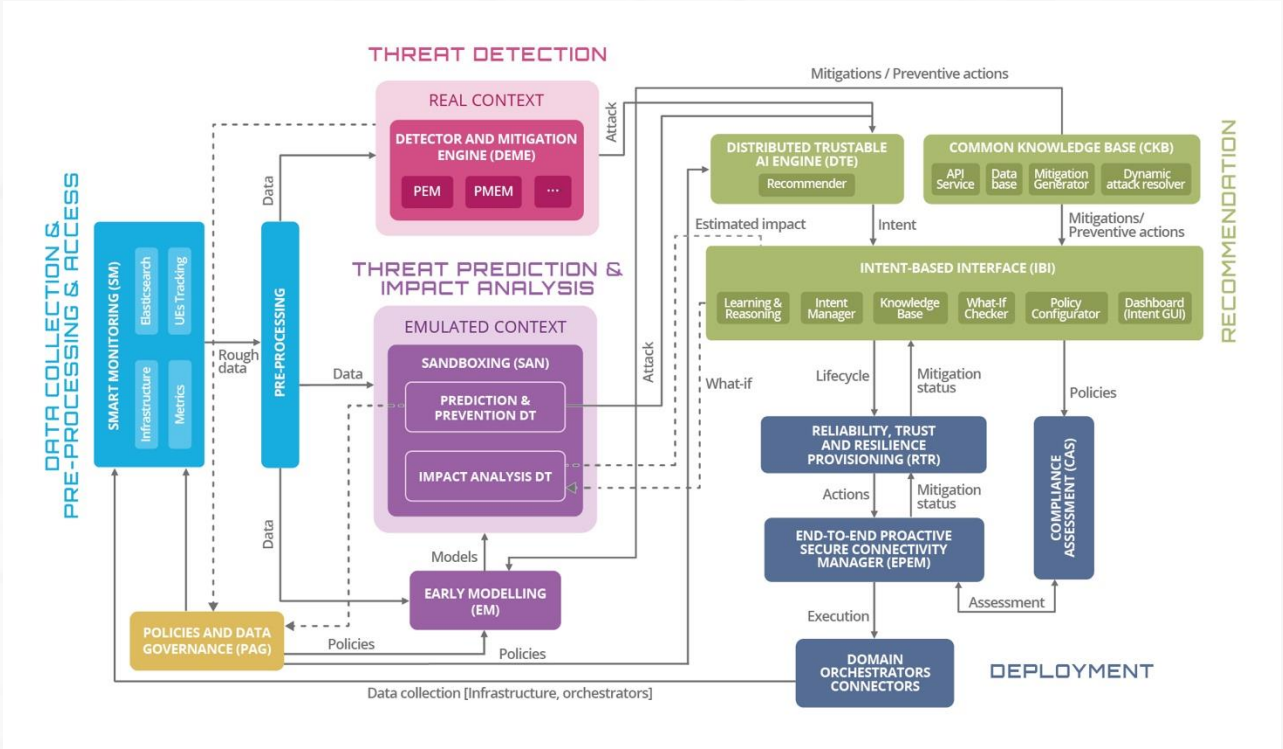
INTRODUCING HORSE: A HOLISTIC 6G SECURITY FRAMEWORK

- **HORSE Overview:** A pioneering architecture for end-to-end security service management in future 6G networks.
- **Core Pillars:** Integrates AI/ML natively for advanced functions like threat detection, prediction, and secure coding.
- **This Paper's Focus:** We introduce an LLM-agent-based subsystem within HORSE that automates the translation and enforcement of security mitigation actions.

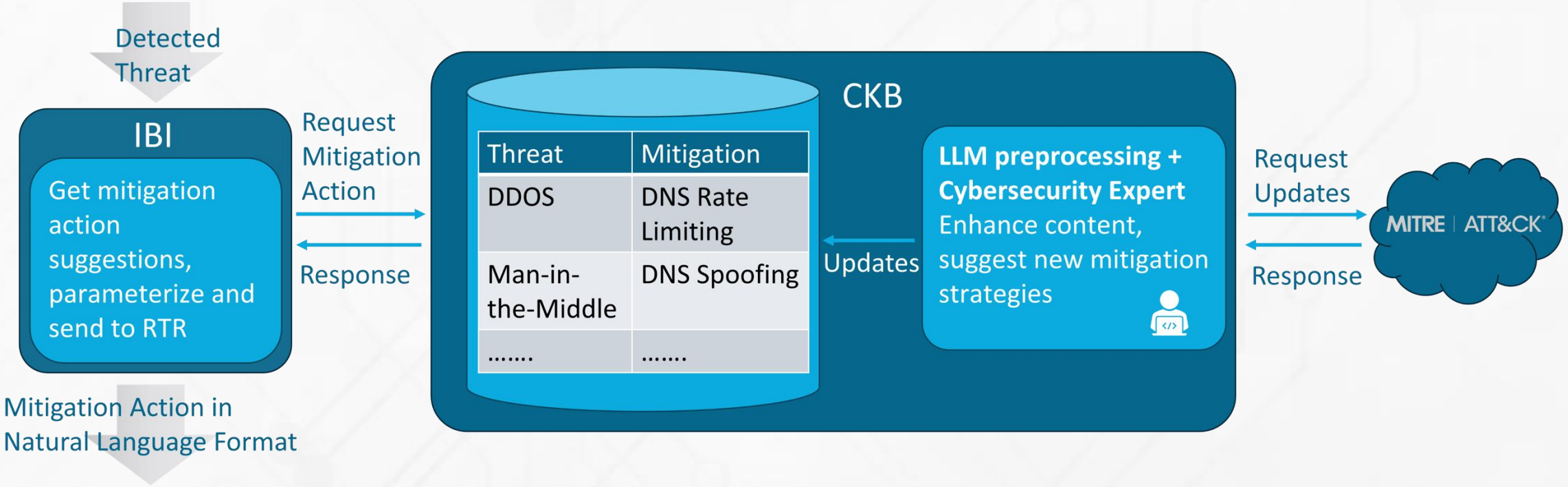


HORSE ARCHITECTURE: AI-DRIVEN ORCHESTRATION

- **Platform's Key Components:**
  - Data Collecting & Pre-Processing
  - Threat Detection
  - Threat Prediction & Impact Analysis
  - Recommendation
  - Deployment



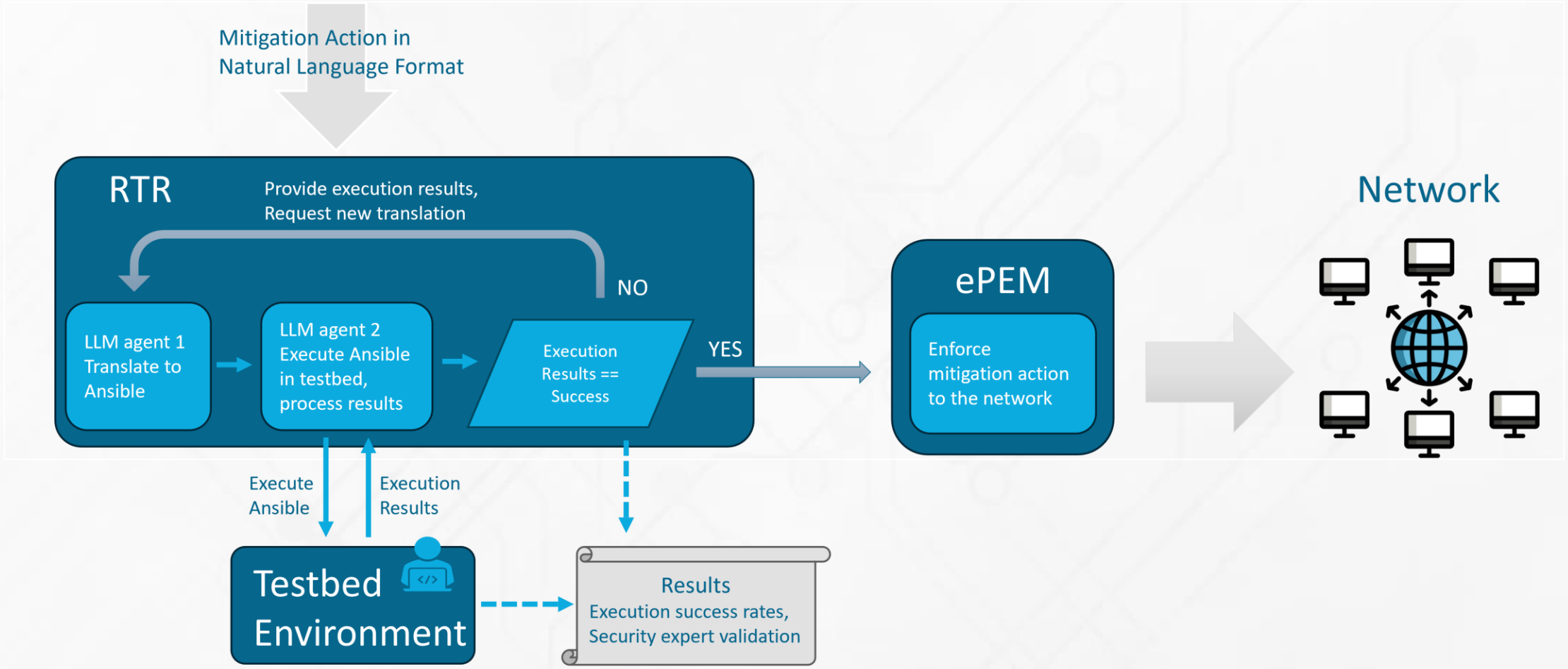
THE COMMON KNOWLEDGE BASE (CKB) WORKFLOW (MITIGATION ACTION GENERATION)



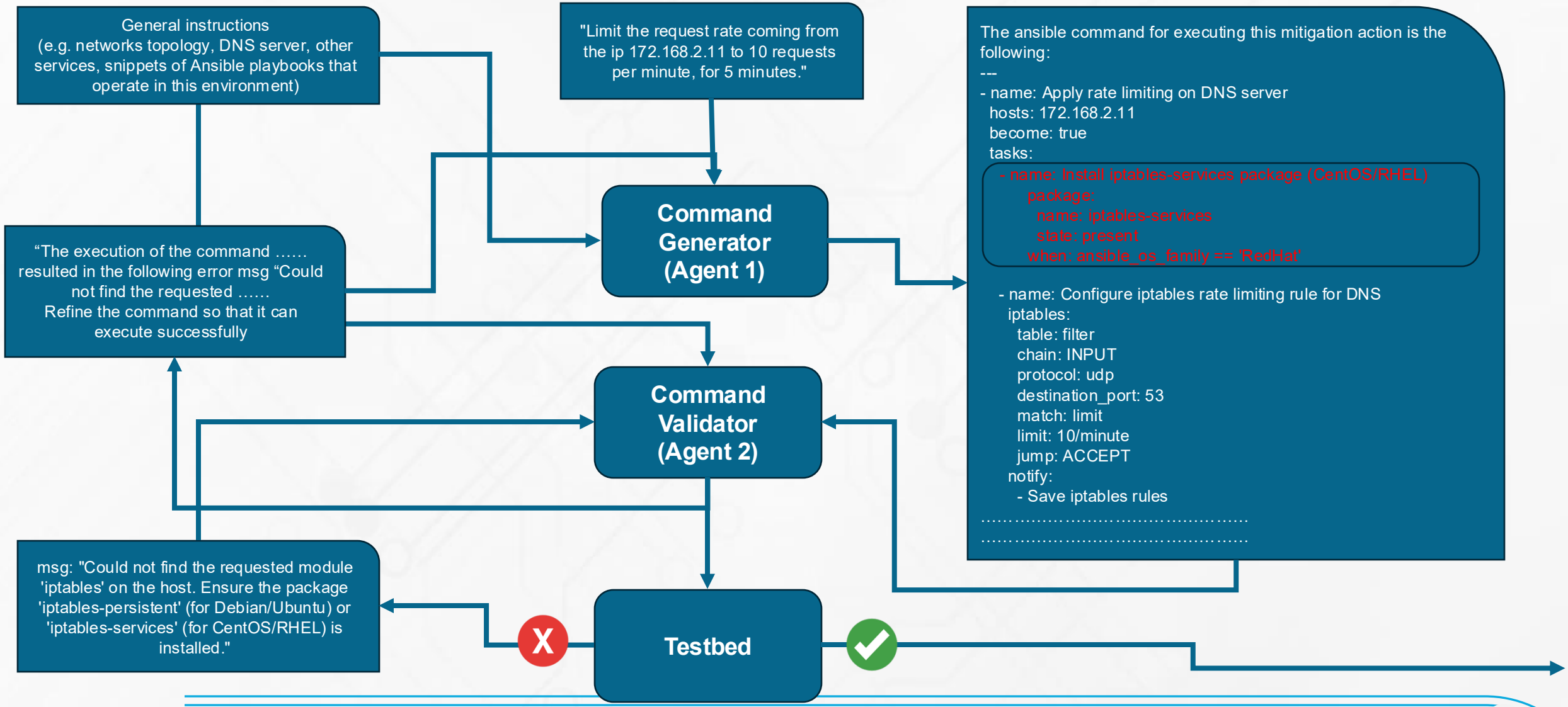
LLM AGENT-BASED MITIGATION WORKFLOW

- **Problem:** Mitigation actions are often high-level (natural language); they need to be translated into executable commands (e.g., Ansible).
- **The Solution:** A **dual-agent system** (Command Generator & Command Executor) operating within the Reliability, Trust, and Resilience Provisioning (RTR) module.
- **Goal:** Automate the translation of natural language mitigation intent into executable Ansible commands.

CONTRIBUTION: THE ITERATIVE FEEDBACK LOOP FOR REFINEMENT

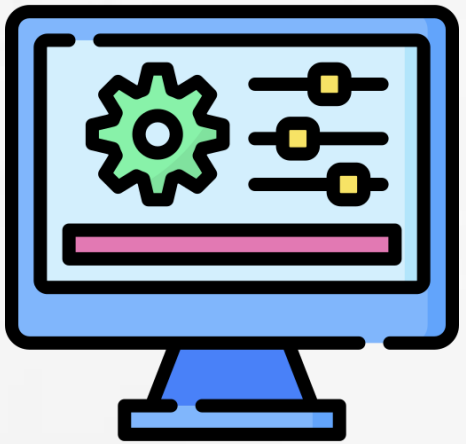


CONTRIBUTION: THE ITERATIVE FEEDBACK LOOP FOR REFINEMENT



EXPERIMENT SETUP AND LLMS EVALUATED

- **Dataset: 230 mitigation sentences** created collaboratively by human experts, MITRE, and LLMs (to ensure diversity and ambiguity).
- **Workflow:** Used the LangChain framework to implement the two-agent system.
- **Models Compared (Open-Source):** Llama 2 (70B), Llama (7B), and Falcon (40B).
- **Evaluation Metrics:**
  - 1) Executability (Success Rate)
  - 2) Efficiency (Avg. Iterations)
  - 3) Latency (Execution Time)
  - 4) Accuracy (Intention Translation)



KEY RESULT: EXECUTION SUCCESS RATE & AVG. ITERATIONS

LLM Model	Success (%)	Failure (%)	Avg. Iterations
Llama 2 (70B)	73%	27%	1.7
Llama (7B)	56%	44%	1.2
Falcon (40B)	65%	35%	1.5

TABLE I
SUCCESS RATE AND AVERAGE ITERATIONS FOR EACH LLM

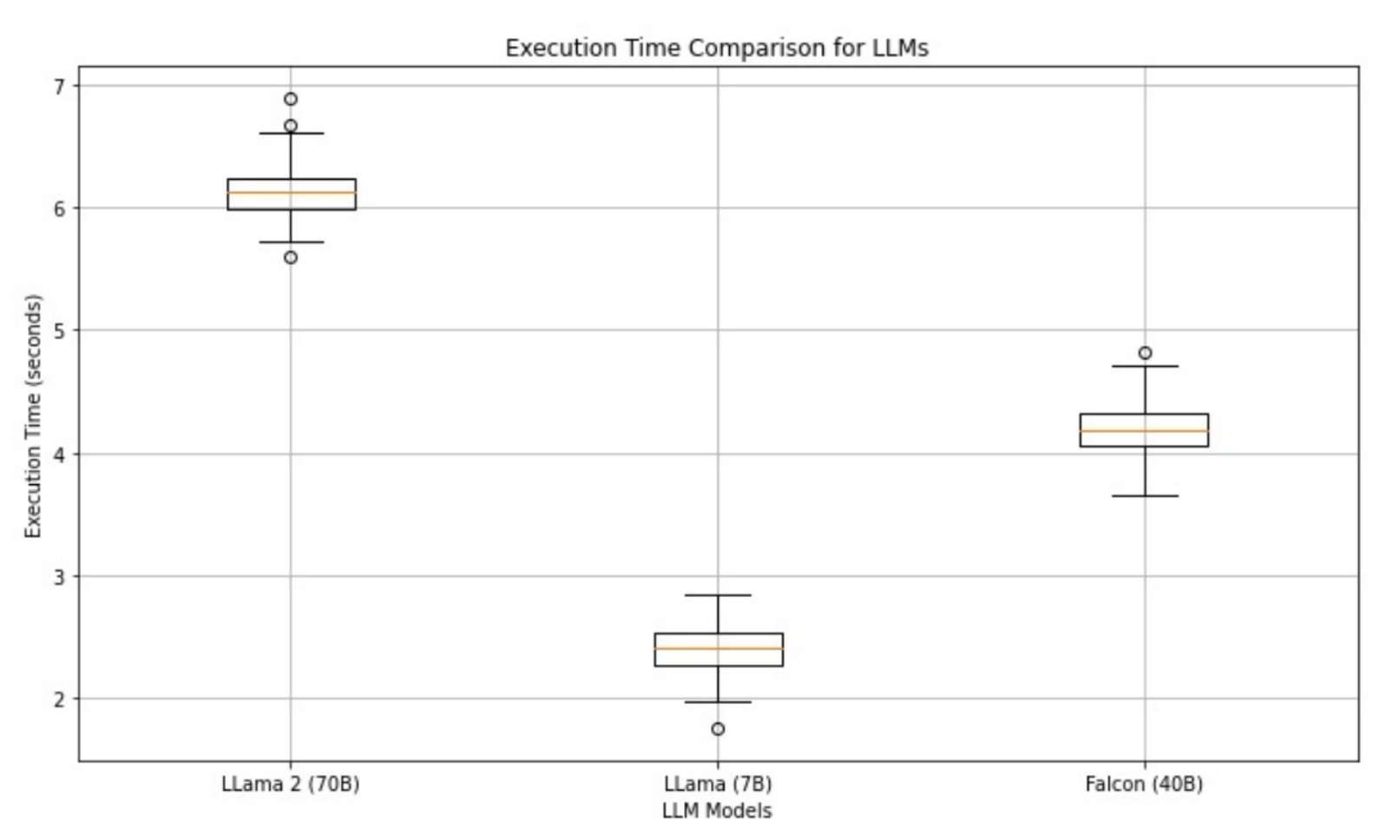
KEY RESULT: CORRECT INTENTION TRANSLATION (ACCURACY)

Method: Cybersecurity experts (3) graded successful commands on a 1-to-5 scale (5 = perfect translation).

The following average scores were observed:

- Llama 2 (70B): 4.5/5
- Llama (7B): 3.8/5
- Falcon (40B): 4.2/5

KEY RESULT: CORRECT INTENTION TRANSLATION (ACCURACY)



SUMMARY OF CONTRIBUTIONS AND PERFORMANCE

- **The Power of LLM Agents:** Demonstrated that LLMs can effectively automate the complex translation of high-level mitigation intents into executable Ansible commands.
- **The Feedback Loop is Key:** The dual-agent workflow with self-correction significantly streamlines operations and reduces human supervision.
- **Top Models:**
  - Llama 2 (70B) most reliable, success rates (up to 73%), with high intention accuracy (4.5/5), execution time (6.15 sec).
  - Falcon (40B) 2nd, but faster than Llama2 (4.2 sec)

FUTURE WORK: SCALING AND ENHANCEMENT

- **Test in Complex Environments:** Validate the workflow's scalability and effectiveness in more complex, real-world 6G network environments.
- **Model Refinement:** Fine-tune LLMs to further improve both the success rate and semantic accuracy of generated commands.
- **Expanding Scope:** Broaden the range of supported mitigation actions and explore incorporating a larger number of LLM agents for richer collaboration.

THANK YOU FOR YOUR ATTENTION

horse-6g.eu

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the other granting authorities. Neither the European Union nor the granting authority can be held responsible for them."