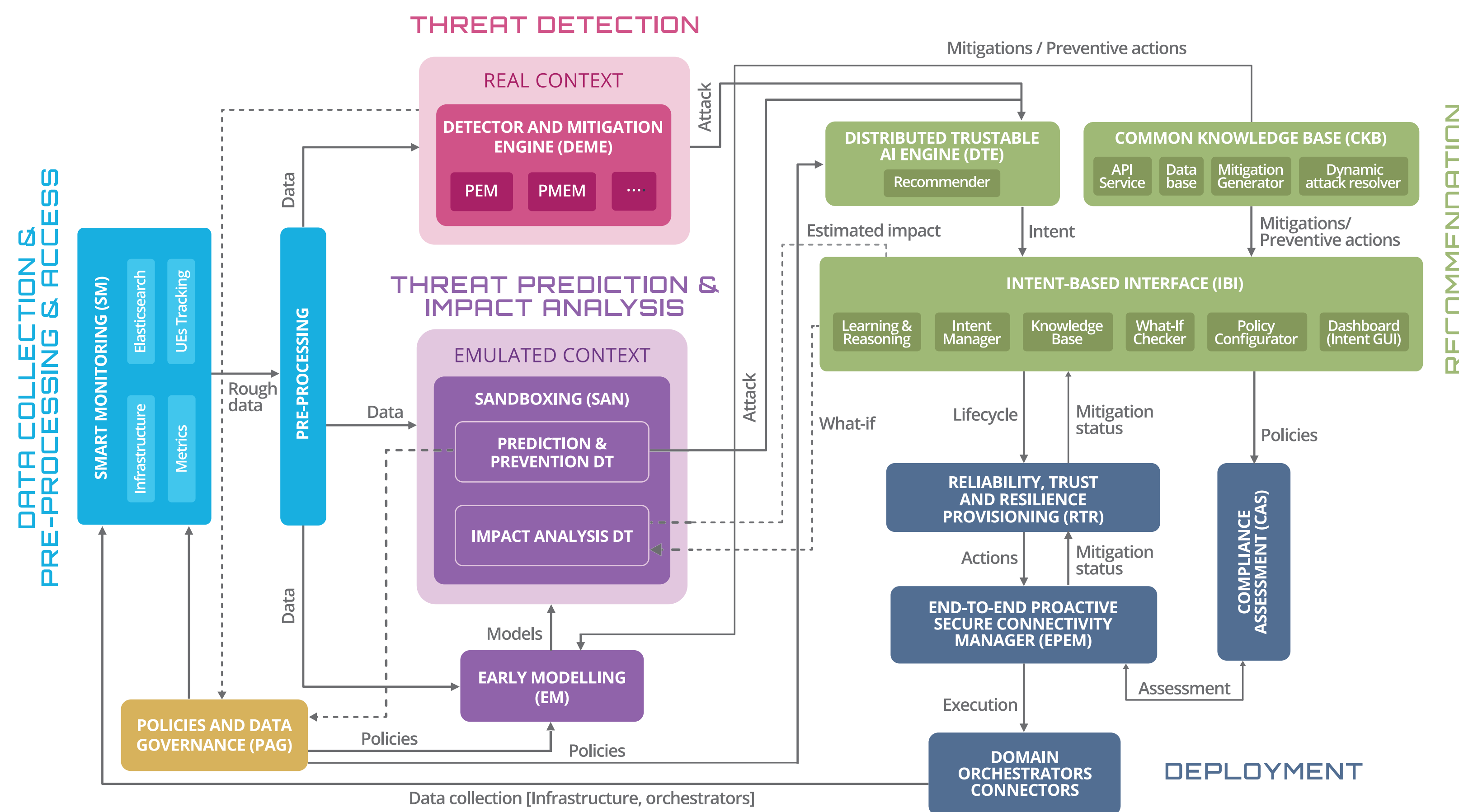


Motivation & Problem

6G networks introduce extreme complexity and dynamism, making traditional security techniques obsolete. Manual mitigation is slow and error-prone. **We propose automated mitigation actions using LLM agents as part of the HORSE cybersecurity framework.**



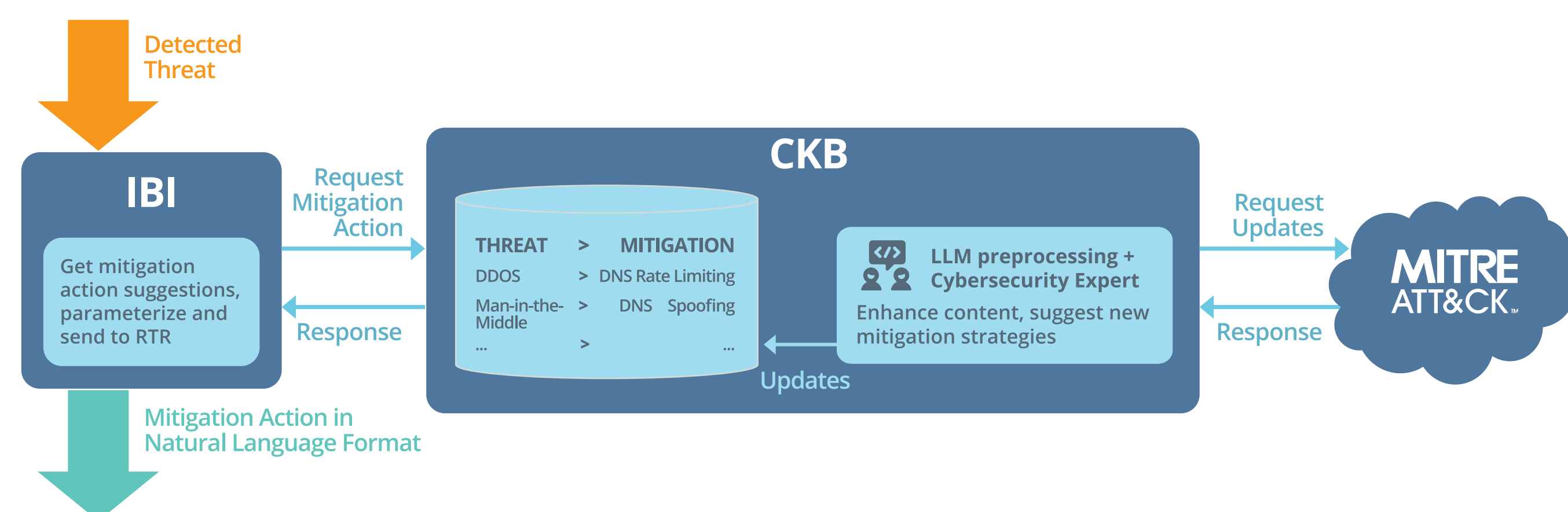
HORSE Architecture

- AI-powered 6G security framework.
- A knowledge base of cyberattacks and mitigations enhanced using generative AI.
- LLM-powered automation of mitigation actions.
- Digital Twin integration for threat prediction and impact analysis before enforcement.
- Intent-Based Interface (IBI) enabling human-centric, policy-compliant security orchestration.
- Trust and compliance layer aligned with regulatory frameworks, including learning from human operator decisions.

Data Collection & Knowledge Base Population

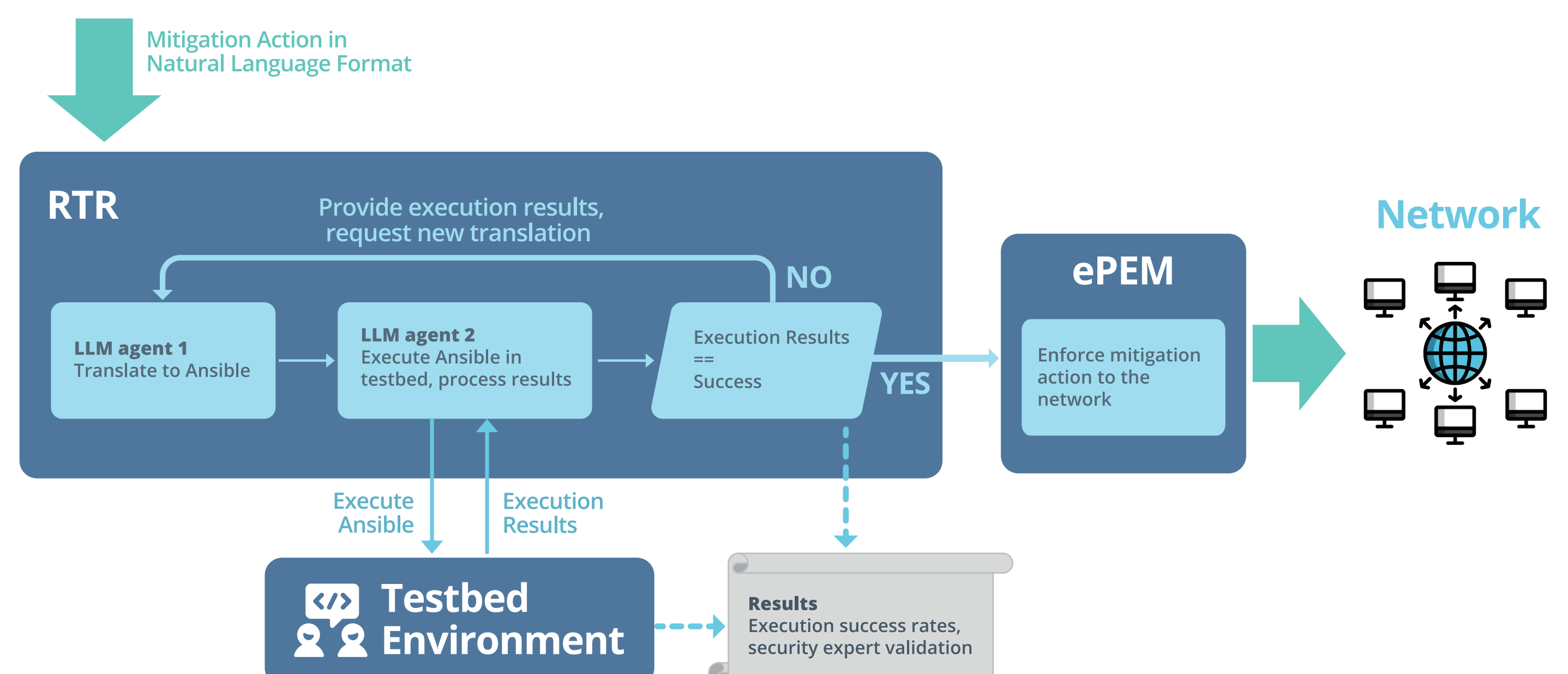
Generative AI to aggregate threat-mitigations pairs from trusted sources and enhance the CKB:

- The primary source of data is the MITRE ATT&CK framework.
- LLM-as-a-Judge to evaluate generated data at large scale.
- A human-in-the-loop strategy enables cybersecurity experts to review generated attack-mitigations pairs during the process.



Mitigation Action Enforcement Workflow

- **Translator LLM** converts natural-language mitigation instructions into Ansible commands.
- **Executor LLM** tests and refines the commands using feedback loops.
- **Repeat up to 3 iterations** > final command deployed via ePEM module.



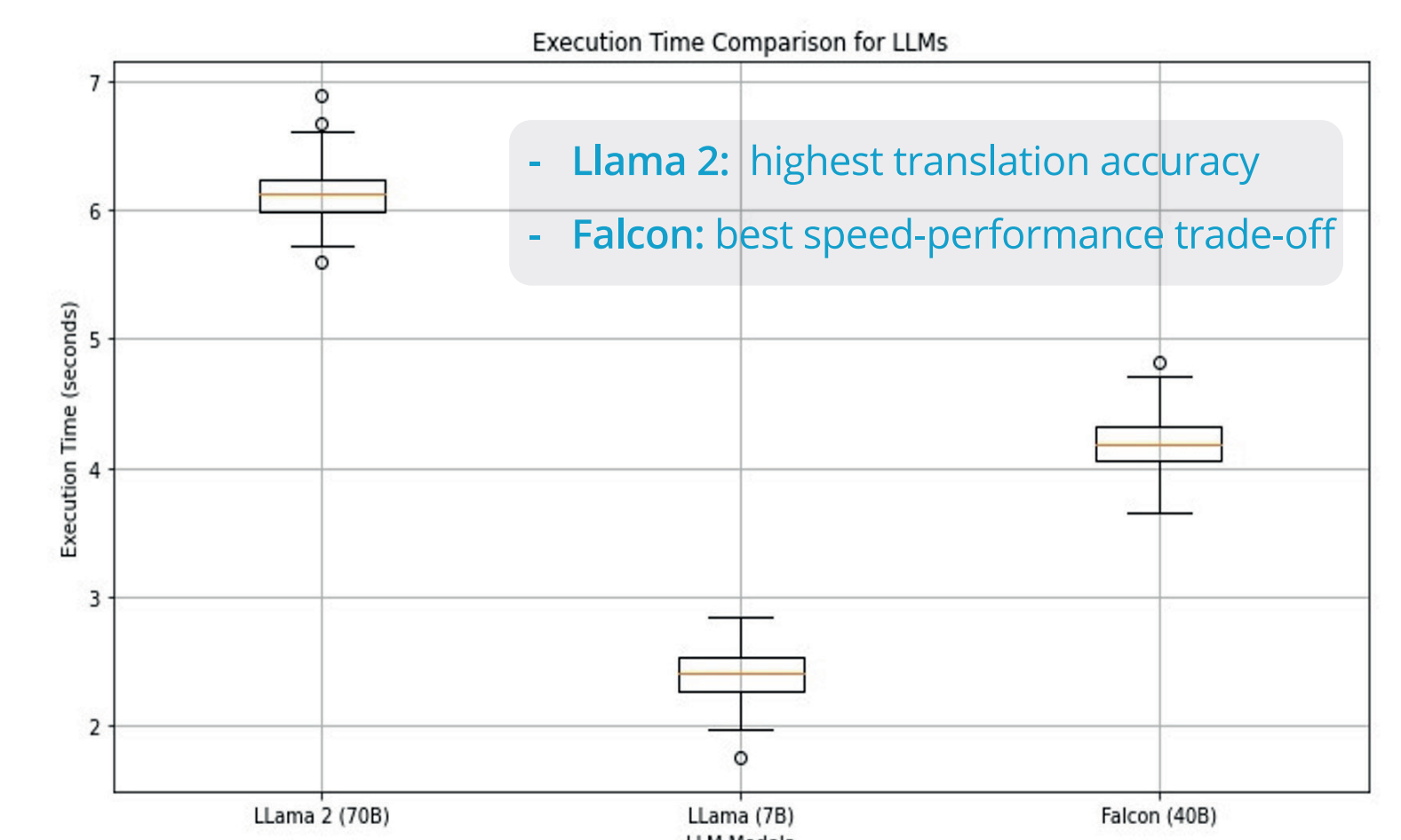
Experimental Setup & Results

- Dataset of 230 mitigation enforcement actions described in natural language. Mitigation actions extracted from MITRE ATT&CK, refined by cybersecurity experts, augmented using LLMs.
- Evaluated **3 LLMs** (Llama 2 70B, Falcon 40B, Llama 7B).

Metrics:

- Execution success rate (%)
- Iterations to success (1, 2 or 3)
- Intention translation accuracy (1 to 5)
- Execution time
(Inference Hardware: Nvidia RTX 3090 GPU)

Model	Success %	Avg. Iterations	Intention Accuracy
Llama 2 70B	73%	1.7	4.5 / 5
Falcon 40B	65%	1.5	4.2 / 5
Llama 7B	56%	1.2	3.8 / 5



Conclusion & Future Work

- LLMs can reliably translate natural language into executable actions
- Enables autonomous and adaptive 6G threat mitigation
- Future Work: Fully Automated Data Collection and Knowledge Base Population using Generative AI workflows

 **Contact person**
mkafetz@8bellsresearch.com



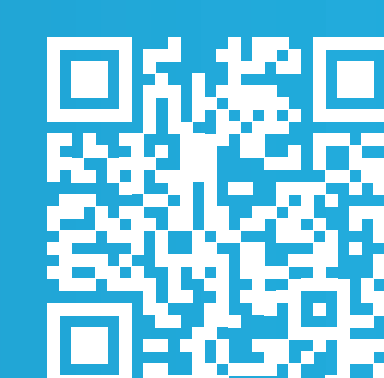
horse-6g.eu



X



in



Co-funded by
the European Union



6GSNS