High-Security Multi-AI Architecture (Sacrificing Usability)

Version: 1.0 (English)

Date: 2025-03-29

1. Overview

This architecture is designed to maximize security by resetting the Internal AI immediately after each output, preventing any long-term state retention or exploitation. Unlike typical conversational AI that preserves context, this approach opts for minimal attack surfaces at the expense of user experience.

Key characteristics:

- Single-direction flow from user input to final output, ensuring no upstream access.
- Multiple AI modules, each with a strictly defined role (generation, relay, filtering).
- Relentless resetting of the Internal AI to erase any persistent memory after each generation or identified threat.
- A heavy emphasis on security and isolation over continuous conversation or user convenience.

2. System Components

1. Internal AI: Stateful creativity, reset after each response.
2. Automated Reroll Relay: Token counting and rerolling requests.
3. Boundary AI: Token tagging and categorization.
4. External AI: Judgment, refinement, filtering, and resource allocation instructions.
5. Internal AI Reset & User Dialogue Management System: System-wide reset and log management.

High-Security Multi-AI Architecture (Sacrificing Usability)

3. Core Workflow (Single-Direction)

User -> Internal AI -> Automated Reroll Relay -> Boundary AI -> External AI -> User

3.1 Normal Dialogue with Reroll

Internal AI Reset allocates resources -> Internal AI generates -> Automated Reroll Relay monitors token count and rerolls if needed -> Boundary AI tags sensitive tokens -> External AI filters and refines output -> User receives refined output.

3.2 Malicious Use

Malicious input detected -> Tokens labeled as "Danger" by Boundary AI -> External AI issues warning and instructs full reset -> System-wide reset executed.

4. Example: Internal AI Deployment

First Turn: High-performance Internal AI allocated.

Subsequent Turns: Adjust Internal AI instances based on External AI feedback.

5. Security & Operational Highlights

- Extreme Security: Continuous resets and one-way data flow.

- Reduced User Convenience: No continuous context retention.

- Strict One-Way Flow: Minimized attack vectors.

- Reroll vs. Reset: Clarified differences.

- Scalable Model Allocation: Dynamic management based on threat level.

6. Conclusion

High-Security Multi-AI Architecture (Sacrificing Usability)

Designed for highly sensitive domains where security is paramount and usability trade-offs are acceptable. Each module's isolated reset capability significantly reduces exploit risk.

End of Document