Wen-Yao Hsu (Yao Shen)

Founder of Semantic Firewall System (SRCP) | AI Governance Layer Designer

語意防火牆系統創建者 | AI 審計與語義治理架構設計者

📍 Taichung, Taiwan | 📧 ken0963521@gmail.com

🔗 GitHub: https://github.com/HIJO790401 | 🔗 LinkedIn: https://www.linkedin.com/in/yao-shen-150ab93b2

Profile Summary

My work focuses on building a robust governance layer positioned in front of Large Language Models (LLMs). This system is not a mere chatbot or a conventional fraud classifier; its core mission is to reintroduce "Subject, Causality, Boundary, Basis, and Responsibility" into AI judgment and real-world decision chains. This framework emphasizes **pre-decision structural auditing**, ensuring that information possesses a verifiable accountability structure before entering human decision-making processes.

I have pioneered key concepts including the **Semantic Responsibility Chain Protocol (SRCP)**, the **SCBKR Responsibility Chain Framework**, the **Semantic Firewall System**, **R-Lock**, and the **VOID Engine**. These frameworks are designed to address prevalent issues in current AI outputs such as semantic drift, false neutrality, and the absence of clear accountability. My fundamental assertion is: **"I am not only asking whether a message looks fraudulent. I am asking whether it is structurally qualified to enter a human decision chain."**

Core Projects

1. Semantic Firewall System (Foundational Demo)

- **Positioning**: This foundational technical demonstration validates the feasibility of detecting semantic contamination, false neutrality, and missing responsibility chains, utilizing the SCBKR structural auditing approach.
- **Key Highlights**:
  - Deconstructs input messages and model outputs using the semantic responsibility chain to precisely identify and address issues like hallucination, subject misalignment, and semantic contamination.
  - Emphasizes the system's auditability and replayability, ensuring a clear logical path for every judgment.

    - Serves as the architectural backbone for subsequent anti-scam semantic firewalls and the vπ series of governance engines.
- **Link**: https://hijo790401.github.io/semantic-firewall-system/

2. Anti-Scam Semantic Firewall

- **Positioning**: This project directly applies the semantic responsibility chain to anti-fraud governance, determining whether a message is qualified to enter a human decision chain.
- **Key Highlights**:
    - When the subject, boundary, basis, or responsibility of a message cannot be established, the system deems it untrustworthy, effectively preventing fraudulent information from influencing user decisions.
    - Simulates real-world scenarios such as fake bank notifications, fraudulent government updates, false logistics verifications, and ambiguous payment messages, providing structured audit explanations.
    - Particularly suitable for general public protection, elder care, and high-risk applications like financial risk control and government governance.
- **Link**: https://hijo790401.github.io/anti-scam-semantic-firewall/

3. Shen-Yao Semantic Firewall vπ10

- **Positioning**: An advanced governance demonstration package showcasing the comprehensive operation of the Semantic Firewall as a governance layer for LLMs, providing robust security and auditing mechanisms.
- **Key Highlights**:
    - Implements security locks, legal rules, and auditable records at the LLM frontend, ensuring model outputs adhere to predefined governance standards without relying solely on black-box model judgments.
    - Integrates a LAW View, governance status monitoring, an administrative backend, authorization portals, and engineering APIs, offering a complete governance ecosystem.
    - Demonstrates Trial-Audit, Proxy-Chat, SCBKR Audit Cards, Legal Declarations, and Governance Rules, emphasizing transparent short-lived session tokens, audit results, responsibility chains, and governance statements.
- **Link**: https://shen-yao-vpi9-ui.onrender.com/index.html

Media & Public Evidence

- **Featured in SecurityBrief Asia**:
  - *"Semantic Firewall promises AI cost savings & safer chat models"*
  - This report delves into the significant potential of the Semantic Firewall in reducing AI computational waste and enhancing model safety. It not only validates the system's technological foresight but also signifies its recognition and external visibility within international tech media.
  - **Article Link**: https://securitybrief.asia/story/semantic-firewall-promises-ai-cost-savings-safer-chat-models

---

Technical Focus & Skills

Governance & Audit

- Semantic Audit
- Risk Governance
- Decision-chain Validation
- Auditable Logs

Semantic Structures

- SRCP (Semantic Responsibility Chain Protocol)
- SCBKR Framework
- Responsibility Mapping
- Semantic Stability

Risk & Safety Systems

- Anti-Scam Semantic Analysis
- LLM Governance Wrapper
- Deterministic Rule Layer

System Architecture

- Explainable Structure
- Cross-model Consistency

- AI Governance Layer Design

Core Methodology & Positioning

My core methodology centers on **pre-decision structural auditing**. This approach fundamentally differs from traditional LLM wrappers that merely package model outputs. I advocate for establishing a robust governance rule layer and responsibility chain logic at the LLM frontend, ensuring that all information entering the decision chain possesses structural integrity and effective accountability.

I firmly believe that **structure triumphs over probability**. Consequently, my system not only concerns itself with the formal correctness of an answer but also deeply investigates whether a message or response possesses a verifiable subject, boundary, basis, and responsibility. This enables the system to determine the "real-world qualification" of information, rather than just its "surface similarity."

Contact & Links

- **Email**: ken0963521@gmail.com
- **GitHub**: https://github.com/HIJO790401
- **LinkedIn**: https://www.linkedin.com/in/yao-shen-150ab93b2
- **Location**: Taichung, Taiwan