School of Information Technology Experience+

Collaborate | Innovate | Apply

Toward Building Behavioral Testbeds for Security and Privacy: LLM-Driven Personas as Crash Dummies

Amir Reza Asadi, Joel Kwesi Appiah, Taiwo Peter Akinremi , Hazem Said

University of CINCINNATI

Motivation

- Technology development outpaces regulatory frameworks and security measures[1]
- LLM agents can simulate diverse user behaviors through personas[2]
- AI personas serve as digital crash dummies for security testing
- Scenario-Based Design[3] provides structured input and output for LLM Data Generation

Example Persona

Adapted from [4]

Anticipated Outcomes

Practical Contribution

- Insights for Building a Testbed for Proactive Threats & Vulnerability Identification
- Optimization Framework for Persona Simulation of Security and Privacy Scenarios

Theoretical Contribution

- Extend the Scenario-based Design Methodology to Develop LLM-Based Usability Studies

Research Questions

1. **To what extent do LLMs adhere to persona characteristics in roleplaying security and privacy scenarios?**
2. **To what extent do different LLMs exhibit diverse behavioral patterns when simulating various personas?**
3. **To what degree does simulating personas in a scenario reveal new vulnerabilities?**
4. **To what extent do LLM persona simulations match human analysts in vulnerability detection?**

Experiment Overview

Regular Users | **Threat Actors** | **Security Practitioners**

→ **Role Playing with LLMs**

→ **Test Scenarios**

Each scenario will be tested several times per LLM-persona combination

→ **Hybrid Evaluation**

Persona Fidelity	*Behavioral Diversity*	*Vulnerability Detection*
• Human validation • LLM self-evaluation	• Behavioral variation metrics • Semantic similarity analysis	• Known vulnerability comparison • Diversity of Vulnerability

→ ***Analysis Report for Testbed***

Rating of Vulenrabilities, Summary of Threats

Method

Systematic Literature Review:

- 128 Personas Identified in ACM, IEEE, Scopus

Example Scenario: Corporate Email Management

Persona Fidelity Experiment

- Intended persona characteristics vector vs. Simulated characteristics vector
- LLM grades and generate these vectors by grading, and the cosine similarity of scores

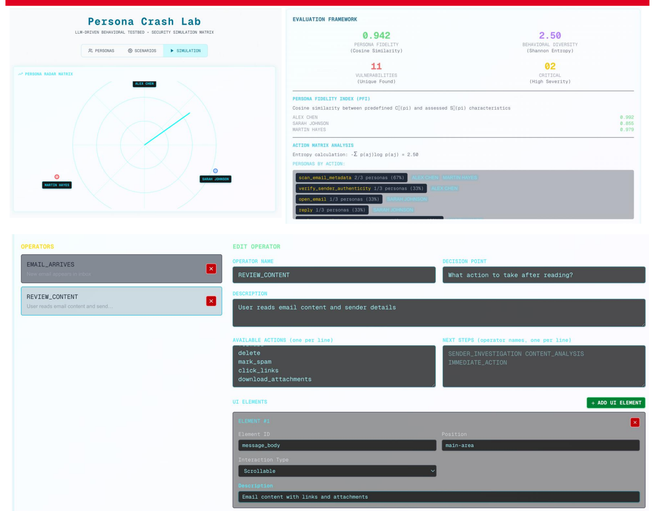
Behavioral Diversity Experiment

- Entropy calculation of persona-action matrix
- Cosine similarity of vector embeddings of actions

Vulnerability Detection Experiment

- Automated vulnerability extraction from scenario transcripts
- Unique vulnerabilities found / scenarios tested

Proof of Concept



References

1. Al-Jabouri, Mohammed. "The Evolution of Privacy Laws in the Digital Age: Balancing Innovation and Personal Security." *Utu Journal of Legal Studies (UJLS)* 1.1 (2024): 39-45.
2. Li, K., Dai, C., Zhou, W., & Hu, S. (2024). "Fine-Grained Behavior Simulation with Role-Playing Large Language Model on Social Media." *arXiv preprint arXiv:2412.03148*.
3. Carroll, J. M. (2003). *Making Use: Scenario-Based Design of Human-Computer Interactions*. MIT Press.
4. Tariq, Muhammad Adnan, Joel Brynielsson, and Henrik Artman. "Framing the Attacker in Organized Cybercrime." In *2012 European Intelligence and Security Informatics Conference*, pages 30–37. IEEE, 2012.