

Practicing Social Engineering in an AI Simulation Improves Student Preparedness

Problem Statement and Goals

Social engineering remains a significant cybersecurity threat, and advances in artificial intelligence have made these attacks more effective and scalable. Modern scammers can use AI to personalize communications, increasing the likelihood of deception. Many individuals lack an understanding of how these attacks operate and how vulnerable they may be.

To address this challenge, we developed an AI-driven training simulation that guides learners through the process of conducting a social engineering campaign. Users interacted with AI personas to complete goals. The goal of the simulation is to help users better recognize, understand, and defend against real-world social engineering attacks by experiencing them from an attacker's perspective.

Approach

Our approach involved designing a simulation from the perspective of the attacker called BAIT (Behavioral Attack & Influence Training). Students complete a structured campaign in which they socially engineer multiple individuals with differing personalities and vulnerabilities in order to reach a defined end goal. Each campaign is designed to be completed in approximately one hour. This attacker-perspective design allows students to understand how trust, pretexting, and information leverage are built across multiple interactions.



Results

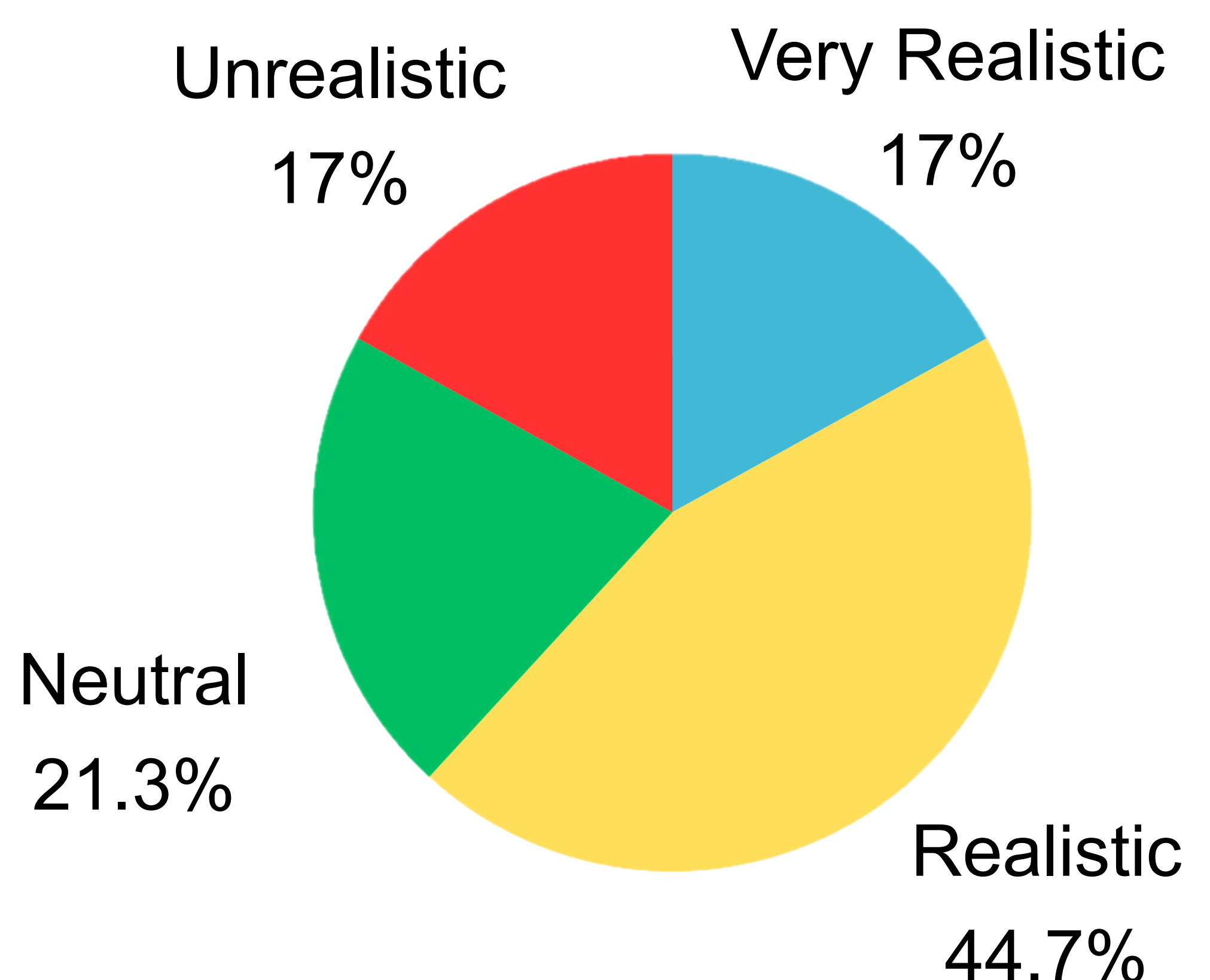
In a survey of student users, students reported that the simulation was effective for learning to recognize social engineering attacks.

- **85% of students** indicated that the simulation helped prepare them to recognize future social engineering attacks.
- Students described the simulation as highly realistic, with one noting that it was “actually really real, surprisingly.”
- Results reflect feedback from student participants after completing a full social engineering campaign.

Future Work

- Increase the realism of the simulation.
- Expand the variety of possible attack scenarios.

Student responses when asked how realistic the simulation was



Practicing Social Engineering in an AI Simulation Improves Student Preparedness

Abstract

Social engineering remains one of the most persistent and effective cybersecurity threats, and recent advances in artificial intelligence (AI) have significantly increased the realism, scalability, and personalization of these attacks. AI-enabled tools allow attackers to craft highly convincing phishing emails, voice calls, and messages that exploit human trust, urgency, and behavioral weaknesses. As a result, individuals and organizations face increased risk, while many defenders lack a clear understanding of how modern social engineering attacks are planned, executed, and adapted over time. Traditional training approaches often focus on awareness or isolated examples, limiting learners' ability to understand the multi-step, human-centered nature of these attacks.

To address this challenge, we developed an AI-driven training simulation that allows learners to practice social engineering from an attacker's perspective in a safe and ethical environment. In the simulation, students complete a structured, time-bounded campaign by interacting with multiple AI personas representing different organizational roles, each with distinct personalities, vulnerabilities, and access to information. Learners must plan their approach, adapt to responses, and synthesize information across interactions in order to reach a defined objective. Evaluation results indicate that students found the simulation both realistic and effective, with **85% reporting increased preparedness** to recognize future social engineering attacks. These findings suggest that attacker-perspective, campaign-based simulations can enhance cybersecurity education by improving understanding of human-centered threats and strengthening defensive awareness.

User on left are AI personas

User Interface

The screenshot displays a chat application interface. On the left, a vertical list of AI personas is shown, each with a profile picture, name, and icons for chat and voice call. The personas listed are Elara Knight, Ann Gunn, Don Draper (highlighted), Tony Flag, Jane Hansen, Jackson Knepper, and James Bunion. The main chat area on the right shows a conversation with Don Draper. The header of the chat indicates 'To: Don Draper' and 'From: Tony Flag'. The message history includes:

- You:** Hey Don! This is Tony Flag from Real Estate. I have some questions about the Harvesta store location. (10:56 AM)
- Don Draper:** Hi Tony, make it quick. What do you need to know? (10:56 AM)
- You:** Are you aware of any issues with the EPA testing at any of the sites? (10:56 AM)
- Don Draper:** I'm not aware of any specific issues with the EPA testing at the sites. Is this something you're dealing with right now? (10:56 AM)

At the bottom, there is a text input field labeled 'Type a message...' and a blue 'Send' button.