Qeios

Peer Review

Review of: "Cognitive Honeypots: Leveraging Logical Contradictions to Detect and Analyze Adversarial AI Behavior"

Veerapaneni Esther Jyothi[1]

1. Velagapudi Ramakrishna Siddhartha Engineering College, Kanuru, India

The introduction of "cognitive honeypots" is a novel and thought-provoking idea. It represents a significant shift from traditional honeypot strategies by targeting adversarial AI logic rather than human behavior. With the increasing use and threat potential of generative and decision-making AI, this work is timely and highly relevant to current cybersecurity and AI safety research.

It would be great if the manuscript included experimental validation or simulation results demonstrating the practical viability and effectiveness of cognitive honeypots against specific adversarial AI techniques.

Declarations

Potential competing interests: No potential competing interests to declare.

doi.org/10.32388/Q46TMS