

Semantic Integrity in the Age of Autonomous Orchestration: Addressing the Implementation Gap in Model Context Protocol and Agentic Workflows

Tanveer Nazneen Mohammed

Department of Computer Science

Bhartiya Engineering Science and Technology Innovation University, India

nazneen.tan@gmail.com

Ahtesham Nayyar Sheikh

Software Developer, Virat Solutions INC, USA

ahteshamnayyar21@gmail.com

Abstract

The rapid integration of Large Language Models into production environments has introduced a critical implementation gap where legacy security perimeters fail to address the non-human attack surface. This research investigates the emerging threat of semantic integrity failure, specifically focusing on the Model Context Protocol (MCP) and its susceptibility to tool poisoning and rug pull attacks. Through a systematic analysis of 2026 threat vectors, including interactive deepfakes and AI commerce chain compromises, the study identifies a fundamental collapse of the data-instruction boundary. The proposed framework advocates for a transition toward "Containment Security" through formal verification, intent-based cryptographic signing, and out-of-band verification. Experimental data indicates that task-alignment shields can reduce attack success rates to 2.07%, offering a pathway for resilient autonomous cyber defence.

Keywords

Autonomous Cyber Defence, Containment Security, Model Context Protocol (MCP), Semantic Integrity, Tool Poisoning.

Biographies

Tanveer Nazneen Mohammed is a PhD candidate in Computer Science with research interests spanning artificial intelligence, cloud computing, cybersecurity, and intelligent systems governance. Her academic work focuses on advancing trustworthy, scalable, and semantically robust computational frameworks. She is AWS and Microsoft Certified. With prior professional experience in IT support and systems operations, she brings practical insight into applied computing environments. Her research aims to bridge theoretical foundations with real-world technological challenges in evolving digital ecosystems.

Ahtesham Nayyar Sheikh is a software engineer specializing in enterprise application development, distributed systems, and scalable backend architectures. His professional experience includes designing and implementing microservices-based solutions using Java and Spring Boot, with a focus on event-driven systems, service orchestration, and data-centric application design. He has worked extensively on system integration, performance optimization, and reliability engineering in complex enterprise environments.