Minseok (Denis) Kim

Department of Computer Science and Engineering
Sungkyunkwan University

for8821@g.skku.edu · deniskim1.com · GitHub · Google Scholar

EDUCATION

Ph.D., Software	*Sungkyunkwan University* Advisor: Hyungjoon (Kevin) Koo	In progress
M.S., AI Systems Engineering	*Sungkyunkwan University* Co-advisors: Hyungjoon (Kevin) Koo and Jinyeong Bak. GPA 4.5 / 4.5.	
B.A., Software	*Sungkyunkwan University*	

RESEARCH INTERESTS

— Adversarial robustness of advanced AI systems — Retrieval-Augmented Generation, agent communication protocols (MCP, A2A), and autonomous-agent trust boundaries.

— Trust-boundary security analysis for autonomous AI agents, with emphasis on the shift from output safety to behavioral safety.

— Large language models for cybersecurity — binary analysis, reverse engineering, and code similarity detection.

— Agent-driven C-to-Rust transpilation: autonomous coding agents that translate legacy C into idiomatic, memory-safe Rust while preserving behavior.

— Defenses against data poisoning and prompt injection: lightweight post-retrieval filtering for RAG and runtime guards for tool-calling agents.

PUBLICATIONS

2026

[1] **Minseok Kim**, Hyungjoon Koo. “Security of Autonomous AI Agents: Trust Boundary-Based Attack Surface Analysis and Trends.” *Review of KIISC, Vol. 36, No. 2*, 2026.

2025

[1] Jiyong Uhm, **Minseok Kim**, Michalis Polychronakis, Hyungjoon Koo. “Fool Me If You Can: On the Robustness of Binary Code Similarity Detection Models against Semantics-preserving Transformations.” *FSE'26 (to appear)*, 2025.

[2] **Minseok Kim**, Hyungjoon Koo. “LLM-Based Drug Term Detection in Korean Messenger Conversations.” *J. Korea Inst. Inf. Security & Cryptology, Vol. 35, No. 6*, 2025.

[3] Giuk Kwon, **Minseok Kim**, Hyungjoon Koo. “Analysis of Watermarking for AI-generated Text.” *CISC-W'25*, 2025.

[4] **Minseok Kim**, Hankook Lee, Hyungjoon Koo. “Rescuing the Unpoisoned: Efficient Defense against Knowledge Corruption Attacks on RAG Systems.” *ACSAC'25*, 2025.

[5] Zhang Yiyue, **Minseok Kim**, Hyungjoon Koo. “Bridging Models and Agents: Protocol Architectures and Security in MCP & A2A.” *CISC-S'25*, 2025.

2024

[1] **Minseok Kim**, Hyungjoon Koo. “Trends in Attacks and Defenses against Retrieval-Augmented Generation (RAG) Systems.” *CISC-W'24*, 2024.

Last updated: July 2026.