Harish Kolla

hkolla03@gmail.com | hkolla.dev | linkedin.com/in/hkolla/ | github.com/Har1sh-k

Experience

Security Research Engineer - Center for Cyber Security Research, UND Jun 2024 – Present

- Built and shipped a **coding-agent security gate** that detects and blocks insecure changes before they reach disk, with application flow review, semantic IDOR and credential-leak guardrails, and Ed25519 identity/API-key infrastructure.
- Shipped an end-to-end **agentic security pipeline** fusing EDR and SIEM data with LLM and RAG embedding-based retrieval for confidence-scored, explainable alert correlation and automated triage, **saving 32+ analyst-hours/week**.
- Led **APT adversary emulation** (MITRE ATT&CK–mapped TTPs) across the Edge AI Infra (**1K+ devices**), working across teams to identify weaknesses, validating and hardening controls to **close 50+ high-risk control gaps.**
- Developed a **meta-agent pentesting orchestrator** that spawns on-the-fly subagents via modular tool adapters, drives testing with a headless browser, and iterates/pivots using confidence scoring; achieved **73% on XBOW benchmarks**.

Security Research Assistant - Penn State(w/ IBM Watson Research, AT&T) Aug 2022 – May 2024

- Automated 20+ security workflows and policies for Virtual Multi-Factor Authentication(MFA), **configured Next-Gen Firewall using Zero-Trust Architecture**, strengthening secure multi-channel communications by 5x.

Senior Software Security Engineer - Capgemini, India Feb 2021 – Jul 2022

- **Performed threat modeling**, static/dynamic code testing, manual code inspection and development, architecture reviews, and penetration testing of web applications and IT architectures to identify vulnerabilities and security defects.

Security Engineer - Siemens & HBL , India May 2019 – Feb 2021

- Mitigated 22+ critical vulnerabilities through **Vulnerability Assessment and Penetration Testing(VAPT)**, static and dynamic analysis, app-layer firewalls& Data Loss Prevention (DLP) measures,resulting 40% increase in app security

Vulnerability Research & Coordinated Disclosure

Disclosures / Bounty : **Anthropic** (Claude SDK), **OpenAI** (Codex), **Google** (Gemini), **Apache** Airflow, **Gitlab**. Published **14 credited CVEs**.

CVE-2026-9198: **Unauthenticated RCE** – IBM Langflow; **CVSS 9.8, CISA KEV**.
CVE-2026-64824 / CVE-2026-64825: **Backup restore RCE and arbitrary file write** – Home Assistant Core.
CVE-2026-14886: **Cross-namespace identity deletion** – HashiCorp Vault Enterprise.
CVE-2026-54305: **Cross-tenant credential takeover via IDOR** – n8n.

Projects

SASTBench – Scanner-Agnostic Benchmark for Agentic SAST | *Python, TypeScript, Rust, Swift* Apr 2026

- Open-sourced **206-case evaluation corpus and harness** across 7 languages, with 147 historical CVEs, safe negative controls, PR-diff tasks, and model based disclosure-date filtering to reduce advisory-memorization risk.

XFire (CrossFire) – Multi-Agent Adversarial Security Review Tool | *Python, CI/CD* Mar 2026

- Built **security tool (PyPI) for automated PR reviews via GitHub Actions**, runs 3 AI agents in parallel blind review to detect vulnerabilities and production-breaking bugs, routes disputed findings via adversarial debate engine.

SecureVibes – AI-Native Multi-Agent AppSec Scanner | *Python, Claude Agent SDK & Skills* Dec 2025

- Co-built an architecture-aware, multi-agent AppSec scanner; implemented Code-review agent(SAST) and DAST agent/skills to validate SAST findings via dynamic requests, producing lower-noise reports across 11+ languages.

Open Source Contributions

NVIDIA Garak – LLM Red-Teaming Framework | *Contributor (PR #1489)* Nov 2025

- Implemented **BadCharacters probe** to jailbreak LLM with perturbations (ZWS, homoglyphs, BiDi, backspace deletion)

Apache Airflow | *Contributor (PR #66751)* May 2026

- Identified, fixed **SQLInjection** in Hive metastore bookkeeping & Presto partition predicates with identifier allow-listing

HashiCorp Vault | *Contributor (PR #32004)* Jun 2026

- Reported, patched **TOCTOU race condition** in ACME order finalize enabling concurrent certificate double-issuance

Skills

Programming & Systems: Python, TypeScript, SQL, C++, AWS, Docker, Linux, PostgreSQL, GitHub Actions.
Product Security: Secure SDLC, CI/CD, Code-review, Threat-modeling, SAST/DAST, API/ Supply-chain, AuthN/AuthZ.
Standards & Protocols: NIST AI RMF, OWASP AI/LLM Top 10, CWE, MITRE ATT&CK, MCP.
AI/ML Sec: Agent/Model evaluation, Evaluation harnesses/ graders, LLM red-teaming, Sandboxed Tool-use, Guardrail Tests

Education

The Pennsylvania State University - M.S., Cybersecurity Analytics and Operations May 2024

Gandhi Institute of Technology & Management - B.Tech., Electronics & Communications May 2020