Pooja Kiran

Seeking ML Security Engineer Roles

Greater Phoenix Area, AZ | +1 (480) 776-7745 | poojakiranbhardwaj@gmail.com | github.com/poojakira
linkedin.com/in/poojakiran | poojakira.github.io

F-1 OPT (EAD Pending) · Avail. July 6, 2026 · Future H-1B sponsorship needed

SUMMARY

ML Security Engineer securing LLM, RAG, and agentic AI pipelines end-to-end across GPU clusters, retrieval layers, and the model supply chain. Specialist in adversarial ML, OWASP LLM Top 10 defense, supply-chain integrity (SafeTensors, Ed25519, SLSA), and secure MLOps. Author of six open-source AI-security tools mapped to MITRE ATLAS, NIST AI RMF, and EU AI Act.

TECHNICAL SKILLS

ML Security: Threat modeling (STRIDE), MITRE ATLAS, NIST AI RMF, OWASP LLM Top 10, AI red teaming

LLM / RAG Security: Prompt-injection defense (direct/indirect), RAG poisoning mitigation, retrieval-layer hardening, output filtering, canary tokens, multi-tenant isolation

Model Supply-Chain Security: Pickle/PyTorch artifact scanning, SafeTensors validation, Ed25519 signing, provenance/SBOM, SLSA, SARIF CI gates

Adversarial ML: FGSM, PGD, C&W, AutoAttack, adversarial training (Madry), TRADES, randomized smoothing

Privacy Attacks & Defenses: Membership inference (Yeom/Shokri), model inversion (Fredrikson), DP-SGD (RDP accounting), encrypted embeddings

Secure MLOps: CI/CD security gates, data lineage, RBAC/JWT, audit logging, drift detection, observability, incident response

Cloud & Infra: AWS (IAM, VPC, encryption), Docker, Kubernetes, FastAPI, Qdrant, Redis, Prometheus/Grafana, GitHub Actions

Programming & Tools: Python, PyTorch, scikit-learn, spaCy/Presidio, Git, pytest

EXPERIENCE

Independent AI/ML Security Researcher — **Aug 2024 – Present**
Self-directed Research — *Tempe, AZ (Remote)*

- Engineered GPU multi-tenant side-channel analysis that eliminated nearly 50% of memory-leakage vectors, enforcing NIST AI RMF–aligned isolation controls.
- Architected multi-agent adversarial ML attack simulations mapped to MITRE ATLAS, exposing 7 high-risk exploitation paths across LLM and RAG pipelines.
- Designed a privacy-preserving RAG retrieval layer combining differential privacy and encrypted embeddings, hardening LLM retrieval against inference and exfiltration attacks.
- Authored 4 open-source model supply-chain scanners enforcing pickle opcode analysis, SafeTensors validation, and Ed25519 signing through SARIF-gated CI.
- Delivered 6 reproducible attack/defense benchmark suites across LLMs, RAG, model artifacts, and privacy, each with documented threat models and failure modes.

Cybersecurity Innovation Researcher — Technology Innovation Lab — **Aug 2025 – Nov 2025**
Arizona State University and Honeywell Aerospace — *Tempe, AZ*

- Conducted threat assessment across Passenger Service Systems and aviation integrations, mapping cloud-security risk gaps in connected-aircraft platforms.
- Identified IoT attack surfaces in connected aircraft and engineered a validated mitigation prototype, delivered 100% on schedule within a 100-day incubator.
- Drove the full innovation lifecycle from problem validation through prototyping to an executive-level technical pitch to Honeywell engineering leadership.

Graduate Teaching Assistant (IT Grader) — **Jan 2025 – Oct 2025**
Ira A. Fulton Schools of Engineering, Arizona State University — *Mesa, AZ*

- Assessed 100+ graduate submissions against secure-coding, NIST/ISO compliance, and configuration-management standards across 3 advanced IT courses.
- Evaluated infrastructure-security and secure-SDLC practices, reinforcing secure system design for graduate engineering cohorts.

Undergraduate Research Fellow — KSCST Funded **Jan 2023 – Feb 2024**
M. S. Ramaiah University of Applied Sciences *Bengaluru, India*

- Secured a competitive KSCST government research grant, selected from 5,900+ proposals (approximately 25% acceptance), for reinforcement-learning CubeSat communication research.
- Co-authored and presented a peer-reviewed IEEE INDICON 2023 paper on RL-driven satellite systems to IEEE committees and industry evaluators.

SELECTED SECURITY PROJECTS

LLM-Guard-Scanner — Multi-layer OWASP LLM Top 10 scanner. *May 2026*

- Detected direct/indirect prompt injection, RAG poisoning, and PII/secret leakage via pattern and embedding similarity, NER, entropy analysis, and canary-token tracking.
- Deployed as a FastAPI service emitting standards-compliant SARIF for GitHub Advanced Security CI gates.

Model-Supply-Chain-Auditor — ML artifact integrity scanner. *May 2026*

- Blocked malicious model loads via AST-walk pickle opcode analysis flagging dangerous `REDUCE`/`GLOBAL`/`BUILD` callables and unsafe `torch.load` usage.
- Enforced SafeTensors validation and Ed25519 signing with a SHA-256 provenance chain, gated in policy-as-code CI rejecting unsigned artifacts.

ML-Privacy-Attacks — Privacy leakage + compliance analysis. *May 2026*

- Implemented Yeom (2018) and Shokri (2017) membership inference plus Fredrikson model inversion, measuring a 0.42 MIA advantage (4x the 0.10 threshold) mapped to EU AI Act Art. 10/15.
- Reduced privacy budget to epsilon=1.16 at sigma=4.0 via a DP-SGD defense with RDP accounting.

Adversarial-Robustness-Toolkit — Robustness benchmarking. *May 2026*

- Benchmarked FGSM, PGD-20/100, C&W, and AutoAttack against CIFAR-10 ResNet-18 at epsilon=8/255 L-infinity.
- Quantified standard-model collapse to 0% robust accuracy under PGD, with Madry adversarial training recovering nearly 45%; implemented TRADES and randomized smoothing.

docquery — Secure financial-document RAG pipeline. *Jun 2026*

- Built a Qdrant + BGE-reranker + Redis + FastAPI stack with tenant-scoped collections, JWT auth, and per-tenant rate limiting.
- Mitigated RAG poisoning via imperative-instruction detection, XML delimiter sandboxing, source-signature verification, and API-boundary PII redaction.

PulseNet-RUL-Forecasting — Secure MLOps pipeline (NASA C-MAPSS). *Mar 2026*

- Modeled a full STRIDE threat surface with per-tenant RBAC, AES-GCM encryption, and an append-only hash-chained audit trail.
- Optimized inference to 2.7 ms mean latency and 52K samples/sec throughput behind SARIF-gated CI; tuned an adversarial telemetry guard to 1.0 recall.

EDUCATION

M.S., Information Technology (Security focus) **Aug 2024 – May 2026**
Arizona State University — GPA 3.87 *Tempe, AZ*

Coursework: Advanced Information Systems Security, Network Forensics, Cloud Security, Secure Cloud Architecture.

B.Tech, Computer Science & Engineering **Aug 2019 – Aug 2023**
M. S. Ramaiah University of Applied Sciences — CGPA 8.44 *Bengaluru, India*

CERTIFICATIONS & PUBLICATIONS

- **Technology Innovation Lab** — Honeywell Aerospace & ASU (Nov 2025): 100-day connected-aviation cybersecurity incubator.
- **AWS Academy Graduate** — Cloud Security Foundations (Nov 2025): IAM, VPC security, encryption in transit/at rest.
- **IEEE INDICON 2023** — "A Personalized E-Learning System Using Reinforcement Learning Through Satellite," NIT Warangal (Doc ID 10440852).