Chanakya Thotakura

Security Engineer | Trust & Safety | Applied AI
Bay Area, California · chanikya1@icloud.com · linkedin.com/in/chanikya6163 · github.com/Chanikya5793

Summary

Security engineer with 5+ years of hands-on engineering experience in Trust & Safety, fraud and abuse, and security investigations. Experienced in investigating and acting on high-impact threats and online threat actors, with a focus on phishing, account compromise, and coordinated abuse. Builds detection systems and agent workflows that connect evidence, policy decisions, and controlled execution.

Technical Skills

Trust and Safety: Fraud and scams, phishing, account takeover, campaign investigation, threat attribution, enforcement policy.
AI and agents: Agent orchestration, context engineering, Model Context Protocol (MCP), tool calling, typed handoffs, RAG, grounding, abstention.
Evaluation: Adversarial testing, abuse classifiers, expert feedback, age assurance, bias reviews, quality and accuracy assessment.
Engineering: Python, SQL, TypeScript, FastAPI, SQLite, Firestore, AWS, Google Cloud, Cloud Run, Cloud KMS, Docker, GitHub Actions.

Core Competencies

- Investigated high-impact threats and online threat actors spanning phishing, account compromise, fraud, and coordinated abuse. Correlated account, content, and infrastructure evidence, and acted on findings through containment, enforcement, and escalation workflows.
- Used Python and SQL to analyze large event datasets, identify behavioral patterns, and trace connections between accounts and infrastructure during campaign investigations.
- Designed detection and evaluation capabilities for abuse workflows, translating open-ended requirements into shared methods, platform decisions, and scoped engineering work.
- Orchestrated agent workflows using scoped context, typed tool contracts, handoffs, and approval gates. Bounded retries and traced execution to separate model proposals from authorized actions.
- Evaluated conversational agents and abuse classifiers against injection, jailbreaks, social engineering, and policy manipulation. Used failure analysis to refine expert feedback and regression cases, reviewing accuracy, bias, and analyst effort.
- Translated enforcement policy into evaluation cases for emerging abuse tactics, coded language, and age-related safety. Calibrated age signals and fallback behavior where confidence was insufficient.
- Coordinated work across policy, operations, engineering, and model research teams, and mentored engineers. Shared investigation methods and evaluation findings through technical guidance, publications, and community contributions.

Experience

ControlForge Security Aug. 2026 - Present

- Released a Python detection platform covering phishing, impossible travel, sensitive-data access, encoded PowerShell, and unexpected privilege grants.
- Normalized security events and evaluated version-controlled Sigma-style rules through CLI and FastAPI interfaces, returning stable alert IDs, matched evidence, and investigation-ready explanations.
- Kept endpoint probes read-only and recorded results in SQLite. Deployed the Dockerized API on AWS, published signed artifacts to S3, and automated release checks with GitHub Actions.

Northwest Student Planner Jun. 2026 - Present

- Built a RAG copilot using Meta Muse Spark 1.3, user-scoped lexical retrieval, and six read-only MCP tools.
- Orchestrated context retrieval, tool calls, citation validation, and typed action proposals; required user confirmation and revision-checked, idempotent writes before execution.
- Evaluated retrieval, grounding, abstention, proposals, and isolation with 30 regression queries; enforced user scopes, Firebase Authentication, and KMS-wrapped keys, and deployed FastAPI on Cloud Run.

ManaSplit Nov. 2025 - Present

- Published an iOS expense, messaging, and calling app on the Apple App Store using React Native and native Swift modules, with linked-device Signal sessions supporting hardware-backed keys.
- Built an opt-in Security Center combining breach, dark-web, infostealer, and phishing intelligence with deterministic account-compromise risk scoring.
- Combined Web Risk, DNS/RDAP, TLS, certificate-transparency, and k-anonymous password checks. Generated evidence-bounded AI explanations without raw passwords, tokens, or provider secrets; protected identities with encryption and scoped blind indexes.

AIMERS | *Machine Learning Engineering Intern* Mar. 2025 - Aug. 2025

- Built and deployed a Python voice assistant on AWS, integrating speech recognition, supervised intent classification, database retrieval, and text-to-speech.
- Secured retrieval with parameterized SQL, query allowlists, schema validation, least-privilege IAM, managed secrets, encryption, output redaction, and audit logging.
- Tested prompt injection, jailbreaks, replay, and unauthorized tools; constrained execution through intent validation, transcript normalization, rate limits, bounded tools, and safe refusals.

Education and Certification

Northwest Missouri State University | *M.S., Applied Computer Science*

V.R. Siddhartha Engineering College | *B.Tech., Computer Science & Engineering*

Certification: Google Cloud Associate Cloud Engineer.