ROSHANI SAH

Cybersecurity Researcher | AI/LLM Security Testing & Red Teaming
Offensive Security Enthusiast | CTF Player | Web & API Pentesting
roshnisah543@gmail.com | LinkedIn | GitHub | Kathmandu

PROFESSIONAL PROFILE

Cybersecurity-focused BScIT student with hands-on experience in web and API security testing, offensive security techniques, and AI/LLM security research. Practised in CTF challenges, penetration testing methodologies, and vulnerability assessment, with a growing specialization in prompt injection, jailbreak testing, and bias discovery in generative AI systems. Experienced in building security-focused tooling, including a CTF platform designed to teach vulnerability identification. Exploring adversarial testing of large language models and agentic AI systems aligned with the OWASP LLM Top 10 framework. Seeking to apply offensive security skills and AI security knowledge to red teaming of advanced AI models.

CORE COMPETENCIES

AI / LLM Security

- AI & LLM Security Testing
- Prompt Injection & Jailbreak Techniques
- Bias Discovery & Adversarial Input Crafting
- OWASP LLM Top 10 Framework
- AI Red Teaming Methodologies
- Risk Assessment & Threat Modelling

Application & Infrastructure Security

- Web Application Penetration Testing
- API Security Testing
- Offensive Security / Ethical Hacking
- Vulnerability Assessment & Reporting
- OWASP Top 10
- Wireless Security Testing (802.11)

Tools & Technologies

Python | Bash | JavaScript | HTML/CSS | Git & GitHub | Linux | ESP-32 | Postman | VS Code | LLM Prompting & Adversarial Testing Tools | Medium (Technical Writing)

SECURITY RESEARCH & PROJECTS

Can-You-Find-The-Flaw — CTF Game for Beginners 2026
Personal Project | Security Education Platform

- Designed and developed a Capture The Flag game aimed at teaching beginners how to identify and exploit common security vulnerabilities through structured challenges
- Created challenge scenarios covering web application flaws, input validation weaknesses, and logical vulnerabilities — demonstrating deep understanding of attacker methodology and vulnerability patterns
- Applied offensive security thinking to design realistic, educational attack scenarios that mirror real-world vulnerability classes

Rogue WiFi Access Point — Wireless Security Research 2025
Personal Project | ESP-32 Hardware

- Built a rogue WiFi access point using an ESP-32 microcontroller to demonstrate wireless network attack techniques and man-in-the-middle scenarios
- Researched 802.11 protocol weaknesses and implemented practical exploitation, reinforcing offensive security skills transferable to AI system attack surface analysis

AI / LLM Security Research & Practice Feb 2024 –Present
Self-Directed Research | Ongoing

- Studying and practising prompt injection techniques, jailbreak methodologies, and bias discovery approaches against publicly available LLM systems
- Researching the OWASP LLM Top 10 framework to systematically understand and test for common vulnerabilities in large language models and AI-powered applications
- Publishing technical security research and writeups on Medium to document findings and contribute to the cybersecurity community

Web Application Development Projects Feb 2024 – March 2025
CodeKickStarters | FinproAdvisors

- Built CodeKickStarters, a roadmap platform for beginner developers, demonstrating full-stack web development skills and understanding of application architecture
- Developed FinproAdvisors, a financial institute website, handling sensitive data presentation and gaining firsthand understanding of web attack surfaces in financial applications

EDUCATION

BSc Information Technology (BScIT) 2024
Presidential Graduate School, Kathmandu

- Focused coursework on cybersecurity, networking, and software development
- Actively pursued AI/LLM security research and offensive security practice alongside academic programme

High School Diploma 2020
Capital College and Research Center | GPA: 3.53

CTF PARTICIPATION & COMMUNITY

- Active CTF player — practising vulnerability identification, exploitation techniques, and flag-based challenge solving across web, crypto, and misc categories
- Published technical security research and vulnerability writeups on Medium (@roshtech855)
- Continuous self-directed learning in offensive security, AI red teaming, and adversarial ML through hands-on practice and community engagement

ADDITIONAL SKILLS

- Languages: English, Nepali, Hindi
- Technical writing and documentation — experience communicating security findings in structured, reproducible formats
- Strong analytical and creative problem-solving skills applied to offensive security challenges and adversarial testing scenarios