Aavishkar Gautam

+971 56-547-1788 | aavishkar.gautam@nyu.edu | Linkedin | GitHub

Education

New York University Abu Dhabi (NYUAD) — Abu Dhabi, UAE
Bachelor of Science in Computer Science, Minor in Applied Mathematics, GPA: 3.84 — *Graduation Date: May 2025*

Research Experience

Haven Lab NYUAD - AI Safety Research — May 2025 – Present
Research Assistant — *Abu Dhabi, UAE*

- Engineered novel weight-poisoning backdoor attacks using weight orthogonalization to embed persistent backdoors in transformer null spaces, achieving stealthy attacks that survive model fine-tuning.
- Identified and analyzed critical "super-weights" in BERT architectures that govern model safety properties, demonstrating how targeted parameter manipulation degrades classification robustness.
- Co-authored research paper on refusal mechanisms in frontier models (work in progress). Working Draft

Modern Microprocessors (MoMA) Lab NYUAD — May 2024 – August 2024
Summer Research Assistant — *Abu Dhabi, UAE*

- Researched backdoor attacks in centralized and federated learning frameworks, analyzing their stealth, efficiency, and resilience under distributed training.
- Implemented multi-trigger backdoor attacks, experimenting with visible, invisible, and label-consistent techniques to study coexistence and interference across triggers.
- Designed and executed experiments on CIFAR-10 and MNIST using ResNet-18, demonstrating that multiple triggers can coexist without degrading accuracy or attack success.

Work Experience

RAIN Agency — July 2025 – September 2025
Software Development (SDE) Intern — *Remote*

- Implemented on-device LLM inference, cutting latency and operational costs while improving overall user experience.
- Integrated speech-to-text APIs for automatic speech recognition (ASR), enabling real-time transcription features.
- Performed extensive model benchmarking for on-device deployment in healthcare applications, evaluating 20+ LLMs for latency, memory efficiency, and clinical accuracy on resource-constrained devices.

PwC Middle East — June 2024 – August 2024
Software Engineering (SWE) Intern — *Dubai, UAE*

- Designed and delivered an AI-powered proposal builder as a PowerPoint add-in, streamlining workflow for over 2,000 consultants.
- Enhanced client experience by developing and debugging the Emerging Technology(EmTech) Lab website.
- Experimented and integrated APIs for Brain Computer Interface(BCI) devices.

Projects

Doc-Chat | *React, FastAPI, LangChain, Python, Streamlit*

- Developed a web application allowing users to upload documents and engage in conversations with them, utilizing advanced RAG techniques.

PortalPeek | *Selenium, Twilio, FastAPI, LangChain, MongoDB*

- Developed a system that scrapes university student portals, uses AI to classify announcements based on user preferences, and sends scheduled email updates directly to users' inboxes.

Technical Skills

Languages: Python, C/C++, SQL, Postgres, JavaScript, HTML/CSS
Frameworks: React, Node.js, Flask, WordPress, FastAPI, LangChain
Developer Tools: Git, Docker, Google Cloud Platform, VS Code, Visual Studio, PyCharm, IntelliJ, Eclipse
Libraries: Pandas, NumPy, Matplotlib, PyTorch