SANJANA NAMBIAR

☎ +971 569454313 Sharjah, UAE

svn9705@nyu.edu sanjana-nambiar1967 Sanjana-Nambiar Portfolio

EDUCATION

Bachelor of Science in Computer Science, New York University Abu Dhabi **August 2021 – May 2025**
Minors: Applied Mathematics and Engineering
Relevant Coursework:

- Deep Learning and LLM-based Generative AI Systems from NYU Courant Institute of Mathematical Sciences New York
- Processing Big Data for Analytics Applications from NYU Courant Institute of Mathematical Sciences New York
- Computer Security and Cryptography from NYU Courant Institute of Mathematical Sciences New York
- Software Engineering from NYU Abu Dhabi
- Applied Machine Learning from NYU Abu Dhabi
- Introduction to Machine Learning from NYU Tandon School of Engineering New York

CBSE Higher Secondary (AISSCE), Sharjah Indian School **April 2019 - March 2021**
Ranked 3rd school-wide, topped Engineering Graphics and Chemistry.

PUBLICATIONS

Nambiar, S., & Pöpper, C. **JailFact-Bench: A Comprehensive Analysis of Jailbreak Attacks vs. Hallucinations in LLMs**. Presented at *SiMLA 2025* Workshop, co-located with ACNS 2025, Munich, Germany. To appear in the ACNS 2025 post-proceedings (Springer LNCS). **June 2025**

Feuer, B., Goldblum, M., Datta, T., **Nambiar, S.**, Besaleli, R., Dooley, S., Cembalest, M., & Dickerson, J.P. **Style Over Substance: Failure Modes of LLM Judges in Alignment Benchmarking**. Published at *ICLR 2025* Conference. **January 2025**

Nambiar, S., Hegde, C., & Cohen, N. **Synthetic Fine-Tuning as a Defense Mechanism in LLM PII Attacks**. Submitted to the NeurIPS 2024 LLM Privacy Challenge – Blue Team Track. **December 2024**

EXPERIENCES

Research Assistant, Cyber Security and Privacy (CSP) Lab **February 2024 - Present**
New York University Abu Dhabi (Capstone Research) - *Prof Christina Pöpper* Abu Dhabi, UAE

- Created JailFact-Bench, a 99-prompt dataset comparing jailbreak and factual prompts by curating, refining, and validating paired inputs, enabling comprehensive analysis of LLM vulnerabilities.
- Uncovered a 25–40% factual drift in jailbreak outputs versus factual prompts by conducting a statistical analysis on semantic similarity and human-annotated accuracy, redefining risk metrics for adversarial inputs.
- Automated data pipelines and toxicity analysis using Python and APIs (OpenAI Moderation, Perspective), delivering actionable insights to improve LLM alignment and mitigate hallucinations.

Research Assistant, Data Intelligence and Computation in Engineering (DICE) Lab **July 2024 - January 2025**
Tandon School of Engineering, New York University - *Prof Chinmay Hegde* New York, US

- Conducted research on defense strategies for AI models against Jailbreak attacks, achieving a significant reduction in attack success rates from 71.42% to 12.24% through synthetic fine-tuning with counter-fake PII datasets.
- Co-authored a paper presented at ICLR 2025, introducing a benchmark to evaluate biases in LLM judges across alignment metrics like helpfulness, honesty, and harmlessness.
- Discovered and analyzed reference stuffing as an injection attack vector, reducing LLM performance in preference ranking systems by 49%.

Software Engineer Intern **May 2024 - August 2024**
Letsrise Academy, Abu Dhabi Abu Dhabi, UAE

- Designed and implemented a scalable user analytics and admin dashboard using Figma, Flask, and PostgreSQL, handling data pipelines for 250+ user datasets to ensure robust insights on entrepreneurial traits.
- Automated data pipelines using PostgreSQL for real-time user data processing, ensuring system performance and scalability.
- Developed and deployed the dashboard using Flask, Nginx, Gunicorn, and Cloudflare, optimizing system performance by 25%, resulting in successful investor pitches and onboarding four new users.

AI Peer Mentor, Design Lab **May 2024 - June 2024**
New York University Abu Dhabi Abu Dhabi, UAE

- Mentored an international team of high school students to develop an AI-based educational curriculum, guiding them through research and project execution.

- Delivered lectures on crafting effective pitches, citing academic papers, and developing innovative ideas, fostering teamwork and critical thinking.

Research Assistant, E-Brain Lab **March 2024 - May 2024**
New York University Abu Dhabi Abu Dhabi, UAE

- Researched backdoor attacks in neural networks using activation clustering, focusing on cybersecurity for large-scale autonomous systems and enhancing surveillance technologies.

Research Assistant - IoT Environmental Station **February 2023 - May 2024**
Mubadala Arabian Center for Climate and Environmental Sciences (ACCESS) Lab Abu Dhabi, UAE

- Designed and deployed 3 IoT-based Environmental Monitoring Stations with Raspberry Pi 4 and advanced sensors (BME280, SCD30, NextPM), achieving 95% data accuracy and reducing assembly time from one month to one day through workflow optimization and 3D-printed sensor bases.
- Implemented real-time monitoring in remote areas by integrating cellular connectivity, reverse tunneling (ngrok), and automated data collection processes using shell scripts.
- Conducted rigorous Python-based testing and debugging to ensure precise, reliable data acquisition, enhancing system scalability and operational efficiency for environmental research.

HONORS & AWARDS

Highly Commended - Centre for Urban Science and Progress (CUSP) Data Dive 2024 **February 2024**

- Analyzed cycling and walking's impact on London's air quality, uncovering that pollutant exposure increases in high-density areas. Proposed policy recommendations for healthier urban transportation.

Second Place - NYUAD International Hackathon for Social Good (Team Qatrah) **April 2023**

- Built a quantum-enhanced water distribution system. Leveraged Python's NetworkX for graph modeling and QUBO for sensor placement to optimize fault detection and enhance system robustness. (GitHub)

Finalist - CSAW'22 Cybersecurity Games and Conference (Hack My Robot) **November 2022**

- Built and tested a ROS-Noetic TurtleBot3 on Ubuntu 20.04. Used RViz for motion tracking and conducted DoS attacks with Kali Linux to uncover critical system vulnerabilities.

Super Achiever - Middle East Education Award **August 2021**

- Recognized for academic excellence at the 7th India Middle East Education Awards. Achieved a 96.2% score in Grade 12, ranking among the top achievers. (Award)

LEADERSHIP & COMMUNITY ENGAGEMENT

Volunteer, 12th Annual International Hackathon For Social Good **May 2024 - June 2024**

- Assisted in managing logistics and participant registration for a global hackathon with 180 participants from 50 nationalities.
- Supported participants across campus, ensuring a collaborative and seamless experience for students and mentors.

Events Board Member, Undergraduate Student Government **February 2023 - May 2024**

- Spearheaded university-wide events, including the 2024 Gala and Valentine's Day programs, coordinating logistics and engaging over 1,000 students.
- Enhanced campus culture by innovating event planning and ensuring smooth execution.

Communications Officer, Melting Pot **September 2022 - June 2023**

- Led cross-departmental collaborations to secure approvals and book venues for campus-wide events.
- Designed promotional materials to boost student engagement and streamlined communication strategies.

Resources Core Team Member, weSTEM (Women Empowered in STEM) **March 2022 - December 2022**

- Developed a comprehensive guide for 500+ CS and Math undergraduates, compiling alumni insights and course pathways.
- Empowered students with actionable resources to navigate academic and career growth.

Sustainability Committee Member, Undergraduate Student Government **February 2022 - June 2022**

- Advocated for sustainability integration into campus programs by collaborating with university leadership.
- Conducted campus-wide surveys and presented data-driven proposals to administrators to implement green initiatives.

SKILLS

Programming Languages	C++ (4 yrs), Python (3 yrs), C (2 yrs), JavaScript (2 yrs), MATLAB (1 yr), GoLang (1 yr), Java (1 yr), VHDL (1 yr)
Machine Learning	PyTorch, TensorFlow, Keras, Hugging Face, Transfer Learning, Fine-tuning (LoRA), Synthetic Dataset Creation, Parameter Optimization (WandB)
Big Data & Analytics	Apache Hadoop, Apache Spark, HiveQL, Presto, Hadoop, Yarn Scheduler
Web Development	Node.js, Express.js, HTML5, CSS3, Flask, PostgreSQL, MySQL, MongoDB, Gunicorn, Nginx, Firebase, Flutter, Dart
Advanced Computing	Expertise in High-Performance Computing (HPC), Cluster Computing, Slurm Scheduler, Parallel Programming, Efficient Memory Management, Large Dataset Handling
Hardware Skills	Raspberry Pi, Arduino, Soldering, 3D Printing, ROS Noetic
Languages	English (Proficient), Hindi (Proficient), Malayalam (Native), Arabic (Beginner)