Neeladri Bhuiya

nbhuiya@umass.edu | linkedin.com/neeladri-bhuiya | github.com/zawedcvg

Education

University of Massachusetts Amherst, MS in Computer Science — May 2026

- GPA: 4.0/4.0
- Relevant courses: Reinforcement Learning, Neural Networks

National University of Singapore, BS in Computer Science — May 2024

Experience

AI Safety Intern, A10 Networks — June 2025 - August 2025

- Reduced peak memory usage of the existing GCG attack by **50%** (2× memory efficiency), enabling attacks with larger number of prompts.
- Designed and implemented **PLAGUE**, a novel agentic multi-turn jailbreak attack framework that improved the attack success rate (ASR) by up to **40 % points** compared to the prior SOTA in models like Claude Opus 4.1.
- Designed an orchestrator to coordinate multiple agents and integrated a RAG-based module for lifelong retrieval of strategies and in-context learning for PLAGUE.
- **Published paper**: *PLAGUE: Plug-and-play Framework for Lifelong Adaptive Generation of Multi-turn Exploits* | **First Author** | **ICLR 2026**

Extern Graduate Researcher, IBM — Jan 2025 - April 2025

- Led a team of 4 to conduct research on creating a correctness classifier using LLM's uncertainty estimation for code generation under the supervision of Dr. Andrew McCallum and Dr Veronika Thost.
- Engineered an optimized TokenSAR implementation with Tree-sitter, improving runtime efficiency and boosting correctness prediction accuracy by 15% over uncertainty-based baselines.

Graduate Researcher, CIIR and IESL — December 2024 - Current

- Conducting research under Dr. Andrew McCallum (IESL) and Dr. Hamed Zamani (CIIR) on leveraging **box embeddings** for evaluating LLM performance.
- Engineered scalable training pipelines for boxes, reducing end-to-end runtime by **2×** and accelerating experimentation.
- Developed an **nD-to-2D visualization technique** for box embeddings, improving interpretability. Also uncovered strong correlation between box volume and instruction complexity.

Researcher, NUS — August 2023 - May 2024

- Conducted a final year research project under the supervision of Dr. Stefan Winkler and Dr. Viktor Schlegel, leading to a deeper understanding of **LLM's performance** under strong **adversarial attacks**.
- Developed a novel adversarial attack which affects state-of-the-art LLM's performance **up to 45%**
- **Published paper**: *Seemingly Plausible Distractors in Multi-Hop Reasoning* | **First Author** | **EMNLP 2024**

Selected Projects

Silica Ranked — Github

- Built a high-performance backend in Rust with PostgreSQL, parsing player logs 24/7 to generate detailed player statistics.
- Optimized log processing to handle **10,000+ line files in <9 ms**, enabling real-time analytics.
- Developed and deployed an interactive React + Go + Neon frontend with CI/CD pipelines, attracting thousands of global impressions.

Skills

Languages: Python, Rust, Go, C++, Java, C#, SQL, JavaScript
GenAI: PyTorch, TensorFlow, Hugging Face, scikit-learn, Prompt Engineering
System and Tools: Linux, PostgreSQL, MongoDB, Git, Vim