# Nengneng Yu

+1 484-995-8987 | ynn1999@umd.edu | College Park, MD

## Professional Summary

- **Focused on data-driven system research:** leading projects that apply and implement data-driven approaches for computer systems, network, distributed systems, and biological data analysis. With hands-on experience on database systems and operating systems. Published and submitted works to conferences such as IEEE S&P and HotNet.
- **Competency:** Six year in computer science/engineering major in the US, fluent in English, and experienced in team collaboration.
- **Technical Proficiency:** C++/C, Java, Python, PyTorch, Pandas, Numpy, Linux, Machine Learning, Deep Learning, system programming, data structures, and algorithms.

## Education

- **University of Maryland College Park** — College Park, MD
  *Doctor of Philosophy in Computer Science* — *Aug 2023 - Present*
  *Advisor: Zaoxing(Alan) Liu https://zaoxing.github.io/*
  *GPA: 3.77/4.00*
- **Boston University** — Boston, MA
  *Bachelor of Science in Computer Engineering* — *Sep 2019 - May 2023*
  *Graduated with Magna Cum Laude*

## Research Projects

- **Data-driven Generative Analysis System for Cancer Detection** — Feb 2024 – Present
  *Mentor: Zaoxing(Alan) Liu, Yuefan Wang* — *Froot Lab, UMD & Johns Hopkins Medicine*
  - Developed an AI-based generative system addressing the limitations of traditional bioinformatics in handling high-dimensional, small-sample proteomics data for gastric cancer detection and stage differentiation.
  - Designed a pipeline combining Diffusion models for synthetic data generation and XGBoost for biomarker identification, implemented using PyTorch.
  - Achieved near 100% accuracy in gastric cancer NAT/Tumor classification. Improved CPTAC dataset performance by 10-20%, Korean dataset from 70% to over 90% on weighted accuracy among cancer stages.
- **Accelerating the mapping system for Micro Aerial Vehicle (MAV) navigation** — Sep 2023 – Present
  *Mentor: Peiqing Chen* — *UMD*
  - The performance of MAV autonomous navigation is limited by the slow updates and queries in mapping systems.
  - Developed a parallelized software cache layer (C++) to accelerate updates and queries in the mapping system.
  - Implemented the cached-mapping system on a ROS-based MAV autonomous navigation simulation platform and achieved up to 45% saving in end-to-end workflow runtime, raising flight speed by 26%.
- **Interactive Research Agents for Internet Incident Investigation** — May 2023 – Nov 2023
  *Mentor: Zaoxing(Alan) Liu* — *Froot Lab, UMD*
  - Developed an LLM-based agent to simulate experienced researchers and automate the investigation process, addressing the inefficiencies of traditional manual and time-consuming Internet incident investigations.
  - Built an agent using Auto-GPT and GPT-4, equipped with autonomous information retrieval, knowledge memory, and self-learning capabilities. Tested it on challenging scenarios such as the impact of hypothetical solar storms on networks.
  - Achieved 87.5% consistency in insights compared to human experts, effectively automating complex Internet incident analysis.
  - Co-first authored paper appeared at **HotNet 2023**
- **Advanced Persistent Threat (APT) Detection and Analysis** — Feb 2022 – Present
  *Mentor: Zaoxing(Alan) Liu, Tuo Zhao* — *Red Hat & Boston University & Georgia Tech*
  - Advanced Persistent Threats (APT) are stealthy and prolonged cyber attacks that target critical information or systems. Traditional detection methods struggle to identify new attack patterns and often lack precision due to data scarcity.
  - Built a flexible data pipeline with Python and Pandas. Developed an APT detection system using a Multi-Classification-Head Transformer with a pre-training and fine-tuning workflow for efficient knowledge transfer.
  - Achieved 10%-29% improvement in fine-grained detection accuracy and 25%-50% reduction in training data needs across 8 real-world APT datasets. Enhanced model transparency, aiding in better post-attack investigation.
  - Submitted to **IEEE S&P 2025**

## Course Projects

- **Concurrency Control Schemes for Database Systems** Jan 2023 – May 2024
  - Implemented and evaluated six concurrency control schemes, including Two-Phase Locking (2PL), Optimistic Concurrency Control (OCC), and Multi-Version Concurrency Control (MVCC). Developed two versions of 2PL with exclusive and shared locks, serial and parallel validation versions of OCC, and a simplified MVCC with Serializable Snapshot Isolation (SSI).
  - Developed and integrated the concurrency control schemes into a main-memory key-value store using C++ and thread management techniques. Built a prototype transaction processing framework with a custom lock manager and multi-threaded execution support.
  - Conducted performance benchmarking with CMake and CTest to assess throughput and latency across varying transaction lengths and contention levels.
- **eBPF Modularity Project** Sep 2022 – Dec 2022
  - Collaborated with professors from Brown University and IBM engineers to advance research on building a comprehensive and reusable eBPF module library.
  - Leveraged OPENED tool to analyze and decompose eBPF programs from open-source projects, extracting reusable modules for improved modularity and maintainability.
  - Designed a framework to transform extracted modules into a format compatible with Bumblebee tools for generating OCI images. Integrated "glue logic" for seamless module compatibility with L3AF and Polycube.
- **Basic Unix-like Operating System** Jan 2022 – May 2022
  - Implemented a custom shell capable of executing commands via system calls, redirecting stdin/stdout, handling multiple command pipelines, and supporting background execution with "&".
  - Developed a subset of the POSIX threads API in user mode, enabling multi-threaded execution with round-robin scheduling for effective resource management.
  - Designed and implemented a copy-on-write (COW) thread-local storage (TLS) mechanism to enable data sharing between threads while ensuring isolation of changes, improving memory management at the thread level.

## PUBLICATION & WORKS

**[1]** Yajie Zhou*, **Nengneng Yu*** , Zaoxing Liu, "Towards Interactive Simulacra of Internet Investigation by Human Researchers", Hot Topics in Networks (HotNets), 2023
**[2]** Yajie Zhou, **Nengneng Yu**, Simiao Zuo, Yue Yu, Haoming Yi, Chao Zhang, Tuo Zhao, Zaoxing Liu, "Fine-Grained, Adaptive Advanced Persistent Threats Detection with SENTINEL", Under Submission

## Services

- **University of Maryland College Park Department of Computer Science** Aug 2023 - Present
  *Graduate Teaching Assistant* *CMSC498B Cloud Computing*
- **University of Maryland College Park Department of Computer Science** Aug 2023 - Present
  *Graduate Teaching Assistant* *CMSC414 Network and Security*
- **University of Maryland College Park Department of Computer Science** Aug 2023 - Present
  *Graduate Teaching Assistant* *CMSC250 Discrete Structure*
- **Boston University College of Engineering** Sep 2022 - Dec 2022
  *Teaching Assistant* *EC440 Operating System*
- **Boston University College of Engineering** Jan 2022 - May 2022
  *Teaching Assistant* *EC414 Machine Learning*