KHIEM TON

Curriculum Vitae

Newark, New Jersey — kt477@njit.edu — Website — Linkedin

EDUCATION

Ph.D. in *Data Science* (in progress)
Expected graduation: May 2029
New Jersey Institute of Technology
Data Science Department

B.S. in *Software Engineering*
Year of completion: 2024
FPT University, Danang, Vietnam
Software Engineering Department

RESEARCH EXPERIENCE

AI4SocialGood Lab, NJIT, Newark, NJ — Ph.D. Student | August 2024 – Present
Advisor: Prof. Hai Phan
Research topics: LLMs, reinforcement learning, software security and privacy

Vulnerability Fixing with Functionality Preserving via Reinforcement Learning (in progress)

- Curated a verifiable dataset of vulnerable code and test cases.
- Developed a reinforcement learning based training pipeline that increased vulnerability patching rate from 44% to 76% while improving functionality preserving rate from 75% to 80%.
- Implemented the approach using verl, and trained the model using multinode distributed training on an HPC cluster, achieving 4× speedup.

NOIR: Privacy-Respecting LLM System

- Designed and evaluated prompt and response reconstruction attacks on LLMs, demonstrating up to 90% success rate.
- Co-developed NOIR, a split-learning-based LLM system leveraging local differential privacy to protect user data while preserving model utility and mitigating the attacks.

SGCode: Secure Code Generation System

- Developed SGCode, an LLM-based system for secure code generation with integrated static security analysis.
- Conducted systematic experiments quantifying functionality-security trade-offs, providing insights into the limits of LLM-based secure code generation.

Qualcomm, San Diego, CA — Research Collaborator | January 2026 – Present
Research topics: Large Language Models, Agentic Systems

Mobile Agentic Systems (in progress)

- Co-designed a mobile agentic system that leverages Language Models to predict user needs.
- Developed a mobile app and server to collect mobile usage data and feedback for training and evaluating the system.

AT&T, Bedminster, NJ — Research Collaborator | November 2025 – Present
Research topics: Reinforcement Learning, Scheduling, Resource Management, Large Language Models.

Cellular Network Maintenance Scheduling (in progress)

- Designed an offline reinforcement learning based scheduling system for cellular network maintenance, optimizing for both operational efficiency and user experience.
- Processed and analyzed large-scale network data to inform model training and evaluation.

PUBLICATIONS

Peer-Reviewed

[1] K. Nguyen*, **Ton, Khiem***, N. Phan, et al., "Noir: Privacy-preserving generation of code with open-source llms," in *Proceedings of the 35th USENIX Security Symposium*, To appear, 2026.

[2] **Ton, Khiem***, N. Nguyen*, M. Nazzal, et al., "Sgcode: A flexible prompt-optimizing system for secure generation of code," in *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '24, Association for Computing Machinery, 2024. DOI: 10.1145/3658644.3691367

Patents

[1] N. Phan, K. Nguyen, **Ton, Khiem**, et al., *System and method for private generation of code*, International Patent Application, No. PCT/US25/042391, 2025.

INDUSTRY EXPERIENCE

OppyAI Inc., Newark, NJ — CTO | November 2024 – November 2025

- Developed NOIR web application and inference engine, enabling users to privately query LLMs and observe security attack attempts in real time.
- Reduced NOIR's inference latency by 3× by extending vllm with split-learning CUDA graph and high-throughput gRPC packet transfer.
- Tech: vllm, CUDA, gRPC, Python, FastAPI, ReactJS

FPT Software, Danang, Vietnam — Data Engineer | November 2023 – June 2024

- Cut risk assessment data collection and processing costs and time by 50% by engineering an end-to-end data pipeline with Microsoft PowerApps for a banking client.
- Designed a Tableau dashboard to visualize risk assessment results, enabling stakeholders to pinpoint process bottlenecks.
- Developed an LLM-based QA system that automatically responds to employee queries about the risk assessment process, reducing manual support and handover load.
- Tech: Oracle SQL, Microsoft PowerApps, Tableau, Python, Huggingface Transformers

Productminds, Denmark — AI Engineer | September 2023 – April 2024

- Built an NLP pipeline to automatically crawl Danish news articles and extract named entities, serving as the core data source for trend analysis products.
- Developed an LLM-based machine translation system for Icelandic-to-English, enabling coverage of a low-resource language.
- Tech: Supabase, Python, OpenAI API, Huggingface Transformers, FastAPI, spaCy

TEACHING EXPERIENCE

- Teaching Assistant, Artificial Intelligence (CS670), NJIT — Spring 2026
- Lecturer, Social Network Analysis (IS333), NJIT — Fall 2025
- Teaching Assistant, Artificial Intelligence (CS670), NJIT — Fall 2025
- Teaching Assistant, Operating Systems Design (CS630), NJIT — Summer 2025
- Teaching Assistant, Big Data (DS644), NJIT — Spring 2025

HONORS, AWARDS

- FPT Education Hackathon Runner-up — Fall 2022
- Golden Toad for Most Excellent Student — Summer 2022
- FPT University 100% Scholarship — 2020
- Odon Vallet Scholarship — 2019

SERVICE

- Volunteer: International Conference on Robotics and Automation — May 2025

SKILLS

- Programming: Python, C/C++
- Tools: Git, Linux, Docker, HPC
- ML Tools: Pytorch, Huggingface Transformers, verl, vllm