Minh A. Nguyen

nguyenanhminh2804@gmail.com | github.com/minhna1112

Aspiring PhD student with four years of experience in applying AI research into downstream applications. Investigating the association between AI and Secure Software Development

---

EDUCATION

PhD Student in *Computer Science*
University College Dublin, Dublin, Ireland — 2024 — Now

Bachelor of Engineering in Electronics and Telecommunication
Hanoi University of Science and Technology**,** Hanoi, Vietnam — 2017 — 2022

European Union's Erasmus+ Exchange Student
Technical University of Munich, Munich, Germany — 2019 — 2020

PROFESSIONAL EXPERIENCE

Researcher at LERO, Dublin, Ireland — Sep 2024 — Now

Lero is the top Irish research center and industry leader in software research and development. I work within the UCD division of LERO, led by Dr. Liliana Pasquale, studying AI and Software Security topics

- **Assisting Code Review with LLMs:** Create datasets for evaluating Large Language Models (LLMs) on subtasks of Code Review, including vulnerability detection and fixing.
- **Secure Code Generation:** Train LLMs to make sure they do not introduce vulnerabilities when generating or refactoring code.
- **Co-evolution of AI and IDEs for Accountable Coding Assistants:** Investigate methods to distinguish between human and AI-generated code. Develop IDE tools to track ownership and modification of AI-generated code.
- **Secure Supply Chain of LLMs:**
  - Perform a security audit the supply chain of open-source LLMs
  - Examining access control mechanism of cloud-based LLMs resources
  - Investigate potential risks in a simulated decentralized LLM inference network

AI Engineer at FPT Software AI Center, Hanoi, Vietnam — Aug 2022 — Sep 2024

FSoft AIC is a national leading AI research lab and AI solutions provider. I work within the AI4Code team that spans both academic research and product development, led by Dr. Nghi Bui.

- **CodeVista – AI Coding Assistant in VSCode & IntelliJ**
  The product performs various coding tasks such as explaining, debugging, and fixing SonarLint issues by activating predefined actions or prompting, provides coding Q&A with Google Search cross-checking, allows users to index code files for retrieval by identifiers and semantic search.
  - Implemented the Python AI server with multiple LLM agents using LangChain

    - Built the inference server for self-hosted Code LLaMA2 34B model with inference serving through the HuggingFace TGI framework
    - Built the client interfaces for Azure OpenAI GPT and Google Code Chat Bison
    - Developed the context architecture for secure storage and retrieval of code snippets and embeddings by combining Zoekt code regex search and Milvus vector database
    - Created pull requests for bug fixes to the public LangChain GitHub repo

- **Docify AI – AI Code Documentation in VSCode, IntelliJ, Eclipse & CLI**
  The product reads a snippet, a file, or a repository of code in 10 programming languages to generate docstrings and comments by using our proprietary CodeT5 CodeSum 700M model.
    - Helped train the in-house model based on Salesforce CodeT5 using The Vault dataset and DistillNet architecture; quantized the model and converted the checkpoint from PyTorch to TensorRT for faster inference; hosted the model on NVIDIA Triton Server; integrated Meta's opensource NLLB model to support translating code comments and docstrings among 13 human languages
    - Implemented the client service using Python FastAPI that serves 3K+ weekly requests

Research Assistant at IVSR Lab, Hanoi, Vietnam Nov 2020 — Jul 2022
IVSR is a lab for autonomous drones. I led the team that developed AI tools for navigation control.

- **Project 1: Package delivery drones –** €19K project with Technical University of Graz, Austria.
    - Built the prototype
- **Project 2: Intelligent Vision System for Robotics –** Project with College of Information Sciences, Ritsumeikan University, Japan
    - Facilitate low-cost, learning-based solutions for robotics vision
- **Project 3: Autonomous Drone Racing**
    - Provided proof-of-concept demonstrations of agile flying robots in both simulated and lab environments. Resulted in papers names “ORB-Net: End-to-end Planning Using Feature-based Imitation Learning for Autonomous Drone Racing”

PUBLICATIONS

1. Khanh Nghiem, **Anh Minh Nguyen**, and Nghi DQ Bui. 2023. “Envisioning the Next-Generation AI Coding Assistants: Insights & Proposals” IDE Workshop at ICSE 2024. arxiv:2403.14592
- Presented experiences on applying AI4SE research and LLM innovations in building two extensions in VSCode and IntelliJ: Docify AI and CodeVista.
- Propose open questions and challenges that academia and industry should tackle to materialize the next-generation AI coding assistants
- Won the “Best Paper” Award

2. Dung Nguyen Manh, Nam Le Hai, Anh TV Dau, **Anh Minh Nguyen**, Khanh Nghiem, Jin Guo, and Nghi DQ Bui. "The Vault: A Comprehensive Multilingual Dataset for Advancing Code Understanding and Generation." Proceedings of EMNLP 2023 arXiv:2305.06156.

- Presented a dataset of 43M high-quality code-text pairs in 10 programming languages for training LLMs extracted and refined from The Stack (3TB of source code) alongside the toolsets and methodologies for creating the data
- Built the static code analysis tool for code-text pair extraction (source code); helped design, implement, and manage the infrastructure for the data pipeline; informed iterations of data cleaning and toolset debugging from results observed in downstream applications

3. Huy Xuan Pham, Micha Heiß, Dung Tran, **Anh Minh Nguyen**, Anh Quang Nguyen and Erdal Kayacan "ORB-Net: End-to-end Planning Using Feature-based Imitation Learning for Autonomous Drone Racing." Proceedings of ISR 2023

- Developed an innovative neural network architecture that enables real-time motion controls over high-speed drones that requires only single-camera RGB image inputs and a lightweight Jetson Nano processor
- Built the continuous imitation training pipeline in PyTorch by automating data collection from rule-based navigation decisions and evaluating AI-based decisions in each training epoch
- Enhanced AI model's real-world robustness by introducing randomness to the simulated training environments for drone navigation

4. **Anh Minh Nguyen,** Dinh Tuan Tran, Dung Duc Tran, et al., "MonoIS3DLoc: Simulation to Reality Learning Based Monocular Instance Segmentation to 3D Objects Localization From Aerial View." IEEE Access 2023 ACCESS.2023.3288027.

- Introduced a new neural architecture in Tensorflow that infers positions of recognized entities in 3D from 2D RGB images, where inferences are performed on the processed semantic masks instead of real world images, which expedited training data synthesis since semantic masks are easily algorithmically created

TEACHING

- *Prompt Engineering for Coding Tasks and Beyond.* Curriculum in development for recurring training at FPT Software. January 2024 (expected delivery)
- *Spatial Intelligence for Flying Robots*. Recurring 4-week training program at IVSR Lab, Hanoi University of Science and Technology. October – November 2022
- Teaching Assistant at Hanoi University of Science and Technology. August 2018 - May 2022
    - *Linear Algebra, Fundamental of Network Engineering, Applied Software Engineering*

HONORS AND AWARDS

PhD Merit-based Scholarship, School of Computer Science, UCD Sep 2024 - Aug 2028

Best Paper Award, 1st IEEE/ACM Workshop on IDE, ICSE 2024 April 2024

Excellent Employee, FPT Software AI Center Jun 2023

Third Prize, Hanoi LLM Hackathon 2023, FPT Software Oct 2023

Erasmus+ KA1 Grant for Exchange Student, European Commision Jul 2019 — Jul 2020