NGUYEN HAI DANG

Birthday: 30/07/2004   **Phone**: 0376762541   **Mail**: haidang.nguyen.cs@outlook.com   **Address**: Thanh My Tay Ward, Ho Chi Minh City

WORKING EXPERIENCE

Research Intern | InSecLab – University of Information Technology (VNU-HCM)
Jul 2025 – Dec 2025
Objective:
Research and develop the *Vulnex Framework*, a Causal XAI approach for interpreting Transformer-based vulnerability detection models such as CodeBERT and GraphCodeBERT.
Key Contributions:
- Implemented post-hoc explainability methods including LIME, SHAP, and Integrated Gradients (IG) to obtain token-level attribution signals from black-box models.
- Designed evaluation scenarios (e.g., Input Truncation, Invisible Vulnerabilities) to assess model robustness and explanation faithfulness.
- Conducted quantitative analysis using MIOU, MSP, and MSR metrics to study the relationship between model attention and ground-truth vulnerabilities.

Achievements:
- Co-authored a Q1 research paper (under review at Elsevier):
- *"From Explanation to Localization: A Causal XAI Framework for Understanding and Improving Transformer-Based Vulnerability Detection"*

Technologies: Python, PyTorch, Git

Software Developer(AI/Backend Focus) | MOONSHOOT (Remote)
Apr 2025 – Jul 2025
- Built RESTful APIs using FastAPI for user authentication (Magic Link, OTP), store management, and chatbot services.
- Integrated AI tools including OpenRouter (chatbot) and Meilisearch (semantic product search) into backend systems.
- Developed rule-based data collection systems for vendors such as Hasaki and Guardian using Selenium and BeautifulSoup.
- Collaborated within an Agile/Scrum team using Jira.
- **Technologies:** Python, FastAPI, MySQL, Selenium, BeautifulSoup, Git, Jira

PROJECTS

Project: Flight Delay Prediction System (GNN & Streaming Data Pipeline)
Repository: https://github.com/haidangnguyen-cs/flight-delay-gnn-system
Objective:
Build a flight delay prediction system on streaming data using Graph Neural Networks (GNN) to learn relational patterns in the aviation network.
Key Contributions:
- Transformed tabular flight data into graph-structured data (airports as nodes, flight routes as edges) with feature scaling and label encoding.
- Applied GATv2 (Graph Attention Networks) to model topological dependencies between airports and flight features.
- Addressed severe class imbalance using a weighted loss strategy to improve performance on delayed flights.

System & Data Engineering:
- Designed a near real-time streaming pipeline with Apache Kafka as the backbone, separating ingestion and inference modules.
- Integrated PyTorch models into a distributed Spark pipeline using PySpark UDFs for scalable inference on large data streams.
- Archived raw streaming data in Cassandra for historical analysis and visualized prediction results with a Streamlit dashboard.

Results:
- Achieved an F1-score of 0.46 on the delayed class using a highly imbalanced real-world dataset.
- Delivered a fully automated pipeline from raw data ingestion to dashboard visualization.

Technologies:
- Python, PyTorch Geometric, Apache Kafka, Apache Spark (Structured Streaming, UDF), Cassandra, Docker, Streamlit.

EDUCATION

Vietnam Aviation Academy (VAA) Bachelor of Information Technology | 2022 – 2025
- **GPA:** 3.08/4.0
- **Certificate:** TOEIC 475 (Able to read technical documentation).

TECHNICAL SKILLS

TECHNICAL SKILLS
- **Languages & Core:** Python (Strong focus), SQL (Basic query).
- **Data Science & AI Libraries:** PyTorch, PyTorch Geometric, Pandas, NumPy, Scikit-learn.
- **Key Concepts:** Feature Engineering, Model Evaluation, Supervised/Unsupervised Learning, Explainable AI (XAI), Graph Neural Networks (GNN).
- **Backend & MLOps:** FastAPI, Apache Kafka, Apache Spark (Structured Streaming), Cassandra, Docker, Git, Linux.