* Phone: +84 865 589 695

## Nguyen Mai Thanh

* LinkedIn: linkedin.com/in/thanh-mai-57480531b
  + Email: [mikeethanh04@gmail.com](mailto:mikeethanh04@gmail.com) • Github: github.com/mikeethanh

# SUMMARY

From the beginning of my university journey, I initially pursued a career as a Backend Engineer. However, in my second year, I had the opportunity to participate in a scientific research program where my professor assigned me to develop a deep learning model using Deep Neural Networks (DNNs) for automated attendance tracking in online classes via Zoom, named "Supervision-for-Online-Learning." This project required me to explore deep into DNNs, and I need to start from the basics things like: Python, NumPy, OpenCV, Machine Learning, Neural Networks, and more. My passion lies in AI and Data Science, and I am currently enrolled in a Full-Stack Data Science **LLM** course, actively researching **AI Agents** and Large Language Models (LLMs).



# EDUCATION

**UNIVERSITY OF TRANSFORT AND COMMUNICATIONS (UTC) Hanoi, VietNam**

*Bachelor of Science in Information Technology - International Training Program*

Graduation Date: 6/2026

**GPA**: 3.34

# TECHNICAL SKILLS

* **Program laguages:** Python, SQL
* **Libraries:** NumPy, Pandas, Matplotlib, OpenCV
* **Frameworks:** PyTorch, TensorFlow
* **Machine Learning Algorithms:** Linear Regression, Multiple Linear Regression, Logistic Regression, Decision Trees, Clustering (KNN, K-Means, Hierarchical Clustering)
* **Deep Learning:** Neural Networks, CNN, RNN
* **LLMs:** Currently exploring BERT, GPT-2, and BART
* **Data Engineering:** Familiar with Apache Spark, Kafka, Flink, Trino, Hive, MinIO, Debezium, Airflow

## Others:

* + Strong teamwork and collaboration skills, high sense of responsibility, effective communication and presentation skills.
  + Time management and organizational skills, ensuring timely completion of tasks and efficient workflow.
  + Continuous learning mindset, keeping up-to-date with the latest advancements in AI models and technologies.

# PROJECTS

## Supervision-for-Online-Learning (Deep Learning Research Project)

* **Link project**: https://github.com/mikeethanh/Supervision-for-Online-Learning.git
* Developed an AI-powered automated attendance system for online learning environments, leveraging deep learning techniques to ensure accurate identity verification and engagement tracking.
* Conducted in-depth research on Neural Networks, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to extract facial features from video streams and match them with stored student ID images.
* Utilized the face\_recognition library to encode and compare facial data stored in Firebase, ensuring efficient and real-time facial verification.
* Implemented MobileNetV2 with a dataset of 5000 images, achieving a 92% accuracy rate in detecting impersonation attempts (e.g., using student ID photos, displaying static images on camera, or deepfake manipulation).
* Enhanced the model by integrating key architectural improvements, such as Dropout Layers for regularization, Convolutional Layers (Conv) for hierarchical feature extraction, and Batch Normalization for stable training.
* Optimized loss functions using techniques like Cross-Entropy Loss and Adam optimizer to improve convergence and model generalization across varying lighting conditions and facial expressions.

## Other Personal Data Engineering Projects:

* **Build CDC system:**
* **Link project**: https://github.com/mikeethanh/continuous-cdc-postgresql-kafka.git

**+** Set up a lab environment using Confluent Kafka, Debezium, and PostgreSQL with Docker Compose.

+ Establish a continuous data writing flow into PostgreSQL.

+ Implement a CDC system to capture changes from a PostgreSQL database and publish these change events to Kafka.

## FlinkStream: Real-Time Data Processing:

* **Link project**: https://github.com/mikeethanh/streaming-processing-with-flink.git

+ Develop a streaming text analyzer application.

+ Create a streaming pipeline to process signals from IoT devices using both DataStream and Table APIs.

## ML Orchestrator with Airflow:

* **Link project**: https://github.com/mikeethanh/pipeline-orchestration-with-airflow.git

+ Deploy the Airflow platform locally using Docker Compose.

+ Gain proficiency with Airflow operators through practical exercises.

+ Develop an end-to-end machine learning pipeline using Airflow, starting from data preparation to model training

# CERTIFICATIONS & AWARDS

## TOEIC (4 Skills)

* + - Listening & Reading: 780 (Valid until Aug 23, 2025)
    - Speaking & Writing: 270 (Valid until Dec 17, 2026)

## Third Prize - Scientific Research Competition

* + - Awarded for the research project "Supervision for Online Learning", which applied Deep Learning models for automated attendance tracking and student behavior analysis**.**

## Samsung Algorithm and Application Program Certificate(SamSung)

* + - Achieved A+ in Samsung's competitive certification program, demonstrating strong problem-solving and algorithmic skills**.**

## Completed a module on MLOps from Full Stack Data Science (fullstackdatascience.com)

* + **Coursera Certifications:**
    - Supervised Machine Learning: Regression and Classification (DeepLearning.AI – Coursera)
    - Advanced Learning Algorithms (DeepLearning.AI – Coursera)
    - Neural Networks and Deep Learning(DeepLearning.AI – Coursera)
    - Convolutional Neural Networks(DeepLearning.AI – Coursera)
    - Python for Data Science, AI & Development (IBM – Coursera)

*All certifications and awards are verified with supporting documents and available upon request.*