

Nhu Ngoc Hoang

Tel: +32 (0)49 298 4791 | [linkedin.com/in/ngocngoahong](https://www.linkedin.com/in/ngocngoahong) | github.com/nhungoc1508 | ngoc.hoang@nyu.edu

EDUCATION

Erasmus Mundus Joint Master in Big Data Management and Analytics

09/2024 – Present
Belgium, Spain, Italy

- Consortium: Université Libre de Bruxelles (Brussels, Belgium; GPA: 18.25/20), Universitat Politècnica de Catalunya (Barcelona, Spain; GPA: 8.99/10), Università degli Studi di Padova (Padua, Italy).
- Awarded a full-ride scholarship by the European Commission.
- Relevant coursework: Data Warehouses, Advanced Databases, Database Systems Architecture, Big Data Management, Semantic Data Management, Deep Learning, Statistical Learning.

Bachelor of Science in Computer Science

08/2019 – 05/2023

New York University Abu Dhabi

Abu Dhabi, United Arab Emirates

- Major in Computer Science, minors in Applied Mathematics and Interactive Media.
- GPA: 3.906/4.0; recipient of NYU Founders' Day Award (awarded to top-ranking graduating seniors).
- Bachelor thesis: *Anomaly Detection Using AutoEncoders: the Advanced Persistent Threats Case*.

PROFESSIONAL & RESEARCH EXPERIENCES

Graduate Research Assistant

07/2025 – 08/2025

Computer and Decision Engineering Department, ULB

Brussels, Belgium

- Co-authored MobilityDuck, a C++ extension for DuckDB to manage spatiotemporal data types based on MobilityDB, leveraging vectorized execution for high-performance analytics.
- Implemented core spatiotemporal data types and operations to enable native DuckDB support for moving object data analysis, contributing to the open-source ecosystem of high-performance analytical databases.

Software Engineer

10/2023 – 08/2024

LG Electronics Development Vietnam

Da Nang City, Vietnam

- Developed a cross-domain module for over-the-air (OTA) firmware updates within in-vehicle infotainment (IVI) systems, synchronizing critical states across a dual-OS environment (Android Automotive OS and QNX).
- Implemented the core logic in C++ under strict MISRA compliance, ensuring secure and reliable solution delivery for global automotive clients.
- Designed modular architectures and APIs to manage complex package lifecycles, integrating third-party solutions to handle data verification, execution, and status telemetry.
- Automated validation pipelines by building custom tools for test package generation, enhancing the efficiency and accuracy of the quality assurance cycle.

Research Assistant

05/2023 – 07/2023

Center for Quantum and Topological Systems, NYU Abu Dhabi

Abu Dhabi, United Arab Emirates

- Conducted preliminary research and experiments in formalizing braid groups as automorphisms of free groups using the dependently typed functional programming language Agda.

Undergraduate Researcher (Bachelor Thesis)

02/2022 – 05/2023

Department of Computer Science, NYU Abu Dhabi

Abu Dhabi, United Arab Emirates

- Developed two AutoEncoders-based models to detect advanced persistent threats (APTs) in system-level provenance data.
- Tackled the challenge of extreme data imbalance (0.004% anomaly rate) within the DARPA Transparent Computing dataset, achieving state-of-the-art ranking performance (nDCG > 0.9) by training models to learn robust representations of normal system behaviors.

- Co-authored a research paper published in Future Generation Computer Systems (Elsevier), validating the methodology across varying attack scenarios on Linux, Windows, and BSD environments.

Research Assistant

06/2022 – 08/2022

Department of Civil and Urban Engineering, NYU Tandon

New York City, United States

- Conducted equity analysis on synthetic population datasets, utilizing Google BigQuery to query large-scale census records and validate demographic consistency against federal baselines, visualizing findings with various maps and data dashboards.
- Developed computational optimizations for EV charging assignment models, assessing the viability of genetic algorithms and Cython integration to accelerate solving times for stochastic user equilibrium problems by 90%.

Research Assistant

05/2021 – 07/2021

Center for Global Sea Level Change, NYU Abu Dhabi

Abu Dhabi, United Arab Emirates

- Developed and evaluated a Bayesian Network model to predict probabilistic coastal landscape responses (inundation vs. dynamic adaptation) given geomorphological and geologic settings under varying sea-level rise scenarios.
- Engineered a geospatial data processing pipeline using GeoPandas and Rasterio, integrating heterogeneous datasets (satellite elevation, land cover shapefiles, vertical land movement) into a unified training set.
- Optimized the model inference engine by implementing a caching mechanism for repetitive observation tuples, reducing prediction runtime per scenario from 6 hours to 20 minutes.

PUBLICATIONS

Nhu Ngoc Hoang, Ngoc Hoa Pham, Viet Phuong Hoang, Esteban Zimányi, "MobilityDuck: Mobility Data Management with DuckDB", preprint. DOI: [10.48550/arXiv.2510.07963](https://doi.org/10.48550/arXiv.2510.07963).

Sidahmed Benabderrahmane, **Ngoc Hoang**, Petko Valtchev, James Cheney, Talal Rahwan, "Hack me if you can: Aggregating autoencoders for countering persistent access threats within highly imbalanced data", *Future Generation Computer Systems*, 2024. DOI: [10.1016/j.future.2024.06.050](https://doi.org/10.1016/j.future.2024.06.050).

Joseph Y J Chow, Xiyuan Ren, and **Ngoc Hoang**. "NY Statewide Behavioral Equity Impact Decision Support Tool with Replica", *Connected Communities for Smart Mobility Toward Accessible and Resilient Transportation for Equitably Reducing Congestion* (C2SMARTER) Tier-1 University Transportation Center (UTC), 1 July 2023, rosap.ntl.bts.gov/view/dot/72373.

Ngoc Nhu Hoang. "A Vietnamese Named Entity Recognition System for COVID-19 Articles". *2022 IEEE MIT Undergraduate Research Technology Conference* (URTC), IEEE, 30 September 2022. DOI: [10.1109/URTC56832.2022.10002170](https://doi.org/10.1109/URTC56832.2022.10002170).

CONFERENCE PRESENTATION

"A Vietnamese Named Entity Recognition System for COVID-19 Articles", *2022 IEEE MIT Undergraduate Research Technology Conference*, Cambridge, Massachusetts, United States. October 2022.

SKILLS

Programming languages: Python, C, C++, JavaScript, Swift

Data science: Pandas, NumPy, scikit-learn, TensorFlow, Matplotlib, seaborn,

Database technologies: PostgreSQL, DuckDB, Apache Solr, Neo4J, MongoDB, Apache Spark, Apache Airflow

Web development: HTML, CSS, Flask, React

Languages: Vietnamese (native), English (fluent), French (beginner)