Minh-Anh To

minhanh.to@duke.edu ● (919) 475-8485 ● LinkedIn: minhanhto ● Github: minhanhto09

EDUCATION

Duke University - The Graduate School *Aug 2023 – Exp. May 2025*
M.S. - Statistical Science *- GPA: 3.91/4.0* *Durham, NC*

- Full Tuition Scholarship.
- Coursework: Predictive Modeling, Deep Learning, Causal Inference, Statistical Computation, R Programming, SQL, Bayesian Statistics, Theory of Inference, Hierarchical Models, Applied Machine Learning.

Hanoi National University of Education *Aug 2014 – May 2018*
B.S. - Mathematics (Honors) *- GPA: 3.65/4.0 (Top 5%)* *Hanoi, Vietnam*
National Mathematics Scholarship (Top 1% Applicants); Vice President of Student Council; Dean's List.

SKILLS

Programming: Python (NumPy, pandas, scikit-learn, PyTorch, TensorFlow), R (dplyr, ggplot2, tidy), SQL, Git, REST APIs.
AI/ML: Generative AI (LangChain, OpenAI APIs, VectorDBs), Deep Learning (CNNs, GANs), Machine Learning (XGBoost, Random Forest, Logistic Regression, SVM).
Statistics: Hypothesis Testing, ANOVA, Causal Inference, Forecasting, Bayesian Modeling.
Tools & Platforms: Jupyter, Cloud Services (AWS, GCP).

WORK EXPERIENCE

ChessTutorAI *Aug 2024 – Present*
AI Developer *Remote, NC*

- Developed a **Chess Tutor Chatbot** by utilizing **Generative AI** and collaborating with a **chess grand master** to provide learners with personalized gameplay feedback and insights.
- Cleaned and mined large datasets of post-game chess analyses (**>100K entries**) using **Pandas** and **NLP packages**, building a structured database for training.
- Built a **Retrieval-Augmented Generation (RAG)** pipeline using **LangChain** to retrieve chess concepts from an expert-curated dataset, increasing model **retrieval accuracy** by **28%**.
- Optimized Large Language Model (LLM) responses through **few-shot learning** and **chain-of-thought prompting**, minimizing hallucinations; validated performance with semantic similarity metrics, achieving over **99% accuracy**.
- Documented and maintained a comprehensive GitHub codebase, including unit testing and automation scripts, reducing **pipeline debugging** and **deployment time** by **40%**.

Duke University School of Medicine *May 2024 – Aug 2024*
Machine Learning Intern *Durham, NC*

- Transformed **10 years** of unstructured antibody production data into a structured database of **5,000 mRNA sequences** and designed an **end-to-end machine-learning pipeline** to predict antibody yields.
- Engineered **35 mRNA structural features** by devising statistical formulas with molecular biologists and coding custom functions in Python (ViennaRNA, Pandas, NumPy), boosting model performance by **75%**.
- Fine-tuned and cross-validated tree-based models with scikit-learn, achieving **90% accuracy** and reducing runtime by **48%**; identified **5 key mRNA properties** using feature importance analysis to enhance yields.
- Authored a detailed report on the model and its **large-scale biotech applications**, leading to selection from 50 candidates to present findings to **200 scientists** at the Duke CFAR Retreat 2024.

Vietnam Academy of Science and Technology *Jun 2022 – Aug 2023*
Assistant Researcher *Hanoi, Vietnam*

- Demonstrated the practicality of the **bootstrap method** for goodness-of-fit testing using estimator properties, increasing adoption in data science; presented findings to **25 researchers** and secured a significant grant from VinIF Vietnam.
- Led a weekly Bayesian Statistics seminar for **30+ students**, featuring live **R** coding on topics such as **Bayesian Regression**, **Gibbs Sampling**, and **Hierarchical Modeling**.