

HIEU NGUYEN, PhD

hieutrongnguyen95@gmail.com • New York, New York, USA

AI Engineer/Data Scientist with 8+ years of experience in AI/ML ('Classical' ML and GenAI/NLP), biostatistics, RWE/RWD, signal processing, and causal inference. Experienced in handling/analyzing various large-scale structured and unstructured data types (EHR/Clinical trials/Claims/Wearables/Imaging), deploying end-to-end ML/LLM production systems at scale, and publishing papers. Experienced in leading projects and working cross-functionally.

EXPERIENCE

Optum | Lead AI/ML Engineer, Data Scientist | New York, NY Feb 2023 – Present

- Led, scoped, planned, implemented, and deployed a clinician-facing patient history summarization product utilizing multi-modal and GenAI/NLP methods to improve chronic care, scaled to 5 businesses, estimated to save the business \$47M (patented).
- Spearheaded the development of a scalable real-time topic extraction and summarization pipeline for audio call center data. Worked with data engineering, business, legal, compliance, and execs. Formulate metrics and build modules to evaluate ML algorithms.
- Implemented RAG optimization techniques to enhance retrieval efficiency and broaden coverage to 3 medical code sets.
- Designed evaluation, validation, and self-corrective feedback collection pipelines for GenAI models
- Built the pipeline for providing explainable risk scores for medical claims that helped increase fraud detection coverage by 300 times.
- Ideated and built ML models for blood glucose prediction and treatment recommendation.
- Provided expert guidance on survival analysis, RAG, and explainable AI for 3 working groups. Mentored 3 FTEs and 2 interns.
- *Models/skills used: Agentic design, (proprietary and open-source) LLMs, VLMs, CNNs, LSTMs, Transformers, RAG, OCR, Mixture of Experts, NER, fine-tuning, prompt engineering, validation, responsible AI, data drift, model monitoring, SQL, MLOps, CI/CD.*
- *Tools/frameworks used: LangGraph, LangChain, LangSmith, RAGAS, Redis Cache, Airflow, MLFlow, Tableau, Pytorch*

JOHNS HOPKINS HOSPITAL | Machine Learning Researcher – Precision Care Medicine | Baltimore, MD Sep 2018 – Jan 2023

- Managed 8 research teams (40 ppl total) on every step of ML-based disease prediction cycle (design, build, ship, iterate).
- Developed personalized real-time early warning models for 9 critical illnesses and injuries in the ICU, leveraging ML methods on various types of high-dimensional biomedical data (e.g., biosignals, EHR/EMR data, images, clinical trials, observational studies), led to the discovery of new biomarkers and insights for better understanding of disease mechanisms to devise life-saving interventions.
- Models won Investigation Awards at RESS'19 & SCCM'20 – among the most important cardiac arrest meetings worldwide.
- *Models/skills used: CNNs, LSTMs, U-net, XGBoost, Random Forest, Logistic Regression, feature engineering, feature selection, transfer learning, explainable AI, hyperparameter tuning, model evaluation, ML pipeline, Keras, PyTorch, regressions, A/B testing, SQL.*

APPLE | Data Scientist PhD Co-op – Health Tech | Cupertino, CA Jun 2021 – Dec 2021

- Led a team of 4 to the finale of Shark-Tank-like internal competition (2% acceptance rate). Ideated, pitched, and built proof-of-concept to the Apple execs on an eye-protection feature that later got shipped as iOS 17's Screen Distance feature.
- Led 2 AI/ML projects investigating human sleep phenotypes using 2+ million days of real-world tracking time series from the Apple Watch, resulting in 1 ML conference abstract, 2 first-authored papers, and helped guide strategic directions.
- Optimized 3 health product features for Apple Watch Series 8, using insights extracted from data and end-users' feedback.
- *Models used: clustering (autoencoders, k-means, DBSCAN, hierarchical clustering) and supervised learning (LSTM), causal inference.*

PERTHERA AI | Biomedical Data Scientist | Boston, MA May 2020 – May 2021

- Built the company's first AI-powered interpretable outcome prediction and phenotyping algorithms from concept to deployment for matching providers and clinical trials with pancreatic cancer patients using multi-omics Real World Evidence (RWE) data (patented). This work has been actively implemented by the company for other diseases and datasets.
- Developed a literature recommendation system that recommended the most relevant papers to oncologists from 48,000+ Pubmed research papers using NLP methods.
- *Models/tools used: ML survival analysis (DeepCox, Nnet-survival, Dynamic-DeepHit, RSF), LSTM, clustering, GloVe, R-Shiny, SQL.*

SKILLS

Languages | Cloud Platforms: Python, R, SQL, Java, Matlab, bash, HTML, CSS | **Azure ML, Databricks, AWS, GCP, Google Vertex AI Libraries | Containerization & Pipeline Ochestration | Databases | Others:** LangChain, LangGraph, AutoGen, OpenAI, XGBoost, Streamlit, Tensorflow, PyTorch, Transformers, PySpark, NLTK, spaCy, scipy, scikit-learn, causalML, altair, dplyr, ggplot2, plotly, seaborn, matchit, caret, R-Shiny | **Docker, Kubernetes, Spark, Airflow, MLflow | MySQL, PostgreSQL, BigQuery, Snowflake, MongoDB | Scrum (Agile, Jira, Asana, Confluence), Git, CI/CD**

EDUCATION

JOHNS HOPKINS UNIVERSITY | Ph.D. in Biomedical Engineering – AI in Healthcare track

- Thesis: Machine Learning Methods for Survival Analysis of High-dimensional, Longitudinal, and Multimodal Data in Heart Disease

TRINITY COLLEGE | B.S. in Mechanical Engineering, President's Fellow, Full-Ride Scholarship

PUBLICATIONS

Google Scholar profile: <https://scholar.google.com/citations?user=iH6PLfgAAAAJ&hl=en>

Selected publications:

1. **Nguyen, H.**, Doria de Vasconcellos, H., Keck, K., Launer, L.J., Carr, J.J., Guallar, E., Lima, J.A. and Ambale-Venkatesh, B., 2024. Utility of Multimodal Longitudinal Imaging Data for Dynamic Prediction of Cardiovascular and Renal Disease: The CARDIA Study. *Frontiers in Radiology*, 4:1269023. doi: 10.3389/fradi.2024.1269023
2. **Nguyen, H.T.**, Vasconcellos, H.D., Keck, K., Reis, J.P., Lewis, C.E., Sidney, S., Lloyd-Jones, D.M., Schreiner, P.J., Guallar, E., Wu, C.O. and Lima, J.A., 2023. Multivariate longitudinal data for survival analysis of cardiovascular event prediction in young adults: insights from a comparative explainable study. *BMC medical research methodology*, 23(1), p.23. doi: 10.1186/s12874-023-01845-4
3. **Nguyen, H.T.***, Kim, H.B.*, Jin, Q.*, Tamby, S., Romer, T.G., Sung, E., Liu, R., Greenstein, J.L., Suarez, J.I., Storm, C. and Winslow, R.L., 2021. Computational Signatures for Post-Cardiac Arrest Trajectory Prediction: Importance of Early Physiological Time Series. *Anaesthesia Critical Care & Pain Medicine*, p.101015. doi: 10.1016/j.accpm.2021.101015
4. **Nguyen, H.T.***, Ambale-Venkatesh, B.*, Reis J., Wu C.O., Carr J., Nwabuo, C., Gidding S., Guallar E., and Lima J.A., 2022. Lifetime vs 10-year Cardiovascular Disease Prediction in Young Adults Using Statistical Machine Learning and Deep Learning: The CARDIA Study. *MedRxiv*. doi: 10.1101/2022.09.22.22280254
5. Gong, K.D., Lu, R., Bergamaschi, T.S., Sanyal, A., Guo, J., Kim, H.B., **Nguyen, H.T.**, Greenstein, J.L., Winslow, R.L. and Stevens, R.D., 2022. Predicting Intensive Care Delirium with Machine Learning: Model Development and External Validation. *Anesthesiology*.
6. Palepu, A.K., Murali, A., Ballard, J.L., Li, R., Ramesh, S., **Nguyen, H.**, Kim, H., Sarma, S., Suarez, J.I. and Stevens, R.D., 2021. Digital signatures for early traumatic brain injury outcome prediction in the intensive care unit. *Scientific reports*, 11(1), pp.1-9.
7. Swamynathan, R., Varadarajan, V., **Nguyen, H.T.**, Wu, C.O., Liu, K., Bluemke, D.A., Kachenoura, N., Redheuil, A., Lima, J.A. and Ambale Venkatesh, B., 2023. Association between Biomarkers of Inflammation and 10-Year Changes in Aortic Stiffness: The Multi-Ethnic Study of Atherosclerosis. *Journal of Clinical Medicine*, 12(15), p.5062.
8. Ciuffo, L., **Nguyen, H.**, Marques, M.D., Aronis, K.N., Sivasambu, B., de Vasconcelos, H.D., Tao, S., Spragg, D.D., Marine, J.E., Berger, R.D. and Lima, J.A., 2019. Periatrial fat quality predicts atrial fibrillation ablation outcome. *Circulation: Cardiovascular Imaging*, 12(6), p.e008764.

INVITED PEER REVIEW SERVICE

Reviewer for Artificial Intelligence in Medicine, Journal of Biomedical Informatics, International Journal of Medical Informatics, Frontiers in Cardiovascular Medicine, Computer Methods and Programs in Biomedicine, and International Journal of Cardiology