PHUONG-AN BUI

bui.huu@northeastern.edu • Boston, MA • +1 (857) 763 7315

linkedin.com/in/phuongan-bui • github.com/phganie • phganie.github.io/phuongan-bui

EDUCATION

Master of Science in Computer Science • *Northeastern University*　　　　　　　　**Jan 2024 - Aug 2026**
GPA: 3.78/4.0

Bachelor of Science in Computer Science and Business Administration • *Northeastern University*　　**Sep 2022 - Dec 2025**
GPA: 3.75/4.0 • Minor in Psychology • John Martinson Honors Program

SKILLS

- **Programming Languages:** Python, Java, JavaScript, TypeScript, SQL
- **Web Technologies:** React.js, Node.js, Express.js, Redux, HTML, CSS, JSON
- **APIs & Frameworks:** REST API, FastAPI, Flask, Socket.io
- **Cloud/Databases:** MongoDB, MySQL, AWS
- **ML & Data Tools:** TensorFlow, PyTorch, Scikit-learn, Pandas, NumPy, Matplotlib, Hugging Face, Google Cloud
- **Tools & Platforms:** Git/GitHub, Docker, CI/CD, Postman, Power BI, Jira, VS Code, Jupyter Notebook, Google Colab

EXPERIENCE

Center of Inclusive Computing @ Northeastern • *Research Assistant - Backend Developer*　　**Sep 2025 – Dec 2025**

- Designed and implemented **large-scale web scraper** using Python to collect and clean AI curriculum data across **500+ US universities**
- Built **ETL pipelines** to automate extraction, parsing, and organization of unstructured course information from PDFs

Khoury College of Computer Sciences • *Teaching Assistant*　　**Sep 2025 – Present; Sep – Dec 2023**

- Guide **200+ students** through complex algorithms and data structures including dynamic programming, graph algorithms, and complexity analysis
- Grade and provide detailed code reviews for **150+ assignments weekly**, focusing on correctness, optimization, and best practices

WAVE Life Sciences • *Research Informatics Systems Co-op*　　**Jan 2025 – Jun 2025**

- Developed and maintained bioinformatics platform serving **50+ researchers**, resolving **60% of bug backlog** and delivering **5+ new features**
- Built automated data pipelines and Power BI dashboards processing **10,000+ data points**, reducing integration time by **30%**
- Collaborated in **Agile sprints** with cross-functional teams to deliver features aligned with scientific workflows

Break Through Tech AI Fellowship @ MIT & Microsoft • *AI/ML Fellow*　　**May 2024 – Apr 2025**

- Built **scalable NLP evaluation workflows** using Python and Hugging Face, reducing model evaluation time by **40%**
- Debugged tokenization pipeline issues, improving model performance by **25%** through systematic root cause analysis
- Developed **automated testing framework** for prompt engineering experiments, enabling rapid iteration on **100+ test cases**

AI & FULL STACK PROJECTS

ReStacked - AI-Powered Q&A Platform • *React.js, TypeScript, Node.js, MongoDB, LLM APIs*　　**Oct 2025 - Dec 2025**

- Built **full-stack Q&A platform** with real-time features using React, TypeScript, Node.js, Express, and MongoDB
- Defined and implemented **AI Summary and Tag Suggestion** features using Gemini 2.0 Flash, **vector embeddings**, and **RAG workflows** for semantic search
- Developed **WebSocket-based real-time updates** using Socket.io, enabling live notifications and collaborative editing
- Designed and optimized **RESTful API architecture** with authentication, rate limiting, and comprehensive error handling
- Implemented **CI/CD pipeline** using GitHub Actions for continuous integration and automated deployment

ADHD Prediction from Brain Imaging • *Python, Scikit-learn, LightGBM, XGBoost*　　**Jan 2025 - Mar 2025**

- Built **ensemble ML pipeline** combining LightGBM, Random Forest, and XGBoost, achieving **F1 score of 0.73** on Kaggle competition
- Engineered data preprocessing pipeline handling **19,000+ features** using KNN imputation, PCA, and SMOTE for class imbalance
- Optimized hyperparameters across **200+ Optuna trials**, reducing training time by **40%** while maintaining prediction accuracy
- Integrated **4 complex brain imaging datasets**, implementing robust validation and cross-validation strategies

Kanbas - Learning Management System • *React.js, TypeScript, Node.js, Express.js, MongoDB*　　**Jun 2024 - Jul 2024**

- Architected and developed **full-stack LMS** with **15+ RESTful API endpoints** supporting course management, assignments, and authentication
- Implemented **Redux state management** with normalized data structures, reducing rendering time by **35%**
- Designed **MongoDB schema** with proper indexing and relationships, optimizing query performance for **50+ concurrent users**