Phong Cao

ptcao@wpi.edu · (774) 701-3932 · www.linkedin.com/in/phong-cao · GitHub: PhongCT1105

EDUCATION

Worcester Polytechnic Institute, Worcester, MA — Expected 2026

M.S. in Artificial Intelligence, B.S. in Computer Science, GPA: 3.94, **Dean's List**.

Relevant Coursework: Machine Learning, Artificial Intelligence, Deep Learning, Software Engineering, Computer Vision, Generative AI, LLM, Object-Oriented Programming, Operating System, Statistical Methods for Data Science.

SKILLS

- **Programming Languages:** Python, C/C++, TypeScript, JavaScript, Java, SQL.
- **Frameworks & Tools:** PyTorch, TensorFlow, Scikit-learn, Keras, LangChain, LanceDB, OpenAI, Azure AI vision, Azure Cosmos DB, Blob Storage, AWS EC2, RDS, React.js, NodeJS, Prisma ORM, Express, Tailwind CSS, Auth0, Flask, PostgreSQL, MySQL, GitHub, Azure DevOps, Linux, Jira, Taiga, Streamlit.
- **Certifications and licenses:** Microsoft Certified Azure AI Engineer Associate (in-progress).

WORK EXPERIENCE

Machine Learning Intern (MLOps) – *FPT Software, Vietnam* — Jun 2024 – Aug 2024

- Developed an AI-powered solution to optimize the storage, retrieval, and organization of wedding album images.
- Implemented an Image Clustering Deep Learning system applying cloud computing Azure AI vision for CNN-based image analysis, Cosmos DB for vector database storage, and Blob Storage for original image storage.
- Collected and designed a processing pipeline for 5000 images, fine-tuned model, and evaluated performance.
- Enhanced model accuracy by 15%, secured Proof of Concept approval, projected a potential $27,000 in revenue.

Data Science Intern – *FPT Software, Vietnam* — Jul 2023 – Aug 2023

- Launched a specialized ground truth database for human pose detection, featuring 10 unique poses tracked in diverse environments through Python frameworks such as OpenCV, Pandas, and Matplotlib.
- Proposed and executed a data collection strategy that resulted in the acquisition of 10,000 labeled images.
- Improved model accuracy by 15% in the initial training phases, achieving approximately 88% overall accuracy and enabling real-time application, the database is now a key resource for the other 2 projects.

Artificial Intelligence Summer Intern – *Vietnam National University, Vietnam* — Jun 2022 – Aug 2022

- Constructed an AI-driven solution to detect, and classify vehicles from video footage, and calculate speed, supporting efficient urban traffic management which potentially reduces 70% of human needs.
- Deployed YOLOv5 model for real-time vehicle detection in image recognition tasks, processed datasets from over 50 videos collected and 7000 images labeled in the surrounding area.
- Generated a vehicle detection and classification AI model with 95% accuracy; this innovation resulted in a 2nd place award at the regional Science and Technology Competition.

PROJECTS

Project Lead – *AI-powered research synthesis tool (SyntheSearch), MA* — Nov 2024 – Present

- Constructed an AI-powered web application integrating OpenAI GPT, LangChain, LanceDB, and a Retrieval-Augmented Generation (RAG) model for vector database search.
- Engineered a full-stack solution with Python backend, React frontend, and Databricks for ML pipeline management.
- Optimized literature review efficiency by 75% and enhanced research synthesis with advanced LLM-driven insights.

Lead Software Engineering – *Web application for Mass General Brigham Hospital, MA* — Mar 2024 – May 2024

- Led a team of 11 members with Agile Methodology, developed a web application for hospital management.
- Implemented core features, including a path finding algorithm, service request forms in PostgreSQL, graph insight visualization, Machine Learning scheduling system with TensorFlow for hospital staff, deployed on EC2 and RDS.
- Administered comprehensive 90 user stories and scenarios, designed 10 UML diagrams to clarify system architecture and facilitate effective communication within an Agile team of 11 members.
- Delivered fully functional hospital management integrating machine learning accuracy at 90% for task scheduling, providing real-time data insights, and a user-friendly interface with React.JS, advertised to Brigham Hospital staff.

Project Lead – *Multimodal machine learning system for Cancer Detection* — Sep 2024 – Dec 2024

- Established a robust machine learning system leveraging 54,676-feature gene expression data to predict cancer types, identify key genes, and support personalized treatment, aiming for white approaching method.
- Applied feature selection with ANOVA, Mutual Information, and Variance Threshold using overlapping; trained models with Lasso Regression and Random Forest via LOOCV and K-Fold, achieving 90% accuracy and 0.9 recall.