

Bao Hoang

+1 517-505-9949 | hoangbao@msu.edu | [Google Scholar](#) | [GitHub](#) | [Personal Website](#)

RESEARCH INTEREST

My research interests focus on trustworthy machine learning, specifically robustness, privacy, and explainability, with applications in healthcare settings.

EDUCATION

Michigan State University

Bachelor of Science, Computer Science and Advanced Mathematics, Honors College

East Lansing, MI

Expected May 2025

- GPA: 4.0/4.0
- Graduate-level Coursework: Machine Learning, Data Mining, Computational Foundations in Artificial Intelligence, Algorithmic Graph Theory.
- Honors and Awards: Honors College Excellence Scholarship, Professorial Assistantships, R.E. Phillips Memorial Scholarship, EnSURE Funding 2023 and 2024, Dean List all semesters, MSU Herzog Mathematics Competition 2024 Honorable Mention.

PUBLICATIONS

- [1] **Bao Hoang**, Yijiang Pang, Siqi Liang, Liang Zhan, Paul Thompson, Jiayu Zhou. "*Distributed Harmonization: Federated Clustered Batch Effect Adjustment and Generalization*". KDD 2024 [Paper][Code]
- [2] **Bao Hoang**, Yijiang Pang, Hiroko H. Dodge, and Jiayu Zhou. "*Subject Harmonization of Digital Biomarkers: Improved Detection of Mild Cognitive Impairment from Language Markers*". PSB 2024 [Paper][Code]
- [3] **Bao Hoang**, Yijiang Pang, Hiroko H. Dodge, and Jiayu Zhou. "*Translingual Language Markers for Cognitive Assessment from Spontaneous Speech*". InterSpeech 2024 [Paper][Code]
- [4] Yijiang Pang, **Bao Hoang**, Jiayu Zhou. "*Cross-modality debiasing: using language to mitigate sub-population shifts in imaging*". arXiv 2024 [Paper]
- [5] Yijiang Pang, Shuyang Yu, **Bao Hoang**, Jiayu Zhou. "*Towards Stability of Parameter-free Optimization*". submitted to ICLR 2025 [Paper]

EXPERIENCE

Research Assistant

ILLIDAN Lab (PI: Dr. Jiayu Zhou)

September 2021 – Present

East Lansing, MI

- Project Contributions:
 1. Improved Detection of Mild Cognitive Impairment (MCI) via Spontaneous Conversations:
 - * Led a multimodal analysis integrating linguistic, facial, and acoustic features from patient-doctor conversations to identify MCI digital biomarkers.
 - * Applied Natural Language Processing techniques to analyze lexical, semantic, and syntactic features.
 - * Utilized the VGG-Face Convolutional Neural Network to extract patients' facial emotion embeddings.
 - * Extracted Mel-frequency Cepstral Coefficients as paralinguistic features from patients' speech signals.
 - * Developed a novel deep learning-based subject harmonization approach, mapping original digital biomarkers into generalizable subject-invariant representations, resulting in a 5% increase in the AUC performance of MCI classifiers.
 2. Efficient Multi-site Brain Imaging Distributed Harmonization:

- * Enhanced the ComBat harmonization technique for brain imaging by integrating the K-means Clustering algorithm, enabling harmonization of new sites' data without retraining, thereby doubling inference time while maintaining similar performance.
- * Extended the proposed harmonization algorithm to a distributed framework using the Federated Averaging algorithm, ensuring the privacy of sensitive health information.
- 3. Machine Learning Research:
 - * Developed an innovative ensemble method combining multiple sentence transformer embeddings using Siamese neural network and Kullback–Leibler Divergence loss, leading to improved In-Context Learning performance for Large Language Models.
 - * Improved the worst-case accuracy of the Contrastive Language-Image Pre-training (CLIP) model by creating effective debiased prompts for the proposed Language-Distributional Robustness Optimization (L-DRO) algorithm.
 - * Conducted a theoretical analysis on the convergence rate of AdaGrad-Norm optimization in non-convex settings, identifying the optimal step size to minimize the upper bound of gradient norms, which informed the design of the proposed parameter-free Adam optimization (AdamG).

Machine Learning Engineer Intern

May 2022 – August 2022

FPT Software AI Center

Hanoi, Vietnam

• Project Contributions:

- 1. Healthcare Chatbot Development:
 - * Developed a multilingual healthcare support chatbot in Python using the Bot Framework SDK v4.
 - * Integrated APIs from Google Cloud and Microsoft Azure to enhance intent recognition, support multilingual conversations, and provide location-based services to find the nearest hospital.
 - * Implemented Dijkstra's algorithm for efficient indoor hospital navigation.
- 2. Smart Workplace Alert System:
 - * Engineered a real-time monitoring system using Streamlit, FastAPI, and pre-trained YOLOv5 to detect and alert on non-compliant behaviors, such as cell phone usage during work hours.

Machine Learning Engineer

January 2021 – August 2021

Projectube

Hanoi, Vietnam

- Developed a machine learning server using Flask for Projectube, an interactive platform that enables Vietnamese high school students to discover and explore extracurricular opportunities.
- Designed a personalized recommendation system leveraging the User-Based Collaborative Filtering algorithm to suggest activities aligned with users' interests, based on real-time browsing behavior analysis.
- Integrated the PhoBERT model using PyTorch to efficiently detect and filter toxic Vietnamese comments.

PROGRAMMING COMPETITIONS

- Ranked 15th, 14th, 26th, and 21st in The 2024, 2023, 2022, and 2021 ICPC East Central NA Regional Contest ([Profile](#))
- Ranked 260th, 196th, 252nd, 281st in Google Kick Start Round A, B, G, H 2022 ([Certificate](#))
- Ranked 1597th and 928th in Meta Hacker Cup Round 2 2022 ([Certificate](#)) and 2023 ([Certificate](#))
- Top 0.25% contestants in LeetCode contest platform ([Profile](#))

TECHNICAL SKILLS

Programming Languages: Python, C++, SQL, JavaScript

Backend Frameworks: Flask, FastAPI, NodeJS

Machine Learning Frameworks: TensorFlow, PyTorch, Scikit-learn

Cloud Platform: Microsoft Azure, Google Cloud

Miscellaneous: Git, L^AT_EX