

# Bao Hoang

517-505-9949 | [hoangbao@msu.edu](mailto:hoangbao@msu.edu) | [Google Scholar](https://scholar.google.com/citations?user=hoangbao) | [github.com/hoangcaobao](https://github.com/hoangcaobao)

## RESEARCH INTEREST

---

My research interest lies in trustworthy machine learning, focusing on robustness, privacy, and explainability with applications in healthcare settings.

## EDUCATION

---

### Michigan State University

East Lansing, MI

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

*Expected May 2025*

- GPA: 4.0/4.0
- Honors and Awards: Honors College Excellence Scholarship, Professorial Assistantships, R.E. Phillips Memorial Scholarship, Global Spartan Scholarship, EnSURE Funding 2023 and 2024, Dean List all semesters

## 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 [Accepted]
- [2] **Bao Hoang**, Yijiang Pang, Hiroko H. Dodge, and Jiayu Zhou. "*Translingual Language Markers for Cognitive Assessment from Spontaneous Speech*". InterSpeech 2024 [Accepted]
- [3] **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]
- [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*". arXiv 2024 [Paper]
- [6] DT Nguyen, CT Tran, TT Nguyen, **CB Hoang**, BN Nguyen, PI Cheong. "*Data augmentation for small face datasets and face verification by generative adversarial networks inversion*". KSE 2021 [Paper]

## EXPERIENCE

---

### Research Assistant

August 2021 – Present

*ILLIDAN Lab*

*East Lansing, MI*

- Project Contributions:
  1. Improved Detection of Mild Cognitive Impairment (MCI) via Spontaneous Conversations:
    - \* Led a comprehensive multimodal analysis integrating linguistic features, facial expressions, and acoustic markers from patient-doctor conversations to identify key MCI biomarkers.
    - \* Applied advanced Natural Language Processing (NLP) techniques to analyze lexical, semantic, and syntactic features, contributing to the understanding of language patterns indicative of MCI.
    - \* Utilized the VGG-Face Convolutional Neural Network (CNN) to extract patients' facial emotion features, enhancing the performance of MCI detection.
    - \* Extracted Mel-frequency cepstral coefficients (MFCC) from patient speech signals to explore the correlations between paralinguistics and MCI.
    - \* Developed a novel deep learning-based subject harmonization approach to map original multimodal features into generalizable subject-invariant representations, achieving a 5% increase in the AUC performance of MCI classifiers.
  2. Multi-site Brain Imaging Distributed Harmonization:

- \* Enhanced the ComBat harmonization technique for brain imaging by integrating a K-means clustering algorithm, allowing for the harmonization of data from new sites without requiring retraining, thus doubling inference time while maintaining similar performance.
- \* Extended the harmonization algorithm to a distributed framework using the Federated Averaging algorithm, facilitating secure multi-site harmonization without the need to share brain imaging data.
- 3. Machine Learning Research:
  - \* Developed debiased prompts to improve worst-case accuracy of Contrastive Language-Image Pre-training (CLIP) model using proposed Language-Distributional Robustness Optimization (L-DRO), enhancing model robustness.
  - \* Contributed to evaluating the upper bound of the convergence rate for a proposed tuning-free Adam optimization (AdamG) in a non-convex setting under L-smooth conditions with affine noise variance.

## Machine Learning Engineer Intern

FPT Software AI Center

May 2022 – August 2022

Hanoi, Vietnam

- Project Contributions:
  - 1. Healthcare Chatbot Development:
    - \* Developed a multilingual healthcare support chatbot in Python using the Bot Framework SDK v4, which facilitated efficient patient interaction and streamlined medical inquiries.
    - \* Integrated APIs from Google Cloud and Microsoft Azure to enhance intent recognition capabilities, support multilingual interactions, and provide location-based services to find the nearest hospital.
  - 2. Smart Workplace Alert System:
    - \* Engineered a real-time monitoring system using Streamlit for the user interface, FastAPI for backend operations, and a pre-trained YOLOv5 model for identifying and alerting about non-compliant behaviors, such as cell phone usage during work hours.

## Machine Learning Engineer

Projectube

January 2021 – August 2021

Hanoi, Vietnam

- Developed a machine learning server using Flask for Projectube, an interactive platform designed to help Vietnamese high school students explore extracurricular opportunities.
- Designed a personalized recommendation system using the K-nearest neighbors algorithm and cosine similarity, improving user experience by suggesting activities that match their interests based on their browsing behavior.
- Implemented the state-of-the-art PhoBERT model using PyTorch to effectively identify and filter toxic comments in Vietnamese, enhancing the positive community interactions on the platform.

## AWARDS

---

- Ranked 14th, 26th, and 21st in The 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 0.3% contestants in LeetCode contest platform ([Profile](#))

## TECHNICAL SKILLS

---

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

**Backend Frameworks:** Flask, FastAPI

**Machine Learning Frameworks:** TensorFlow, PyTorch, Scikit-learn

**Cloud Platform:** Microsoft Azure, Google Cloud

**Miscellaneous:** Git,  $\text{\LaTeX}$