

Thai-Hoang Pham

+1 (937) 242-3329 | ✉ hoangpham@txstate.edu | 🌐 <https://pth1993.github.io>

🔍 <https://scholar.google.com/citations?user=bht4j9oAAAAJ> | 🐙 <https://github.com/pth1993>

RESEARCH INTERESTS

Trustworthy machine learning; fairness; robustness; distribution shift; machine learning for healthcare; multimodal medical imaging; biomedical foundation model; large language model, large vision-language model

RESEARCH SUMMARY

I am an Assistant Professor of Computer Science at Texas State University, where I direct the Probabilistic and Human-Aligned Modeling (PHAM) Lab. I received my Ph.D. in Computer Science and Engineering at The Ohio State University (OSU) in 2025. My research lies at the intersection of **machine learning (ML)** and its applications in **healthcare** and **biomedicine**. My interdisciplinary work has been published in top-tier venues, including *Nature Machine Intelligence* (Impact Factor: 25.898), **NeurIPS**, **ICLR**, **AAAI**, **UAI**, **ICCV**, **KDD**, **ICDM**, and **SDM**. It has also been featured by Ohio State News, ReachMD, Technology Networks, and DeepLearning.AI. I was named a **Best Paper Award finalist** at **ICDM 2021** and received **Student Scholarships** from the **SDM 2022**, **UAI 2024**, **AAAI 2025**, and **NeurIPS 2025**. At OSU, I was honored with the **Exemplary Graduate Student Researcher Award** from the **College of Engineering** and the **Graduate Student Research Awards** from the **Department of Computer Science and Engineering** and the **Department of Biomedical Informatics**.

APPOINTMENTS

- **Texas State University** Since 08/2026
Tenure-Track Assistant Professor in Computer Science San Marcos, Texas, USA

EDUCATION

- **The Ohio State University** 2025
Ph.D in Computer Science and Engineering Columbus, Ohio, USA
 - Advisor: Ping Zhang
 - Thesis title: *Machine Learning for Biomedicine Under Dataset Shifts*
 - Committee: Ping Zhang, Lang Li, Wei-lun Chao, Xueru Zhang
- **The Ohio State University** 2024
M.Sc in Computer Science and Engineering Columbus, Ohio, USA
 - On the Basis of Candidacy for the Ph.D.
 - Committee: Ping Zhang, Lang Li, Wei-lun Chao, Xueru Zhang
- **FPT University** 2015
B.Sc in Computer Science Hanoi, Vietnam
 - Honor: *Valedictorian*
 - Thesis title: *Building a Semantic Role Labelling System for Vietnamese*

HONORS AND AWARDS

- **Exemplary Graduate Student Researcher Award.** College of Engineering, OSU. 2025
- **Graduate Student Research Award.** Department of Computer Science and Engineering, OSU. 2025
- **Graduate Student Research Excellence Award.** Department of Biomedical Informatics, OSU. 2025
- **Top Reviewer.** NeurIPS 2025. 2025
- **NeurIPS Scholar Award.** NeurIPS 2025. 2025
- **AAAI Scholarship.** AAAI 2025. 2025
- **UAI Scholarship.** UAI 2024. 2024
- **Top Reviewer.** AISTATS 2023. 2023
- **SIAM Student Travel Award.** SDM 2022. 2022
- **Best Paper Finalist.** ICDM 2021. 2021
- **VEF 2.0 Program's Scholarship.** Vietnam Education Foundation. 2017
- **Best Paper Finalist.** IALP 2016. 2016
- **Valedictorian.** FPT University. 2015
- **The Most Excellent Student (top 1).** FPT University. 2013
- **Second Prize in 21st National Mathematics Olympics.** Vietnam Mathematical Society. 2013
- **Excellent Student (top 3%).** FPT University. 2011-2015
- **Full-tuition Scholarship for Talented Student.** FPT University. 2011-2015

PROFESSIONAL EXPERIENCE

- **AIMed Lab, The Ohio State University** 2020 - Now
Graduate Research Associate Columbus, Ohio, USA
 - **PI:** Ping Zhang
 - **Topic:** Machine Learning for Biomedicine under Dataset Shifts
 - **Achievement:** Published 12 first- and co-first-author papers in top venues including *Nature Machine Intelligence*, *NeurIPS*, *ICLR*, *AAAI*, *UAI*, *ICCV*.
- **Astute Solutions** 2019
NLP Research Intern Columbus, Ohio, USA
 - **Project:** Building the chatbot framework. In particular, building the named entity recognition system that can detect named entities required by the chatbot

(SELECTED) RESEARCH PUBLICATIONS

The full list of my publications (25+ papers) is available at [Google Scholar](#) (Summary: 600+ citations, h-index: 11, i10-index: 14, * denotes equal contribution).

- [**NeurIPS 2025**] **Thai-Hoang Pham**, Jiayuan Chen, Seungyeon Lee, Yuanlong Wang, Sayoko Moroi, Xueru Zhang, and Ping Zhang (2025). The Boundaries of Fair AI in Medical Image Prognosis: A Causal Perspective. *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*.
- [**AAAI 2025**] **Thai-Hoang Pham**, Yuanlong Wang, Changchang Yin, Xueru Zhang, and Ping Zhang (2025). Open-set heterogeneous domain adaptation: Theoretical analysis and algorithm. *Proceedings of the Annual AAAI Conference on Artificial Intelligence (AAAI)*.
- [**ICCV 2025**] Jiayuan Chen*, **Thai-Hoang Pham***, Yuanlong Wang, and Ping Zhang (2025). Integrating Biological Knowledge for Robust Microscopy Image Profiling on De Novo Cell Lines. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. (**Highlight paper**).
- [**HEALTH**] Seungyeon Lee*, **Thai-Hoang Pham***, Zhao Cheng, and Ping Zhang (2025). Domain Invariant Representation Learning and Sleep Dynamics Modeling for Automatic Sleep Staging. *ACM Transactions on Computing for Healthcare (HEALTH)*. (**Impact factor: 8.0**).
- [**UAI 2024**] **Thai-Hoang Pham**, Xueru Zhang, and Ping Zhang (2024). Non-stationary domain generalization: theory and algorithm. *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*.
- [**ICLR 2023**] **Thai-Hoang Pham**, Xueru Zhang, and Ping Zhang (2023). Fairness and accuracy under domain generalization. *Proceedings of the International Conference on Learning Representations (ICLR)*.
- [**KAIS**] **Thai-Hoang Pham**, Changchang Yin, Laxmi Mehta, Xueru Zhang, and Ping Zhang (2023). A fair and interpretable network for clinical risk prediction: a regularized multi-view multi-task learning approach. *Knowledge and Information Systems (KAIS)*. (**Impact factor: 2.7**).
- [**SDM 2022**] **Thai-Hoang Pham**, Lei Xie, and Ping Zhang (2022). FAME: Fragment-based Conditional Molecular Generation for Phenotypic Drug Discovery. *Proceedings of the SIAM International Conference on Data Mining (SDM)*. (**Oral presentation**).
- [**ICDM 2022**] Seungyeon Lee*, **Thai-Hoang Pham***, and Ping Zhang (2022). Dream: Domain invariant and contrastive representation for sleep dynamics. *Proceedings of the International Conference on Data Mining (ICDM)*. (**Oral presentation**).
- [**Patterns**] **Thai-Hoang Pham**, Yue Qiu, Jiahui Liu, Steven Zimmer, Eric O'Neill, Lei Xie, Ping Zhang (2022). Chemical-induced gene expression ranking and its application to pancreatic cancer drug repurposing. *Patterns*. (**Impact factor: 6.7**. **Preview and Editorial on Patterns**).
- [**ICDM 2021**] **Thai-Hoang Pham**, Changchang Yin, Laxmi Mehta, Xueru Zhang, and Ping Zhang (2021). Cardiac complication risk profiling for cancer survivors via multi-view multi-task learning. *Proceedings of the International Conference on Data Mining (ICDM)*. (**Oral presentation**. **Selected as one of the best papers and invited to KAIS Journal Special Issue**).
- [**Nat Mach Intell**] **Thai-Hoang Pham**, Yue Qiu, Jucheng Zeng, Lei Xie, and Ping Zhang (2021). A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. *Nature Machine Intelligence*. (**Impact factor: 25.898**. **Selected Media Coverage: Ohio State News, ReachMD, Technology Networks, DeepLearning.AI**).
- [**KDD 2025**] Changchang Yin, Shihan Fu, Bingsheng Yao, **Thai-Hoang Pham**, Weidan Cao, Dakuo Wang, Jeffrey Caterino, and Ping Zhang (2025). SepsisCalc: Integrating Clinical Calculators into Early Sepsis Prediction via Dynamic Temporal Graph Construction. *Proceedings of SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*. (**Oral presentation**).

PRESS APPEARANCE & MEDIA OUTREACH

- **Artificial intelligence-assisted drug repurposing via “chemical-induced gene expression ranking”** 2022
Patterns
- **AI in drug discovery: Applications, opportunities, and challenges** 2022
Patterns
- **Automating Drug Discovery With Machine Learning** 2021
Technology Networks
- **Novel AI tool takes a DeepCE dive into potential drugs for COVID-19** 2021
Drug Target Review
- **Computer model makes strides in search for COVID-19 treatments** 2021
Ohio State News
- **AI, COVID-19 and the long haul** 2021
Nature Machine Intelligence
- **Computer Model Combats COVID-19: Finding a Solution Through Deep Learning** 2021
ReachMD
- **New AI Study Finds Drugs for COVID-19 Based on Phenotypes”** 2021
Psychology Today

(SELECTED) INVITED TALKS & PRESENTATIONS

- **Fairness and accuracy under domain generalization** 2023
 - FPT Corporation, Hanoi, Vietnam
 - Aimesoft, Hanoi, Vietnam
- **Fragment-based conditional molecular generation for phenotypic drug discovery** 2022
 - SDM 2022, Virtual
- **Cardiac complication risk profiling for cancer survivors via multi-view multi-task learning** 2021
 - ICDM 2021, Virtual
- **A deep learning framework for high-throughput mechanism-driven phenotype compound screening** 2021
 - Department of Biomedical Informatics, Ohio State University, USA
- **Transicd: Transformer based code-wise attention model for explainable icd coding** 2021
 - AIME 2021, Virtual

ACADEMIC SERVICES

- **Conference Reviewer**
 - **Machine Learning:** ICLR, ICML, UAI, NeurIPS (**Top Reviewer**), AISTATS (**Top Reviewer**)
 - **Artificial Intelligence:** AAAI, IJCAI
 - **Data Mining:** KDD, ICDM, SDM, PAKDD
- **Journal Reviewer**
 - **Computer Science Journals:** IEEE Transactions on Pattern Analysis and Machine Intelligence; IEEE Transactions on Neural Networks and Learning Systems; ACM Transactions on Knowledge Discovery from Data
 - **Biomedical Informatics Journals:** Bioinformatics; Bioinformatics Advances; Artificial Intelligence in Medicine; Journal of Biomedical and Health Informatics; Scientific Report
- **Journal Sub-Reviewer**
 - Nature Medicine; Nature Biomedical Engineering; Nature Method; Nature Communications; Science Advances
- **Advisory Board**
 - Heylion (2023-2025)

MENTORING

- **Xuwei Tan, PhD Student** 2025
Mentored and co-authored a paper on fairness in machine learning, accepted to AAAI 2026
- **Jiayuan Chen, PhD Student** 2025
Mentored and co-authored a paper on microscopy image profiling, accepted to ICCV 2025
- **Seungyeon Lee, PhD Student** 2022
Mentored and co-authored 2 papers on automatic sleep staging, accepted to ICDM 2022 and ACM Transactions
- **Biplob Biswas, PhD Student** 2021
Mentored and co-authored a paper on automatic ICD coding, accepted to AIME 2021

TEACHING EXPERIENCE

- **BMI 5553 - Predictive Analytics in Electronic Health Records** Autumn 2024
Instructor. Designed lectures, assignments, and final project; taught graduate students across CSE and BMI.
- **CSE 2111 - Modeling and Problem Solving with Spreadsheets and Databases** Autumn 2020
Teaching Assistant. Led lab sessions, graded assignments, held office hours.
- **CSE 6331 - Algorithms** Autumn 2019, Spring 2020
Teaching Assistant. Led lab sessions, graded assignments, held office hours.
- **CSE 3521/5521 - Survey of Artificial Intelligence I: Basic Techniques** Autumn 2018, Spring 2019
Teaching Assistant. Led lab sessions, graded assignments, held office hours.

SKILLS

- **Machine Learning:** Transformers, Pytorch, Scikit-learn, Keras, Tensorflow
- **Natural Language Processing:** Gensim, NLTK, Spacy, Stanford CoreNLP
- **Computer Vision:** TorchVision, OpenCV
- **Bioinformatics:** RDKit, RCDK
- **Programming Languages:** Python, Java, C/C++
- **Mathematics:** Calculus, Linear Algebra, Probability and Statistics, Optimization