TRANG M. NGUYEN

(319)333-6995 ⋄ https://nguyentr17.github.io
Amherst, MA 01002 ⋄ tramnguyen@umass.edu

RESEARCH INTERESTS

Multimodal Machine Learning (Foundation Generative Models, Fusion Mechanism Design), Explainable AI, Model Adaptation & Drift.

EDUCATION

University of Massachusetts Amherst, Amherst MA *Expected May 2027*
M.S./Ph.D in Computer Science (GPA: 4.0)
Relevant Coursework: Optimization, Advanced NLP, Intelligenct Visual Computing (3D shapes)

Grinnell College, Grinnell IA *May 2017*
B.A in Mathematics & Statistics (GPA: 3.9)

HONORS & AWARDS

2023 Honorable Mention, Ford Foundation Predoctoral Fellowship
2023 Lori A. Clarke Scholarship, University of Massachusetts Amherst
2023 Finalist, Three Minute Thesis Competition
2017 Phi Beta Kappa, Academic Honor Society
2016 First Place, National Undergraduate Project Competition (American Statistics Association)
2015 Summer Business Scholars Program Full Scholarship, University of Chicago

WORK EXPERIENCE

Lead DataOps Engineer, Tamr Inc. 2019 – 2022

· Engineered and deployed scalable data and machine learning pipelines with Apache Spark and Airflow, processing and unifying millions of enterprise records to build analysis-ready datasets.
· Led the full lifecycle of human-in-the-loop ML projects for major finance and biopharma clients, creating a single source of truth for their enterprise data to improve data consistency and analytical capabilities.

Data Scientist Assistant, American Institutes for Research 2017 – 2019

· Co-authored two published R packages (EdSurvey and WeMix) that connect to big survey databases and analyze complex survey data with imputation variance and multilevel weights.
· Implemented deep learning algorithms to extract behavioral features from videos, and natural language processing to extract features from video transcripts for classroom analytics.

ACADEMIC RESEARCH EXPERIENCE

Research Assistant, University of Massachusetts Amherst 2022 – Present
PI: Prof. Madalina Fiterau (Information Fusion Lab)

· Investigate information fusion and conflict resolution in multimodal foundation models (e.g., LLaVA, CLIP), and develop metrics to quantify model's internal cross-modal conflict signal.
· Designed fusion mechanism that can capture various types of relationship among data modalities.

Independent Researcher, Harvard University Medical School 2021 - 2022
PI: Dr. Oleg Pianykh

· Designed and implemented experiments to investigate the phenomenon of temporal ML model degradation in healthcare and non-healthcare settings.

PUBLICATIONS

* All authors are first co-authors.

· **T. Nguyen***, S. Zhang, S. Shankar, A. Fanelli, M. Fiterau. **"Video2Reaction: Mapping video to audience reaction distribution in the wild."** *Submission under review.*

· **T. Nguyen***, J. Michaels*, M. Fiterau, D. Jensen. **"Challenges in Understanding Modality Conflict in Vision-Language Models."** *Actionable Interpretability Workshop at ICML* (2025)

· S. Zhang, S. Shankar, **T. Nguyen**, A. Fanelli, M. Fiterau. **"Audio-Visual Speech Separation via Bottleneck Iterative Network."** *AI Heard That! Workshop at ICML* (2025)

· **T. Nguyen**, E. Amponsah, A. Campbell, A. Kumar, M. Fiterau, L. Shahriyari. **"Optimal Fusion of Genotype and Drug Embeddings in Predicting Cancer Drug Response."** *Briefing in Bioinformatics* (2024)

· D. Vela, A. Sharp, R. Zhang, **T. Nguyen**, A. Hoang, O. S. Pianykh. **"Temporal quality degradation in AI models."** *Scientific reports*, 12(1), 11654 (2022)

· P. Bailey, C. Kelley, **T. Nguyen**, H. Huo, C. Kjeldsen. **"WeMix: Weighted mixed-effects models using multilevel pseudo maximum likelihood estimation."** *R package version* (2021)

· P. Bailey, M. Lee, **T. Nguyen**, T. Zhang. **"Using EdSurvey to analyse PIAAC data."** In: *Large-Scale Cognitive Assessment: Analyzing PIAAC Data*, Springer International Publishing (2020)

· Y. Long*, **T. Nguyen***, I. Tareque*. **"Logistic Regression and Classification Tree on Customer Churn in Telecommunication."** In: *National Undergraduate Statistics Class Project Competition*, American Statistics Association (2016).

TECHNICAL SKILLS

Languages	Python, R, SQL, Java, Javascript, C
ML Frameworks	PyTorch, TensorFlow, diffusers
Key ML Areas	Multimodal Learning (Vision, Language, Audio), Foundation Generative Models (LLMs, VLMs, T2I, GAN), Deep Learning, Mechanistic Interpretability
Tools	Git, Docker, Slurm/HPC Environments, Cloud Platforms

SELECTED UNPUBLISHED PROJECTS

Studying the Priming Effect of Visual Large Language Models Jun - Sep 2024

· Investigated the mechanism behind the priming effect phenomenon in which Visual Large Language Models are primed to perceive a happy face as sad using in-context prompts.

· Poster was presented at New England Computer Vision (NECV) Workshop 2024.

Analyzing Eating Disorders through Social Media Anecdotes Feb - May 2023

· Analyzed different triggers and recovery patterns of people who suffer from eating disorders and self-report on Reddit from 2020 to 2023 using topic modeling and large language models (GPT-3).

Informal Network Analysis of Staff in a College Sep - Dec 2016

· Designed survey to collect data on informal communications among staff and proposed a cost-effective method to group staff in an organization using network analysis.

· Delivered a white paper and web application to the Human Resources Department of Grinnell College.

SERVICE & OUTREACH

2023 - 2024	Advisor for Voices of Data Science
2022 – 2023	Co-Chair for Voices of Data Science Conference
2020 - 2021	Data Science Talent Fellow at Open Avenues Foundation