Viet Quoc Duong

(585) 733-5492
vqduong@wm.edu
vduong143.github.io

EDUCATION

WILLIAM & MARY **WILLIAMSBURG, VA**
Doctor of Philosophy in Computer Science *January 2022 - Present*

- Advisor: Dr. Huajie Shao.
- Ph.D. Dissertation: Advancing Reliable and Concept-Based Interpretable Machine Learning with Less Supervision.
- Cumulative GPA: 3.96.
- Honors: Stephen K. Park Graduate Research Award (2025).
- Research Topics: Interpretable Machine Learning (ML), Concept-based Models, Human-in-the-Loop Large Language Models (LLMs), Out-of-Distribution (OOD) Detection, Multi-modal Learning.
- Relevant Coursework: Deep Representation Learning, Machine Learning Systems, Advanced Software Engineering, Analysis of Algorithms, Data Analysis and Simulation, Compiler and Parallel Computing, Data Visualization.

UNIVERSITY OF ROCHESTER **ROCHESTER, NY**
Master of Science in Computer Science *August 2018 - August 2021*

- Advisors: Dr. Lenhart Schubert, Dr. Jiebo Luo.
- Master's Thesis: Towards Schema-based Knowledge Extraction from Social Media Texts through Semantic Parsing.
- Concentrations: Artificial Intelligence and Computational Linguistics.
- Scholarship: Tuition Award (all eligible semesters).
- Relevant Coursework: Logical Foundation of Artificial Intelligence, Machine Learning, Natural Language Understanding, Statistical Speech and Language Processing, Formal Semantics, Markov Chain and Random Processes.

Bachelor of Arts in Data Science *September 2014 - May 2018*
Bachelor of Science in Mathematics

- Minor: Financial Economics.
- Honors: Distinction in Both Majors.
- Scholarship: Dean's Scholarship (all eligible semesters).
- Projects: Time-series Analysis on Paychex, Inc. Sales Data (industry-sponsored), Applications of the Autoregressive Conditional Model on NASDAQ Intraday Database.

PROFESSIONAL EXPERIENCE

WILLIAM & MARY **WILLIAMSBURG, VA**
Graduate Research Assistant, Department of Computer Science *August 2023 - Present*

- Supervisor: Dr. Huajie Shao.
- Developed weakly-supervised multi-modal OOD detection frameworks to identify real-world data anomalies caused by domain shifts, sensor noise, and modality misalignment, achieving high detection accuracy across multiple benchmarks and improving the reliability of deep code models in practical software engineering settings.
- Designed an interpretable Concept-based Taylor Additive model that integrates metadata-based concept grouping, additive modeling, Taylor series expansions, and tensor decomposition to deliver an efficient white-box architecture, enabling faithful prediction explanations with minimal annotation cost; this work received the KDD 2024 Best Paper Award.
- Led the development of ConceptAlign, a human-in-the-loop LLM framework that employs hybrid uncertainty estimation (combining prompt variation and model stochasticity) to selectively query expert annotations, substantially reducing labeling cost while improving concept quality and downstream task performance.
- Collaborated with AT&T Chief Data Office, the University of Illinois Urbana–Champaign, and Iowa State University on applied research projects spanning customer care analytics and healthcare AI.

AT&T CHIEF DATA OFFICE **BEDMINSTER, NJ**
Research Intern, Data Science and AI Research *July 2024 - October 2024*

- Mentors: Dr. Qiong Wu, Dr. Zhengyi Zhou.
- Built interpretable time-series forecasting models on large-scale, multi-channel customer interaction data using state-space models under data sparsity and irregular sampling.
- Applied causal inference to identify relationships between customer behavior, satisfaction, and upgrade propensity, supporting data-driven business decisions.
- Worked with real-world constraints such as noisy and imbalanced data, as well as deployment-oriented evaluation settings.

Student Intern, Data Science and AI Research *May 2023 - July 2023*

- Mentors: Dr. Qiong Wu, Dr. Zhengyi Zhou.
- Analyzed large-scale AT&T customer interaction and usage data to systematically quantify customer satisfaction and behavioral patterns.
- Investigated explainable neural classifiers and out-of-distribution detection techniques for customer account churn prediction, leading to the development of a concept-based, interpretable additive model that improves prediction accuracy while providing meaningful explanations.
- Collaborated with business stakeholders to design proactive customer retention strategies, leveraging insights from key predictive features and metadata to identify and support at-risk customers.

UNIVERSITY OF ROCHESTER **ROCHESTER, NY**

Graduate Research Assistant, Artificial Intelligence Research Lab *August 2018 - December 2021*

- Contributed to the design of Unscoped Logical Forms (ULFs), a novel semantic representation for natural language, by expanding English-language corpora and annotation schemas to cover complex linguistic phenomena.
- Built advanced semantic parsing systems for ULFs by integrating rule-based syntactic reasoning with state-of-the-art neural sequence-to-graph models, enabling expressive and interpretable logical inference for natural language understanding.
- Developed scalable language inference engines that combine deductive and probabilistic reasoning, supporting both general-purpose and domain-specific applications.
- Published three peer-reviewed papers at international conferences in artificial intelligence and computational linguistics.

Research Assistant, Visual Intelligence and Social Multimedia Analytics Lab, *September 2017 - December 2021*

- Collected and analyzed large-scale social network data to study emerging societal issues, including sexual harassment in college environments, community responses to the COVID-19 pandemic, and public attitudes toward COVID-19 vaccination.
- Applied data mining, statistical analysis, and machine learning techniques to extract trends and behavioral insights from multi-modal social media data.
- Published two peer-reviewed papers at international conferences in social network analysis and medical sciences.

TEACHING EXPERIENCE

WILLIAM & MARY **WILLIAMSBURG, VA**

Teaching Assistant, Department of Computer Science *January 2022 - May 2023*

- CSCI-241 Data Structures
- CSCI-301 Software Development

UNIVERSITY OF ROCHESTER **ROCHESTER, NY**

Teaching Assistant, Department of Computer Science *August 2016 - December 2017*

- CSC-172 Data Structures and Algorithms

PUBLICATIONS AND PRESENTATIONS

Published

*(*Co-first Authorship)* Yanfu Yan*, **Viet Duong***, Huajie Shao, and Denys Poshyvanyk. "Towards More Trustworthy Deep Code Models by Enabling Out-of-Distribution Detection". International Conference on Software Engineering (ICSE), 2025.

(Best Paper Award) Viet Duong, Qiong Wu, Zhengyi Zhou, Hongjue Zhao, Chenxiang Luo, Eric Zavesky, Huaxiu Yao, and Huajie Shao. "CAT: Interpretable Concept-based Taylor Additive Models." ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2024.

Viet Duong, Qiong Wu, Zhengyi Zhou, Eric Zavesky, Jiahe Chen, Xiangzhou Liu, Wen-Ling Hsu, and Huajie Shao. "General-Purpose Multi-Modal OOD Detection Framework." Transactions on Machine Learning Research (TMLR), 2024.

Gene Louis Kim, Mandar Juvekar, Junis Ekmekciu, **Viet Duong**, and Lenhart Schubert. "Monotonic Inference with Unscoped Episodic Logical Forms: From Principles to System." Journal of Logic, Language and Information 33, 2024

Hanjia Lyu, Junda Wang, Wei Wu, **Viet Duong**, Xiyang Zhang, Timothy D. Dye, and Jiebo Luo. "Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination." Intelligent Medicine 2, 2022

Gene Kim, **Viet Duong**, Xin Lu, and Lenhart Schubert. "A Transition-based Parser for Unscoped Episodic Logical Forms." International Conference on Computational Semantics (IWCS), 2021.

Gene Kim, Mandar Juvekar, Junis Ekmekciu, **Viet Duong**, and Lenhart Schubert. "A (Mostly) Symbolic System for Monotonic Inference with Unscoped Episodic Logical Forms." Workshops on Natural Logic Meets Machine Learning (NALOMA), 2021.

Viet Duong, Phu Pham, Tongyu Yang, Yu Wang, and Jiebo Luo. "The Ivory Tower Lost: How College Students Respond Differently than the General Public to the COVID-19 Pandemic." IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2020.

Gene Kim, Benjamin Kane, **Viet Duong**, Muskaan Mendiratta, Graeme McGuire, Sophie Sackstein, Georgiy Platonov, and Lenhart Schubert. "Generating discourse inferences from unscoped episodic logical formulas." International Workshop on Designing Meaning Representations (DMR), 2019.

Under Review

Viet Duong, Nhat Le, Qiong Wu, Zhengyi Zhou, Han Zhao, Huajie Shao. ConceptAlign: A Human-in-the-Loop Large Language Model Framework for Concept Generation and Alignment. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026.

AWARDS

Best Paper Award, ACM SIGKDD Conference on Knowledge Discovery and Data Mining (2024)
Stephen K. Park Graduate Research Award, William & Mary (2025)

SKILLS AND QUALIFICATIONS

Technical Skills

- **ML & LLM Frameworks:** PyTorch, HuggingFace, OpenAI API, TensorFlow, scikit-learn.
- **Multi-modal & LLM Techniques:** Vision–Language Models, LLM Alignment, Prompt Engineering, Active Learning, Uncertainty Estimation, Concept-Based Learning.
- **Systems & Data:** Apache Spark, Databricks, Large-Scale Dataset Curation.
- **Languages:** Python, Java, C++, SQL.

Certificates

Deep Learning, *April 2020*
DeepLearning.AI, Coursera

- **Coursework:** Neural Networks and Deep Learning, Improving Deep Neural Networks, Structuring Machine Learning Projects, Convolutional Neural Networks, Sequence Models.