Dzung V. Pham

Address: Computer Science Laboratories Rm. E250, 130 Governors Drive, Amherst, MA 01003, USA
Mobile: +1 (413) 629-9759 | *E-mail:* dzungvpham@gmail.com | *Website:* `dzungvpham.github.io`

Research Interests

I am interested in **securing user privacy and safety in AI-powered systems**. I enjoy thinking about how to **design and develop practical LLM-integrated methods** that can protect everyday users from privacy leakage and security breaches.

Keywords: security, privacy, machine learning, large language models, AI agents, federated learning, differential privacy, the web.

Education

University of Massachusetts Amherst **2022–Present**
PhD Candidate in Computer Science (GPA: 4.00). Graduation expected in 2028.
Advisors: Prof. Amir Houmansadr and Prof. Eugene Bagdasarian.

Williams College **2016–2020**
B.A. in Computer Science & in Statistics with Honors. GPA: 3.90 (Magna Cum Laude).
Thesis: *Exploring Variable Importance with Stacked Models*. Advisor: Prof. Richard D. De Veaux.

Publications

Hyejun Jeong,* Dzung Pham,* Amir Houmansadr, Eugene Bagdasarian, "AI Snitches Get Glitches: Towards Evading Agentic Surveillance." Preprint, 2026. `https://arxiv.org/abs/2606.25836`.

Dzung Pham, Kleomenis Katevas, Ali Shahin Shamsabadi, Hamed Haddadi, "AgentStop: Terminating Local AI Agents Early to Save Energy in Consumer Devices." In *ACM Conference on AI and Agentic Systems (CAIS)*, 2026. `https://arxiv.org/abs/2605.15206`.

Dzung Pham, Peter Kairouz, Niloofar Mireshghallah, Eugene Bagdasarian, Chau Minh Pham, Amir Houmansadr, "Can Large Language Models Really Recognize Your Name?" In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, 2026. `https://arxiv.org/abs/2505.14549`. An early position-version of this paper was accepted to *7th Annual PrivaCI Symposium* (2025).

Chau Minh Pham, Jenna Russell, Dzung Pham, Mohit Iyyer, "Frankentext: Stitching Random Text Fragments into Long-form Narratives." In *the 64th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2026. `https://arxiv.org/abs/2505.18128`.

Dzung Pham, Jade Sheffey, Chau Minh Pham, Amir Houmansadr, "ProxyGPT: Enabling User Anonymity in LLM Chatbots via (Un)Trustworthy Volunteer Proxies." `https://arxiv.org/abs/2407.08792`. Preprint, under review. Early non-archival version accepted and presented at the NDSS 2025 *Workshop on Measurements, Attacks, and Defenses for the Web (MADWeb)*.

Dzung Pham, Shreyas Kulkarni, Amir Houmansadr. "RAIFLE: Reconstruction Attacks on Interaction-based

Federated Learning with Adversarial Data Manipulation." In *Network and Distributed System Security (NDSS) Symposium,* 2025. https://dx.doi.org/10.14722/ndss.2025.240363.

Quan Do, Kiersten Campbell, Emmie Hine, Dzung Pham, Alex Taylor, Iris Howley, Daniel Barowy. "Evaluating ProDirect Manipulation in Hour of Code." In *Proceedings of the 2019 ACM SIGPLAN SPLASH-E Symposium (SPLASH-E '19).* dl.acm.org/doi/10.1145/3358711.3361623.

Research Experiences

AI Security Lab at UMass Amherst **2022–Present**

Conducting research on building secure and private AI systems with Prof. Amir Houmansadr and Prof. Eugene Bagdasarian. Led 3+ first-author projects on privacy and LLMs.

Research Intern | Brave Software, Inc. **Summer 2025**

Developed AgentStop (published in ACM CAIS '26), an efficient method to terminate local AI agents early to help save energy on consumer devices, achieving up to 20% wasted energy reduction with $< 5\%$ task utility loss on agentic question-answering and coding tasks (SimpleQA, FRAMES, and SWE-Bench Verified). Built a comprehensive measurement system for profiling local agents' power consumption on Apple Silicon and Nvidia Jetson devices. Mentors: Minos Katevas and Ali Shahin Shamsabadi.

Undergraduate Research | Williams College **2018-2020**

Conducted three different research projects:

- **Senior Thesis** (Statistics): *Exploring Variable Importance with Stacked Models.* Evaluated the robustness of feature importance metrics in stacked models (built with the H2O framework) and found that ensemble diversity can moderate the elevated importance of correlated features.
  Advisor: Prof. Richard De Veaux. Thesis link: https://doi.org/10.36934/t2020-099.
- **FallNet**: Developed a two-stream convolutional neural network with transfer learning from MobileNetV2 to detect people falling in video input using Keras, TensorFlow, and OpenCV, and Motion History Image. Won the 2019 Ward Prize for Best Project in Computer Science.
  Advisor: Prof. Duane Bailey. github.com/dzungvpham/fall-detection-two-stream-cnn
- **SWELL**: Designed and implemented major parts of the SWELL graphical programming language, including the parser, interpreter, and programming user interface. Taught the language to 5th-grade students at Williamstown Elementary School and evaluated its effectiveness in teaching beginner coding.
  Advisor: Prof. Dan Barowy. http://swell-lang.org/index.html

Teaching & Mentorship Experiences

Teaching Associate | UMass Amherst **Fall 2025, Spring 2026**

Instructor of Record for COMPSCI 110: Foundations of Programming. Taught fundamental concepts of programming with Python to 80+ undergraduate Computer Science majors and non-majors. Reshaped the course's examination structure to encourage logical reasoning and de-emphasize rote memorization by introducing the use of "cheat sheet" and to promote learning from mistakes by pioneering an in-person exam revision mechanism.

Undergraduate Research Volunteer Mentor | UMass Amherst **Summer 2026**

Mentoring undergraduate students in Computer Science research, with a focus on privacy and LLMs.

COMPSCI 692PA: Seminar on AI Security and Privacy | UMass Amherst **2025**
Organized a seminar on GenAI security and privacy with Prof. Eugene Bagdasarian at UMass. Invited and hosted leading experts to give talks, created and maintained course website at `https://aisec.cs.umass.edu/seminar`.

PhD Mentor | UMass Center for Data Science **2024–2025**
Mentored a group of UMass Computer Science Master's students in collaboration with industry researchers from Google Research and Meta's FAIR Lab on evaluating the privacy property of zero-th order machine learning algorithms via membership inference attacks.

Teaching Assistant | UMass Amherst **2022–2024**
Courses: Intro to Computer and Network Security, Programming Methodology, Web Programming.

Teaching Assistant | Williams College **2017–2020**
Courses: Statistical Learning & Data Mining, Regression & Forecasting, Principles of Programming Languages, Algorithm Design & Analysis, Computer Organization, Data Structures & Advanced Programming, Introduction to Computer Science.

Engineering Experiences

Meta Platforms (Facebook), Inc.

Software Engineer – Machine Learning **2020–2022**

- Trained and deployed a wide variety of ML models to identify and prevent fraud, scams, and harassment on Facebook Marketplace.
- Built and maintained scalable and reliable ML backend infrastructure to support the integrity of Marketplace for billions of daily interactions.

Software Engineer Intern **Summer 2018 & 2019**

- Built two internal tools to support software release engineers with managing version releases and debugging software build failures.
- Trained and deployed a ranking ML model to help Facebook employees discover internal job opportunities based on their career history, skills and preferences.

Honors & Awards

The Internet Society NDSS Symposium Fellowship **2025**

Sigma Xi, Phi Beta Kappa, Mu Sigma Rho | Williams College **2020**
Elected to Sigma Xi (national honorary scientific research society), Phi Beta Kappa (academic honor society), and Mu Sigma Rho (national honorary society for statistics).

Ward Prize for Best Student Project in Computer Science | Williams College **2019**
Won the prize for the FallNet research project. Also nominated for the 2020 Ward Prize for a project on

building a distributed reinforcement learning system.

Invited Talks

Can Large Language Models *Really* Recognize Your Name? **2025**
Google's Contextual Integrity Reading Group, UMass COMPSCI 690F Fall 2025

Guest Lecture | UMass COMPSCI 360: Intro to Computer Security **2025**
Towards Anonymous Chatbots with (Un)Trustworthy Browser Proxies

Guest Lecture | UMass First Year Seminar: Exploring Modern Computing **2024**
Large Language Models and User Privacy

Computer Science Colloquium | Williams College **2023**
On the Challenges of Building Privacy-preserving Chatbot Services

Services

Paper Review: NeurIPS, ICLR, USENIX Security, The WebConf, IMC, IEEE TDSC.

PhD Applicant Support Program and Application Review | UMass Amherst **2024**
Supported applicants from under-represented communities in their applications to UMass CS PhD program (e.g., resume and statement of purpose review, one-on-one counseling). Reviewed Security/ML System applications to UMass CS PhD program.

Computer Science Student Advisory Committee | Williams College **2018–2019**
Organized the first mock tech interview program in the Computer Science department at Williams College to help fellow students apply for software engineering jobs in the tech industry.

Personal Projects

CSLAC: Computer Science at Liberal Arts Colleges **2026**
Creator of the cslac.org website (open-sourced on GitHub), which showcases data on 700+ Computer Science faculty and 1000+ courses at 130+ liberal arts colleges (LACs) in the US. CSLAC aggregates faculty data from each LAC's department and combines with Google Scholar citation metrics to shed light on the landscape of Computer Science education and research in LACs.

Skills

Programming & Tools: Python, Java, JavaScript, C/C++, SQL, PyTorch, Claude Code
ML Methods: Large Language Models, Federated Learning, Adversarial Machine Learning, Membership Inference & Reconstruction Attacks
Areas: AI Security & Privacy, AI Agents, Differential Privacy, ML Systems & Infrastructure