

Tai D. Nguyen

Home: taidnguyen.github.io
Email: taing@seas.upenn.edu
Phone: (425) 246-3778

Education	University of Pennsylvania Philadelphia, PA <i>Master of Science in Engineering</i> , Computer and Information Science 2021-2024 <i>Thesis</i> : Attribute in-context learning examples with influences <i>Coursework</i> : Machine Reasoning, Theory of Computation, NLP Efficiency, NLP & Programming Languages, Software Systems, Operating Systems, Algorithms
	Haverford College Haverford, PA <i>Bachelor of Arts</i> , Economics & Statistics 2015-2019 <i>Thesis</i> : Quantifying the impact of home sharing on rental, hotel, and welfare in New York City
Publications	In-context example selection with influences [1] N. Tai , E. Wong <i>On Arxiv; In submission</i>
	Explanation-based finetuning makes models more robust to spurious cues [2] J. Ludan, Y. Meng*, T. Nguyen* , S. Shah*, Q. Lyu, M. Apidianaki, C. Callison-Burch Association for Computational Linguistics (ACL 2023)
	Software entity recognition with noise-robust learning [3] T. Nguyen , Y. Di, J. Lee, M. Chen, T. Zhang IEEE/ACM International Conference on Automated Software Engineering (ASE 2023)
Research Experience	H2Lab @ University of Washington Remote <i>Research assistant</i> March 2024 - now - Advisor: Ben Bogin - Building a data attribution benchmark for tracing LLM capabilities from pretraining
	Chris Callison-Burch Lab Philadelphia, PA <i>Student lead</i> May 2023 - Feb. 2024 - Advisors: Marianna Apidianaki, Chris Callison-Burch (UPenn) - Lead UPenn team of 3 students on IARPA HIATUS project, building lexical features for interpretable authorship attribution. - Proposed an interpretable system that outperforms neural embeddings on cross-domain evaluation.
	IBM Burlington, VT <i>Data scientist</i> 2019 - 2021 - Built analysis tools to support the development of IBM System-z servers for 1200+ engineer users, including algorithms and interactive visualization tools that helped discover and confirm 100+ erroneous machine parts. - Mentored 3 interns and peers in obtaining the IBM Data Science Professional Certificate.

	Human-centered Software Systems Lab <i>Student lead</i> - Advisors: Muhao Chen (USC), Tianyi Zhang (Purdue) - Led project on building a massive corpus and an efficient noise-robust model for doing named entity recognition (NER) in the software domain. - First-authored publication accepted to ASE 2023.	Remote Summer 2022
Teaching & Service	NeurIPS 2023 Workshop: ATTRIB <i>Program committee member</i> - Reviewed and recommended papers for acceptance.	2023
	STEAM for Vietnam <i>Data science lead</i> - Co-taught workshop for Vietnamese teachers to build TJBOT, a Raspberry Pi project powered by IBM Watson. - Made recommendations to improve student's graduation rate by 20%. - Developed data pipeline and dashboards to support internal teams.	2021
	Haverford College Admissions <i>Senior interviewer</i> - Conducted 200+ interviews with prospective students to help recommend them for admissions to Haverford.	2018-2020
Projects	NFL Big Data Bowl <i>Finalist</i> - Awarded \$15,000 as one of 5 finalists among 200+ submissions in a competition hosted by the National Football League (NFL). - Designed regression model simulating a kick returner's most optimal return path; proposed new metrics to evaluate returner's decision making. Presentation / submission	2022
	Underthesea NLP <i>Open-source contributor</i> - Trained models on aspect-based sentiment classification and NER for Vietnamese, improving 15% F1 from existing models. Repository	2021 - 2022
Talks	UPenn ML & Formal Methods Seminar Influence-guided example selection for in-context learning	Feb. 2023
Honors	Vietnam Education Foundation (VEF) Fellow IBM Volunteer Excellence Award IBM New Developer Jumpstart Winner	2021 2021 2020
Skills	Programming: Python, C++, C, Java, Bash, git, RunAI/slurm Framework: PyTorch, Docker, Google Cloud, AWS, plotly, d3.js Languages: Vietnamese (native), English (native), Mandarin (intermediate)	

Last updated April 9, 2024.