Tai D. Nguyen

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

Education

University of Pennsylvania Philadelphia, PA
Master of Science in Engineering, Computer and Information Science 2021-2024
Thesis: In-context data attribution via prompting language and language-vision models
Coursework: Machine Reasoning, Theory of Computation, NLP Efficiency, NLP & Programming Languages, Software Systems, Operating Systems, Algorithms

Haverford College Haverford, PA
Bachelor of Arts, Economics & Statistics 2015-2019
Thesis: Quantifying the impact of home sharing on rental, hotel, and welfare in New York City

Publications

Explanation-based finetuning makes models more robust to spurious cues
[1] 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
[2] **T. Nguyen**, Y. Di, J. Lee, M. Chen, T. Zhang
IEEE/ACM International Conference on Automated Software Engineering **(ASE 2023)**

In-context example selection with influences
[3] **T. Nguyen**, E. Wong
On Arxiv; In submission

Research Experience

Chris Callison-Burch Lab Philadelphia, PA
Student lead May 2023 - now
- 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.

Human-centered Software Systems Lab Remote
Student lead Summer 2022
- Advisors: Tianyi Zhang (Purdue), Muhao Chen (USC)
- 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.

Penn Computational Social Science Lab Philadelphia, PA
Research assistant Sep. 2021 - May 2022
- Advisor: Homa Hosseinmardi
- Assisted research on bias & misinformation in mainstream media from a decade of closed-caption news.

- Experimented with range of methods including Latent Dirichlet Allocation (LDA) for topic labeling and text segmentation of political topics.

IBM Burlington, VT
Data scientist 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.

Teaching & Service

NeurIPS 2023 Workshop: ATTRIB 2023
Program committee member
- Reviewed and recommended papers for acceptance.

STEAM for Vietnam 2021
Data science lead
- 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.

Haverford College Admissions 2018-2020
Senior interviewer
- Conducted 200+ interviews with prospective students to help recommend them for admissions to Haverford.

Projects

NFL Big Data Bowl 2022
Finalist
- 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

Underthesea NLP 2021 - 2022
Open-source contributor
- Trained models on aspect-based sentiment classification and NER for Vietnamese, improving 15% F1 from existing models.
Repository

Talks

UPenn ML & Formal Methods Seminar Feb. 2023
Influence-guided example selection for in-context learning

Honors

Vietnam Education Foundation (VEF) Fellow 2021
IBM Volunteer Excellence Award 2021
IBM New Developer Jumpstart Winner 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 February 20, 2024.