

Huy Thanh Vu

Updated September 2022

✉ hvu@cs.stonybrook.edu

🌐 [linkedin.com/in/huy-vu-94874a130](https://www.linkedin.com/in/huy-vu-94874a130)

☎ 631-310-7194

EDUCATION

Stony Brook University

New York, U.S.

Doctor of Philosophy Candidate in Computer Science (GPA: 3.94/4)

2018 - Expected 2023

Ho Chi Minh City University of Science

Ho Chi Minh City, Vietnam

Bachelor of Mathematics and Computer Science (GPA: 9.2/10 - ranked 1st in university graduation batch)

2012 - 2017

Tokyo University of Agriculture and Technology

Tokyo, Japan

International Exchange Program with Full Sponsored Scholarship

2014 - 2015

KEY SKILLS

- **Languages and Tools:** Python (fluent), C++ (fluent), Pytorch (fluent), Tensorflow (fluent), Slurm (fluent), Git (fluent), Docker (fluent), SQL (fluent), Hadoop (familiar), Spark (familiar)
- **Research Interests:** Machine Learning, Natural Language Processing, Computational Social Science, Psychology Research-applied Machine Learning, Data Mining

WORKING EXPERIENCE

Meta (Facebook)

California, U.S.

Machine Learning Engineer Intern

June 2022 - August 2022

- Creating synthetic training data using Text-to-Speech pipeline (based on Neural Transducer RNN-T) to improve Automatic Speech Recognition task. The main goal is to reduce the performance gaps across age-groups, making the model perform more evenly across different demographics.
- Reduced the performance gaps in ASR task across age-groups up to 13% by: Doing literature review to looking for the right dataset to create synthetic data from. Analyzing speed and pitch features distribution on real data to make sure the generated augmented data simulate closely the real data.

NVIDIA

California, U.S.

Deep Learning Engineer Intern

June 2021 - August 2021

- Thorough surveying literature landscape and implementing strong models such as Distill-FiD for Open Domain Question Answering (ODQA) task. Working on tuning and accelerating models (running on NVIDIA's DGX A100 and Selene supercomputer) to minimize computational costs while still obtaining high performance.
- Reduced total training time up to 34% for the Distill-FiD pipeline using a variety of optimization techniques including half-precision training, parallelizing data processing, fused Adam optimizer. Proposed the idea of utilizing retriever's training iterations to reduce reader's training iterations, which significantly decreases total training costs.

Amazon Alexa AI

Washington, U.S.

Applied Scientist Intern

June 2020 - August 2020

- Working on distilling strong, powerful, but heavy transformers-based language model into a more compact, fast-running model while still maintaining reasonable performances.
- Proposed taking advantage of large models' high performances (RoBERTa, XLM) to run in real-time on devices with small computing capability by combining fast, simpler low-level features and pre-computed deep embeddings with on-the-fly embeddings computing when processing coming data.

Brookhaven National Lab

New York, U.S.

Research Assistant

September 2019 - May 2020

- Working with material scientists, implementing text mining algorithms for information retrieval from material science research literature database.
- Built a search engine using BERT-based contextualized embeddings customized specifically for material science research literature, in which chemical formulas (e.g., HClO) are broken down into separate elements tokens (H, Cl, and O) for better learning these elements' embeddings. The engine received highly rated feedbacks from material scientists.

University of Pennsylvania

Pennsylvania, U.S.

- Working with psychologists, exploring novel ideas of applying contextualized word embeddings into psychology research. Building a model that takes embeddings of a social media user's text posts and embeddings of a question text (from a psychological questionnaire measuring participants' personality) to predict his/her response to the question.
- Proved validity of the proposed hypothesis, with correlation up to 0.39 (significant in psychology field). Paper accepted to EMNLP Findings 2020.

Ho Chi Minh City University of Science**Ho Chi Minh City, Vietnam**

Research Assistant

August 2016 - December 2017

- Proposing novel method segmenting brain images by combining Gaussian Mixture Model and Deep Learning approach, getting the benefits of the two methods regarding accuracy and running time.
- Improved up to 8.6(/100) points and beat state-of-the-art methods. Paper accepted to WACV 2017 conference.

RESEARCH**Stony Brook University****New York, U.S.***Texts Generation Conditioned on Personality and Mental Health Variables*

April 2019 - Present

- Working with psychology professors, social scientists from Stanford Uni., Uni. of Melbourne, and Uni. of Pennsylvania to build language generative models generating texts conditioned on Big5 personalities or mental health variables (e.g., depression, life-satisfaction score). These models help understanding the language of different psychological traits as well as are potential for humanized chatbots for psychology research purposes.
- Successfully created models built upon GPT-2 simulating language of humans having different psychological traits. Conducted deep analysis in many aspects of the models from technical and psychology research perspectives. Now in process of writing paper.

University of Pennsylvania**Pennsylvania, U.S.***Texts Embeddings Approach to Psychology Questionnaire Research*

January 2020 - Present

- Proposing a novel approach to study psychological questionnaire research, in which using contextualized texts embeddings to capture semantics of questions in a questionnaire. These embeddings then can be used as features to train a model that predicts responses of a participant to that question. After training, we use that model to predict responses to a new, out-of-sample set of questions without needing the participant to manually read and answer them. This work potentially can save significant costs and time in developing a new psychological questionnaire.
- Successfully produced proof of concept experiments on a variety of questionnaires such as Big5 personality, HSQ, 16PF. Obtained correlation up to 0.45 (significant correlation in psychology research), which matches predictions from human annotators. Now in process of writing paper.

PUBLICATIONS

- Huy Vu**, S. Giorgi, J. Clifton, N. Balasubramanian, A. Schwartz, "Modeling Latent Dimensions of Human Beliefs", Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), 2022
- G. Pasquini, G. Ferguson, I. Bouklas, **Huy Vu**, et. al., "The where and when of COVID-19: Using ecological and Twitter-based assessments to examine impacts in a temporal and community context", PLOS One, 2022
- A. Ganesan, M. Matero, A. Ravula, **Huy Vu**, A. Schwartz, "Empirical Evaluation of Pre-trained Transformers for Human-Level NLP: The Role of Sample Size and Dimensionality", North American Chapter of Association Computational Linguistics (NAACL), 2021
- Huy Vu**, S. Abdurahman, S. Bhatia, L. Ungar, "Predicting Responses to Psychological Questionnaires from Participants' Social Media Posts and Question Text Embeddings", Findings of Empirical Methods in Natural Language Processing (EMNLP), 2020
- M. Matero, Y. Son, S. Giorgi, **Huy Vu**, M. Zamani, P. Limbachiya, S. Guntuku, A. Schwartz, "Suicide Risk Assessment with Multi-level Dual-Context Language and BERT", Proceedings of the 6th Workshop on Computational Linguistics and Clinical Psychology, 2019
- D. Nguyen, **Huy Vu**, H. Ung, B. Nguyen, "3D-Brain Segmentation Using Deep Neural Network and Gaussian Mixture Model", IEEE Winter Conference on Applications of Computer Vision (WACV), 2017

AWARDS AND ACHIEVEMENTS

- Merit Fellowship of \$5000 2018 – 2019 (Computer Science Department, Stony Brook University)
- Completion of Business Administration Course 10/2017 - 2/2018 (University of Economics Ho Chi Minh City)
- National Program for Mathematics Development Award 2013, 2016 (Vietnam Institute for Advanced Mathematics Study)
- International Exchange Program Full Scholarship 2014 – 2015 (Tokyo University of Agriculture and Technology, Japan)
- Excellence in Academic Activities Award 2013 (Ho Chi Minh City University of Science)
- KumHo Asiana Scholarship 2012 (for 1st-ranked students in university's entrance, KumHo Asiana Vietnam)