TUAN ANH BUI

tuananh.bui@monash.edu
Website ◇ Scholar ◇ Github ◇ Linkedin

INTRODUCTION

- A **Researcher**: Strong track record of publishing A/A* papers on cutting-edge and competitive topics, i.e., Trustworthy ML and Generative AI.
- An **Engineer**: Hands-on experience in various ML/AI projects in diverse domains and applications, i.e., Healthcare (sample), Defence (demo), and Fintech.
- An **Educator**: Extensive teaching experience in ML/AI, i.e., (Head) Teaching Associate @Monash, Lecturer @VietAI, ML/AI Blogger at tuananhbui89.github.io.
- A Hard-Working person trying to be more Productive: Often taking responsibility for multiple (above) roles simultaneously (my humble principles).

EDUCATION

Faculty of Information Technology, Monash Universtiy 2019 - 2023
PhD in Computer Science *Australia*

- Thesis title: "Enhancing Adversarial Robustness: Representation, Ensemble, and Distribution Approaches"
- Research interests: Trustworthy Machine Learning, Generative models, Representation Learning.

Hanoi University of Science and Technology 2007 - 2012
B.Sc. in Electronics and Telecommunications (Honour) *Vietnam*

- Thesis title: "Research on optimal trade-off between power saving and QoS". Best Thesis Award.
- GPA: 8.44/10 (top 5%, Merit for Excellent Graduate Student).

WORKING EXPERIENCE

Monash University 2023 -
Research Fellow & Project Co-Lead *Australia*

- Leading multiple research projects in Trustworthy AI, collaborating with Defense Science and Technology Group (DST):
  - Led project "Personalization and Anti-Personalization GenAI", protecting users data from being used in GenAI.
  - Led project "Machine Unlearning", removing or editing unwanted information/concepts from GenAI.
  - Led project "Adversarial Attack and Defense", making ML models more robust and reliable against adversarial attacks.

RapidAI - A Healthcare Startup in US 2021 - 2021
Machine Learning Consultant *Remote*

- Led development of ML-based stroke detection system using gaze deviation analysis.
- Designed and implemented end-to-end pipeline from data processing to model deployment.
- Achived 80%+ accuracy on clinical validation dataset (sample).

TrustingSocial - A Fintech Startup 2019 - 2019
Research Engineer on Computer Vision *Vietnam*

- Developed face matching and face recognition algorithm for credit scoring system.

Temasek Lab, Singapore University of Technology and Design (SUTD) 2017 - 2019
Research Engineer on Computer Vision *Singapore*

- Improved Generative Adversarial Networks (GANs), esp. in the mode collapse problem.
- Developped ML modules to detect and track Undefined Flying Objects (i.e., Drones) in the Sky Surveillance - Flying Object Detection (SSFOD) project (sample 1)(sample 2)(demo)(demo 2).
- Scaled image retrieval modules handling large large-scale real-world dataset (appr. 200k images) in the Urban-area Scene Based Localization (USBL) project (system design)(paper)(project).
- Developed neural network compression techniques for FPGA deployment (uncompleted patent).

Viettel R&D Institute, Viettel Group 2012 - 2016
Digital Signal Processing Engineer *Vietnam*

- Designed waveforms for VHF and HF tactical radio on Software Defined Radio Platform

· Developed baseband processing algorithms for embedded systems (simulation on Matlab and implement on C55x, TI Fixed-Point processor)
· Received multiple Creative Idea Awards for technical innovations.

PUBLICATIONS

Total 10+ first-author publications, 300+ citations covering GenAI (i.e., ECCV18, AAAI19,), Trustworthy ML (i.e., NeurIPS23, ICLR22, ECCV20,) and their intersection (i.e., NeurIPS24, Preprint [1], [2]) . **Google Scholar**.

A. Bui, T. Vu, T. Le, J. Kim, T. Abraham, R. Omari, A . Kaur, D. Phung, *"Mitigating Semantic Collapse in Generative Personalization with a Surprising Simple Test-Time Embedding Adjustment"*. Under submission. [paper]

A. Bui, T. Vu, L. Vuong, T. Le, P. Montague, T. Abraham, J. Kim, D. Phung, *"Fantastic Targets for Concept Erasure in Diffusion Models and Where To Find Them"*. Accepted to ICLR 2025. [paper, code]

A. Bui, L. Vuong, K. Doan, T. Le, P. Montague, T. Abraham, D. Phung, *"Erasing Undesirable Concepts in Diffusion Models with Adversarial Preservation"*. Accepted to NeurIPS 2024. [paper, slide, code, project]

A. Bui, T. Le, Q. Tran, H. Zhao, D. Phung, *"A Unified Wasserstein Distributional Robustness Framework for Adversarial Training"*. Accepted to ICLR 2022. [paper, code]

A. Bui, T. Le, H. Zhao, P. Montague, O. de Vel, T. Abraham, D. Phung, *"Improving Adversarial Robustness by Enforcing Local and Global Compactness"*. Accepted to ECCV 2020. [paper, code]

NT Tran, **A. Bui**, NM Cheung, *"Dist-gan: An improved gan using distance constraints"*. Accepted to ECCV 2018. [paper, code]

TEACHING EXPERIENCE

Deep Learning - (Head) Teaching Associate 2020 -
Monash University *Australia*
· Topics: Basic and Advanced Deep Learning, i.e., Optimization, CNNs, RNNs, Tranformers, Generative models.

Intelligent image and video analysis - Head Teaching Associate 2023 -
Monash University *Australia*
· Topics: Classic Computer Vision (i.e., Edge/Keypoint detection, Morphology, SIFT) and Recent CV Applications (i.e., Object Detection/Segmentaion).

Advanced Computer Vision - Lecturer 2023 -
VietAI - An Education Nonprofit Organization in Vietnam *Remote*
· Topics: Deep Generative Models (sample lecture) and Multimodal Learning (sample lecture).

HONOURS & AWARDS

2024	Future of Life Institute's Fellowship, US.
2023	DAAD PostdocNetAI Fellowship in Generative Models, Germany.
2022, 2025	Top 10% Reviewer at AISTATS 2022, Notable Reviewer at ICLR 2025.
2019-2023	Faculty of Information Technology's Scholarship, Monash University, Australia.
2012-2015	Creative Idea Award for Technical Innovations at Viettel R&D Institute, Vietnam.
2012	Merit for excellent graduate student, HUST, Vietnam.
2012	Best Thesis Award in Thesis Defence, SET, HUST, Vietnam.
2012	Third prize in Student Conference on Scientific Research, HUST, Vietnam.
2007	Second prize in National Physics Olympiad for High School Students, Vietnam.

GRANTS

2024	Trustworthy Generative AI: Towards Safe and Aligned Foundation Models ($800K, Funded by Department of Defence, Australia). (Role: Chief Investigator/Project Co-Lead).
2024	Machine Learning via Adversarial Confusion (Under submission). (Role: Contributor).
2024	Erasing Undesirable Concepts from Foundation Models (Under submission). (Role: Co-PI).