

HUNG VINH TRAN

+61-433-532-892 (AU) | tranvinhhung99@gmail.com |  [chenxing1999](#) | [Google Scholar](#)

EXP SUMMARY

PhD candidate at UQ (submitting Dec 2026) working on efficient recommender systems, with 4 years of prior industry experience at Cinnamon AI building production document-AI pipelines for clients in Japan and Taiwan. Two first-author papers at WWW (2025 oral, 2026) and one at TOIS. My research focuses on making recommendation models cheaper to train and deploy – 90% embedding storage reduction, 75% training-time reduction, 99% data reduction with 93–95% performance retention.

EDUCATION

The University of Queensland

Doctor of Philosophy in Computer Science

Jul. 2023 – Dec. 2026 (Expected)

Brisbane, QLD, Australia

- *Research Topic:* Scalable and Lightweight On-Device Recommender Systems
- *Supervisor:* Assoc. Prof. Rocky Tong Chen (primary) & Prof. Hongzhi Yin

Ho Chi Minh City University of Science (HCMUS)

Bachelor of Science in Information Technology (Honors) | GPA: 8.68/10 – Top 5%

Sep. 2017 – Sep. 2021

Ho Chi Minh City, Vietnam

SELECTED RESEARCH OUTPUTS

- [1] **Tran, H. V.**, Chen, T., Gao, X., Yu, J., Monteil, J., & Yin, H. (2026). FOSTER: First-order Dataset Distillation for Text-based Sequential Recommendation. (In submission)
 - Developed FOSTER framework combining first-order bi-level optimization, stochastic item subset sampling, and a distributional hypothesis regularizer for efficient dataset distillation in text-based sequential recommendation.
 - Approximated full-dataset performance on Amazon Games, Foods, and Yelp-Nashville using only 20-60 synthetic sequences (<0.1% of training data), while reducing distillation time by $10\times$ over prior state-of-the-art.
 - Demonstrated cross-architecture transferability of distilled data to Qwen3-4B, retaining 87-95% of full-dataset Recall@10 with under 10% of the training time.
- [2] **Tran, H. V.**, Chen, T., Wen, H., Nguyen, Q. V. H., Cui, B., & Yin, H. (2026). Efficient Content-based Recommendation Model Training via Noise-aware Coreset Selection. WWW 2026 (CORE A*, CCF A).
 - Developed NaCS framework combining submodular optimization with progressive label correction for efficient training of both traditional and LLM-based recommendation systems.
 - Recovered 93-95% of full-dataset AUC with 1% training data, reducing 75% of model training time.
- [3] **Tran, H. V.**, Chen, T., Ye, G., Nguyen, Q. V. H., Zheng, K., & Yin, H. (2025). On-device Content-based Recommendation with Single-shot Embedding Pruning: A Cooperative Game Perspective. WWW 2025 (CORE A*, CCF A, Oral)
 - Proposed cooperative game theory approach for embedding parameter attribution, enabling single-shot pruning without expensive retraining cycles.
 - Achieved 90% storage reduction while maintaining competitive accuracy across multiple memory budgets.
- [4] **Tran, H. V.**, Chen, T., Nguyen, Q. V. H., Huang, Z., Cui, L., & Yin, H. (2025). A Thorough Performance Benchmarking on Lightweight Embedding-based Recommender Systems. TOIS 2025. (Q1, CCF A)
 - Conducted systematically benchmarks for various light-weight embedding-based recommender systems.

EMPLOYMENT HISTORY

Cinnamon AI, <https://cinnamon.ai>

Project Technical Leader & AI Researcher

Jul. 2021 – June 2023

Ho Chi Minh City, Vietnam

- Led 3 digitisation projects – 10-person team across three countries (Vietnam, Japan, and Taiwan) – from PoC to production across diverse document types (traditional, handwritten forms).
- Introduced Optuna hyperparameter tuning and custom LayoutLM clustering model to address data quality challenges and improve entity grouping accuracy.

- Developed augmentation techniques to mitigate annotation errors, achieving 5-10% model performance improvement.
- <https://www.kline.co.jp/en/news/drybulk/drybulk-20231023.html>

Cinnamon AI, <https://cinnamon.ai>

AI Researcher

Jul. 2019 – June 2021

Ho Chi Minh City, Vietnam

- Developed augmentation method addressing train/inference gap in annotation workflow, improving model accuracy by 5-10%; adopted as standard across all projects.
- Built AI template repository (FastAPI serving, Docker, CircleCI, AWS) reducing project deployment setup from 2 days to 2 hours; became inference pipeline for nearly all client projects.
- Mentored 2 junior researchers from bootcamp to full-time positions.

VOLUNTEER ACTIVITY

Reviewer for: KDD 2025, WWW 2026, AAAI 2026, RecSys 2025-2026

SKILLS

Research areas: Recommender systems, embedding compression, coreset selection, dataset distillation, bilevel optimisation, LLMs for recommendation.

Languages & frameworks: Python, PyTorch, CUDA, HuggingFace, FastAPI, Docker.

Infra & tooling: AWS, CircleCI, Optuna.

RESEARCH GRANTS AND RELEVANT AWARDS

Australian Government Research Training Program (RTP) Scholarship	2023 – 2027
Nomination for Excellent Student of Ho Chi Minh City Vietnamese News Link	2021
Dean's List of HCMUS Honor Program	2020
Merit-based Scholarship of HCMUS	2017 – 2021

ACADEMIC REFEREES

Release upon contact.