

DAE YON HWANG

Toronto, ON, Canada

Email:  LinkedIn:  Personal Website: 

EDUCATION

University of Toronto	Ph.D. in Electrical & Computer Engineering, 4.0/4.0	Nov 2022
Texas A&M University	Master of Science in Electrical Engineering, 4.0/4.0	May 2016
Hanyang University	B.S. in Electronic Engineering, 3.56/4.0 (<i>Cum Laude</i>)	Feb 2014

WORK EXPERIENCE

Layer 6 AI @TD - Senior Research Machine Learning Scientist *Trustworthy AI* May 2025 - Present

- **Black-Box Agent Optimization**

- Developed a Sequential Monte Carlo framework for optimizing black-box LLM agents, connecting reinforcement learning and Bayesian inference for test-time policy optimization
- Built a lightweight critic to guide black-box LLM policies, reducing optimization cost by **20%** while improving performance by **16%** over the GRPO baseline across agentic tasks

- **Conversational RAG with Safety Guardrails**

- Deployed a production RAG system balancing model capability with banking compliance requirements
- Designed query classification, rewriting, and abstention mechanisms to enforce policy-aligned responses
- Implemented a three-layer safety framework to mitigate adversarial inputs and jailbreak attacks
- Led automated agentic red-teaming and human-in-the-loop evaluation to identify and mitigate model safety failures
- Launched a publicly accessible customer-facing GenAI chatbot on TD Insurance's website, with **zero observed** jailbreak incidents during post-deployment monitoring

Amazon Science - Applied Scientist II *Artificial General Intelligence*

Sep 2022 - Apr 2025

- **Multimodal Foundation Model Post-Training**

- Generated synthetic multi-step reasoning data for post-training multimodal foundation models
- Developed LLM-as-a-Judge pipelines for data filtering and quality control in post-training
- Designed and evaluated fine-tuning and preference-optimization methods for domain knowledge and reasoning
- Contributed to post-training **Amazon Nova** foundation models, improving performance across general and domain-specific tasks including legal and finance

- **Foundation-Model Recommendation System**

- Developed and deployed a foundation-model recommendation system combining pre-training, instruction tuning, preference optimization, and RAG
- Designed online evaluation and continuous-feedback mechanisms and validated the system through customer-facing experiments
- Deployed the system on Prime Video, achieving a **20%** improvement in positive-feedback rate over the control model in online experiments

- **Information Retrieval for Alexa**

- Developed LLM-based synthetic-data generation and representation-learning methods to improve zero-shot retrieval across diverse scenarios
- Designed dimensionality-reduction methods that improved retrieval efficiency and generalization, achieving **5×** faster retrieval with comparable performance

University of Toronto - Research Assistant *Biometrics Security Lab*

Sep 2018 - Sep 2022

- **Robust User Verification from Physiological Signals**

- Developed multimodal physiological-signal models for robust user verification and adversarial resilience

- Collected and released physiological signals from 170 participants, creating a large public dataset for biometric research
- **Human Activity Recognition with Wearable Device**
 - Developed hierarchical multimodal models combining inertial and physiological sensors for wearable activity recognition

Hyundai MOBIS - Research Engineer *DAS Control Engineering*

Jul 2016 - Feb 2018

- Developed and evaluated driver-monitoring and attention-warning systems through simulation and on-road testing across diverse driving environments

SELECTED PUBLICATIONS

D. Y. Hwang , et al., "Agentic Monte Carlo: Simulating Reinforcement Learning for Black-Box Agents"	ICML 2026
Y. Jiang, D. Y. Hwang , et al., "Chartographer: Counterfactual Chart Generation for Evaluating Vision-Language Models"	EMNLP 2026
D. Y. Hwang (<i>Contributor</i>), "The Amazon Nova Family of Models: Technical Report and Model Card"	Amazon Science 2024
D. Y. Hwang , et al., "Link, Synthesize, Retrieve: Universal Document Linking for Zero-Shot Information Retrieval"	EMNLP 2024
D. Y. Hwang , et al., "EmbedTextNet: Dimension Reduction with Weighted Reconstruction and Correlation Losses for Efficient Text Embedding"	ACL 2023
D. Y. Hwang , et al., "GAN-LM: Generative Adversarial Network using Language Models for Downstream Applications"	INLG 2023
D. Y. Hwang , et al., "PBGAN: Learning PPG Representations from GAN for Time-Stable and Unique Verification System"	IEEE TIFS 2021
D. Y. Hwang , et al., "Evaluation of the Time Stability and Uniqueness in PPG based Biometric System"	IEEE TIFS 2020

SELECTED RESEARCH TALKS

From Generative AI to Agentic AI: Building Trustworthy AI	Kyungpook National University
Introduction to Mechanistic Interpretability	TD Bank
Integrating GANs with Language Models	Amazon

PROFESSIONAL SERVICE

Reviewer: EMNLP, ACL, ARR, NeurIPS, IEEE JBHI, IEEE TIFS

Program Committee: EMNLP 2023 Industry Track

SKILLS

Research: Foundation Models, LLM Post-Training, Reinforcement Learning, Agentic AI, Reasoning, Multimodal Learning, Trustworthy AI, Model Evaluation

Engineering: Python, PyTorch, C++, AWS

Languages: English, Korean