

Yee Man (Kath) Choi

PHD STUDENT · DAVID R. CHERITON SCHOOL OF COMPUTER SCIENCE

University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

✉ ymchoi@uwaterloo.ca | 🐦 @yeemanchoi

Education

University of Waterloo

PHD COMPUTER SCIENCE

- Advisor: Dr. Freda Shi, Dr. Krzysztof Czarnecki

Waterloo, ON, Canada

Sep 2025 - present

Language Technologies Institute - Carnegie Mellon University

MS COMPUTATIONAL DATA SCIENCE (MAJOR IN ANALYTICS TRACK)

- Advisor: Dr. Lei Li

Pittsburgh, PA, United States

Aug 2023 - Dec 2024

University of Toronto - St. George Campus

BASc COMPUTER ENGINEERING (MINORS IN ARTIFICIAL INTELLIGENCE)

- Undergrad Thesis: Identify Subsets of Bugs in Mozilla Dataset of interest for Machine Learning
- Advisor: Dr. Shurui Zhou, Dr. David Lie

Toronto, ON, Canada

Sep 2017 - May 2022

Research Interests

- Large Language Models, Agent Alignment, Model Security, Code Generation, Computational Linguistics, AI4Research

Professional Experience

2024-2025 **Graduate Research Assistant**, Carnegie Mellon University

Full-time position, worked on code generation project (HardTests) under the supervision of Prof. Lei Li

2022-2023 **Software Development Engineer (Full-time)**, Amazon

Full-time position, initialized and built the False-positive deployment detection mechanism, to detect short-lived deployment issues, hence preventing false passes in approval workflow for pre-production environments

2021 **Software Development Engineer Intern**, Amazon

3-months internship, trained models with over 95% accuracy for predicting the performance of advertisement keywords and created an API to expose data sources of keyword recommendations for performance analysis

2020-2021 **Digital Design & Verification Engineer**, Rambus

12-months internship, Worked on automating the workflow of design specifications and timing constraints verification for new serializer/deserializer (SerDes)

2019 **Undergraduate Research Assistant**, LegUp, University of Toronto

Summer Research Student, Worked on dynamic memory allocation with High-Level Synthesis for Field Programmable Gate Arrays (FPGAs) under the supervision of Prof. Jason Anderson

2018 **Lab Programmer**, Finn LAND Lab, University of Toronto

Part-time Work Study program, Worked on designing experiments, database management, and analysis on f-MRI imaging data, under the supervision of Prof. Amy Finn

Publications

Kexun Zhang, **Yee Man Choi**, Zhenqiao Song, Taiqi He, William Yang Wang, Lei Li

Hire a Linguist!: Learning Endangered Languages with In-Context Linguistic Descriptions.

In Association for Computational Linguistics(ACL) 2024

Responsible for the experiment and ablation studies for four of the eight low-resource languages studied, co-authored the relevant part in the manuscript.

Zhongmou He*, **Yee Man Choi***, Kexun Zhang*, Jiabao Ji, Junting Zhou, Dejia Xu, Ivan Bercovich, Aidan Zhang, Lei Li

HARDTESTGEN: A High-Quality RL Verifier Generation Pipeline for LLM Algorithmic Coding.

NeurIPS 2025 Deep Learning for Code Workshop (DL4C). To appear in ICLR 2026 * denotes equal contribution

Responsible for the rejection sampling and teacher distillation experiments and the post-training evaluations, co-authored the relevant part in the manuscript.

Yee Man Choi, Xuehang Guo, Yi R. (May) Fung, Qingyun Wang

CiteGuard: Faithful Citation Attribution for LLMs via Retrieval-Augmented Validation.

In review for ACL Rolling Review (ARR) Jan 2026.

Awards, Fellowships, & Grants

2025	Cohere Graduate Scholarship in Computer Science , University of Waterloo	\$ 10,000
2025	David R. Cheriton Graduate Scholarship , University of Waterloo	\$ 20,000
2025	Provost Doctoral Entrance Award for Women , University of Waterloo	\$ 5,000
2017-2022	Dean's List Scholar , University of Toronto	
2018	Shou-yun Pang Scholarship , University of Toronto	\$ 4,500
2017	Kowk Ming King Scholarship , University of Toronto	\$ 260,000

Teaching Experience

Wint '2026	CS 489/698 Introduction to Natural Language Processing , Teaching Assistant	UWaterloo
Wint '2026	CS 241 Foundations of Sequential Programs , Teaching Assistant	UWaterloo
Fall '2025	CS 246 Object-Oriented Software Development , Teaching Assistant	UWaterloo
Spr '2025	11868 Large Language Model Systems , Head Teaching Assistant	CMU

Mentoring & Outreach

2024-2025	Zhongmou He , Undergraduate Research Intern, Carnegie Mellon University
2018	Google IgniteCS Program , Volunteer Instructor

Other Relevant Research/Projects

Mechanism Design with Large Language Models

COLLABORATION WITH JIAYUAN LIU

Sep - Dec 2024

- Formulated mechanism design as a code generation task.
- Leveraged LLMs capabilities and evolution methods for better mechanism design.

Model-Based Evaluation for Code Summarization

COLLABORATION WITH ALEX XIE, BRIAN TUFTS AND DIVYANSHU SHETH

Jan - May, 2024

- Demonstrated the limitation of current n-gram based evaluation metrics for code summarization task
- Proposed the use of new metrics, including execution-based, model-based evaluation, and Minimum Bayes Risk(MBR) Quality Estimation

Developing Chatbot Personalities by Adjusting its tone of voice

ADVISED BY PROF. BAOCHUN LI, COLLABORATION WITH SOPHIE MAO AND TINA YANG

Sep 2021 - May 2022

- Trained text style transfer model using back-translation, to output different tones of speech as requested

Photo2Impressionism

COLLABORATION WITH SOPHIE MAO AND TINA YANG

Sep - Dec 2021

- Collected paintings from the impressionism style with filtering to remove sketches and keep only paintings
- Trained a model with CycleGAN to do the style transfer

Skills

PROGRAMMING LANGUAGES

- C, C++, Python, Rust, Java, CUDA

TOOLS AND FRAMEWORKS

- Pytorch, DeepSpeed, Git, Vim, Bash, SSH

LANGUAGES

- Cantonese (native), Mandarin(native), English(fluent), Japanese(beginner)