

YUK-KIT CHAN

contact@ucsd.edu ♦ [GitHub](#)

EDUCATION

University of California, San Diego, La Jolla, CA
Master of Science, Department of Computer Science and Engineering
Computer Science

Sep 2026 - June 2028 (Exp.)

The Chinese University of Hong Kong, Hong Kong
Bachelor of Engineering, Department of Computer Science and Engineering
Artificial Intelligence – Systems & Technologies
First Class Honours
ELITE Stream student
Minor in Mathematics

Aug 2022 - Jul 2026

RESEARCH EXPERIENCE

Research Assistant, The Chinese University of Hong Kong
Supervised under Professor Wenxuan Wang, Professor Michael R. Lyu
NoisyToolBench & AwN-framework [ENMLP'25 Main]

May 2024 - present

- Investigated the performance of LLM tool-use under imperfect and ambiguous user instructions.
- Designed and constructed NoisyToolBench, a challenging benchmark with noisy and ambiguous queries.
- Proposed the Ask-when-Needed (AwN) framework, which improves LLMs' ability to handle unclear instructions.
- Automatically Evaluated the tool-use performance of LLMs using ToolEvaluator

HalluHunter, Adversarial test-case generator [ACL'26 Findings]

- Developed HalluHunter, a fully automated framework to systematically elicit factual errors in LLMs using knowledge-graph-based fact triplets.
- Built rule-based NLP pipelines to generate diverse single- and multi-hop question templates from extracted triplets, enabling broad factuality coverage.
- Designed an iterative testing loop with adaptive triplet selection algorithm, prioritizing areas where models frequently err to amplify error discovery.

Fineva, LLM financial Benchmark [ACL ARR]

- Designed Fineva, a comprehensive benchmark of 25 tasks across 8 investment-banking departments and 7 evaluation dimensions to reflect real-world workflows.
- Curated novel datasets for portfolio optimization and regulatory compliance, capturing real-time data handling and decision-making under constraints.

Research Assistant, The Hong Kong University of Science and Technology
Supervised under Professor Jialun Cao

June 2024 - Dec 2024

How2Bench [ICML'26 Position Track]

- Proposed How2Bench, a 55-criteria checklist offering systematic guidelines for building high-quality, reliable, and reproducible code-related LLM benchmarks.
- Audited 672 benchmarks from the past decade, uncovering widespread issues in data quality, openness, and documentation.
- Identified critical flaws in prominent benchmarks, including duplicate samples, incorrect reference code and leaked sensitive/confidential data.
- Ran a human study with 49 participants, revealing gaps in awareness of data quality, reproducibility, and transparency practices.

PRESENTATION

Learning to Ask: When LLM Agents Meet Unclear Instruction. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, Online.

PUBLICATIONS

* Denote equal contribution.

Conference Papers

[P2] Jialun Cao, **Yuk-Kit Chan***, Zixuan Ling*, Wenxuan Wang, Shuqing Li, Mingwei Liu, Chaozheng Wang, Boxi Yu, Pinjia He, Shuai Wang, Zibin Zheng, Michael R. Lyu, Shing-Chi Cheung. 2026. Position: Code Benchmarks Should Prioritize Rigor, Reliability, and Reproducibility. In Proceedings of the Forty-third International Conference on Machine Learning, ICML 2026, 6-11 July 2026, Seoul, South Korea. PMLR 306.

[C1] Wenxuan Wang, **Yuk-Kit Chan***, Zixuan Ling*, SHI Juluan*, Youliang Yuan, Jen-tse Huang, Wenxiang Jiao, Zhaopeng Tu, Michael R. Lyu. 2026. Identifying the Achilles’ Heel: An Iterative Method for Uncovering Factual Errors in Large Language Models. In Findings of the Association for Computational Linguistics: ACL 2026, pages 34288–34309, San Diego, California, United States. Association for Computational Linguistics.

[C2] Wenxuan Wang, Shi Juluan*, Zixuan Ling*, **Yuk-Kit Chan***, Chaozheng Wang, Cheryl Lee, Youliang Yuan, Jen-tse Huang, Wenxiang Jiao, and Michael R. Lyu. 2025. Learning to Ask: When LLM Agents Meet Unclear Instruction. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 21784–21795, Suzhou, China. Association for Computational Linguistics

Preprint papers

[P1] Raco Lau, Xingyun Ma, Yuetong ZENG, Jingqi Yuan, Hongji Sun, **Yuk-Kit Chan**, Zixuan Ling, Jinquan Wang, Cheryl Lee, Wenxuan Wang, 2025. Fineva: Benchmarking LLMs on Investment Banking

TECHNICAL PROJECT

Internship in Legal Expression, Hong Kong June 2024 - Dec 2024
Legal Context Retrieval System

- Developed hierarchical document retrieval system improving legal query search accuracy.
- Built backend API to process user queries based on retrieval result and deliver context-aware legal advice.

Final Year Project, The Chinese University of Hong Kong, Hong Kong Sep 2025 - May 2026
Supervised under Professor Liwei Wang
X-CodeBench

- Developed a comprehensive code-related benchmarks, covering in total 2,517 questions in five dimensional tasks.
- Evaluate four SOTA models performance, uncovering the weakness of LLMs in programming.

SKILLS

Programming Languages	Python, C, Java, JavaScript, TypeScript, SQL
Languages	Cantonese (Native), Fluent English, Fluent Mandarin

HONORS AND AWARDS

Honours at Entrance, The Chinese University of Hong Kong	2022
Dean’s List, Department of Computer Science and Engineering	2022-2023, 2024-2025, 2025-2026
Admission Scholarship, Faculty of Engineering	2022-2026
ELITE Stream Scholarship, Faculty of Engineering	2022-2023, 2024-2025
Chong Kin Wo Scholarship, CW Chu College	2023-2024
CW Chu Foundation Scholarship, CW Chu College	2022-2023, 2024-2025