Contact: +85260527486 (Mobile) Email: ttchungac@connect.ust.hk Website: https://ttchungc.github.io

Tsz Ting, Chung

EDUCATION

The Hong Kong University of Science and Technology
Doctor of Philosophy in Computer Science and Engineering — 2021 - Present

The Chinese University of Hong Kong
Bachelor of Science (Hons) in Computer Science [1st Hons, ELITE Stream] — 2017 - 2021

WORKING EXPERIENCE

Pattern Recognition Center, WeChat AI, Tencent
Research Intern (AgentRL) — May 2025 - Present

Tencent AI Lab
Research Intern (token compression) — Nov 2023 - Sept 2024

Hospital Authority AI Lab
Research Assistant (search engines with QA system) — Jan 2021 - July 2021

Stanley Ho Big Data Decision Analytics Research Centre
Research Assistant (ASR model) — Jun 2020- Sept 2020

AWARDS & SCHOLARSHIP

2021-Now	**Hong Kong Ph.D. Fellowship, Hong Kong Research Grants Council**
2024-2025	**Outstanding PG Teaching Assistant Honorable Mention, HKUST**
2021-2022	**Professor Samuel Chanson Best PGTA Award, HKUST**
2021-2022	**RedBird Ph.D. Scholarship, HKUST**
2020-2021	**Dean's List Of The Engineering Faculty, CUHK**
2020-2021	**Silver Award For Outstanding Academic Performance, CUHK**
2018-2020	**ELITE Stream Student Scholarship, CUHK**

RESEARCH

On the Role of Chain of Thoughts in the Long In-Context Learning
Tsz Ting Chung, Lemao Liu, Mo Yu, Dit-Yan Yeung – *In Submission.*
- (Many-shot CoTs) It exhibits patterns that contrast with many-shot ICL and with previous findings in demonstration selection. Further experiments highlight the importance of model-data distribution alignment.

PRELUDE: A Benchmark Designed to Require Global Comprehension and Reasoning over Long Contexts
Mo Yu*, **Tsz Ting Chung***, Chulun Zhou*, Tong Li*, Rui Lu*, Jiangnan Li*, Liyan Xu*, Haoshu Lu, Ning Zhang, Jing Li, Jie Zhou – *In Submission.*
- (Measure of AGI) Introduce a long-context benchmark requiring global comprehension and deep reasoning. Experiments show ICL, RAG, SFT, and DeepResearch systems fall behind humans and even more in reasoning.

DivLogicEval: A Framework for Benchmarking Logical Reasoning Evaluation in Large Language Models
Tsz Ting Chung, Lemao Liu, Mo Yu, Dit-Yan Yeung – *EMNLP 2025 Findings.*
- (Logic Evaluation) Introduce a new benchmark to assess LLMs' logical reasoning while minimizing external influences, address data distribution bias, and propose a metric to reduce evaluation bias and uncertainty.

Unified Triplet-Level Granularity Hallucination Evaluation for Vision Language Models
Junjie Wu*, **Tsz Ting Chung***, Kai Chen*, Dit-Yan Yeung – *TMLR 2025.*
- (LVLM Hallucination) Introduce a new framework to evaluate LVLMs' hallucination on the triplet level, with a benchmark dataset for evaluation and a mitigation method proposed based on the paper's findings.

The Stochastic Parrot on LLMs Shoulder: A Summative Assessment of Physical Concept Understanding
Mo Yu*, Lemao Liu*, Junjie Wu*, **Tsz Ting Chung***, Shunchi Zhang*, Jiangnan Li, Dit-Yan Yeung, Jie Zhou – *NAACL 2025 (Oral).*
- (Measure of AGI) Investigate the stochastic parrot phenomenon and propose a task that alleviates the memorization issue via the usage of grid-format inputs that abstractly describe physical phenomena.

Selection-p: Self-Supervised Task-Agnostic Prompt Compression for Faithfulness and Transferability
Tsz Ting Chung, Leyang Cui, Lemao Liu, Xinting Huang, Shuming Shi, Dit-Yan Yeung – *EMNLP 2024 Findings.*
- (Token Compression) With simple tuning and small additional parameters, LLMs can achieve a better or similar level of performance in natural language understanding tasks with compressed demonstrations.