

Tsz Ting, Chung

EDUCATION		
	The Hong Kong University of Science and Technology	2021 - Present
	Doctor of Philosophy in Computer Science and Engineering	
THE CHINESE UNIVERSITY OF HONG KONG		
	Bachelor of Science (Hons) in Computer Science [1st Hons, ELITE Stream]	2017 - 2021
<hr/>		
WORKING EXPERIENCE	Pattern Recognition Center, WeChat AI, Tencent	May 2025 - Present
	Research Intern (AgentRL)	
	Tencent AI Lab	Nov 2023 - Sept 2024
	Research Intern (token compression)	
	Hospital Authority AI Lab	Jan 2021 - July 2021
	Research Assistant (search engines with QA system)	
	Stanley Ho Big Data Decision Analytics Research Centre	Jun 2020- Sept 2020
	Research Assistant (ASR model)	
<hr/>		
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
<hr/>		
RESEARCH	On the Role of Chain of Thoughts in the Long Test-time In-Context Learning	
	Tsz Ting Chung, Lemao Liu, Mo Yu, Dit-Yan Yeung – <i>In Submission</i> .	
	<ul style="list-style-type: none"> ○ (Many-shot CoTs) Explored the unique dynamic of CoT-ICL, reframing ICL as test-time learning by proposing the principle of comprehensibility and curricular smoothness and proposing a method for building effective CoT-ICL. 	
	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 – <i>In Submission</i> .	
	<ul style="list-style-type: none"> ○ (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 – <i>EMNLP 2025 Findings</i> .	
	<ul style="list-style-type: none"> ○ (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 – <i>TMLR 2025</i> .	
	<ul style="list-style-type: none"> ○ (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 – <i>NAACL 2025 (Oral)</i> .	
	<ul style="list-style-type: none"> ○ (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 – <i>EMNLP 2024 Findings</i> .	
	<ul style="list-style-type: none"> ○ (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. 	