Lunyiu Nie

lynie@utexas.edu | (737) 600−2774 | flitternie.github.io | [Google Scholar]

EDUCATION

The University of Texas at Austin, United States — Aug 2023 – Present

Ph.D. in Computer Science, Advisor: Professor Swarat Chaudhuri

- CGPA: 4.00 / 4.00

Tsinghua University, Beijing, China — Sep 2020 – Jul 2023

M.S. in Computer Science and Technology, Advisor: Professor Jidong Zhai and Professor Juanzi Li

- CGPA: 3.76 / 4.00

The Chinese University of Hong Kong, Hong Kong — Aug 2015 – Jul 2020

B.B.A. with Honours in Accountancy, with Computer Science as Second Major

- CGPA: 3.60 / 4.00, Computer Science GPA: 3.70 / 4.00

SELECTED PUBLICATIONS

[1] Online Cascade Learning for Efficient Inference over Streams.
Lunyiu Nie, Zhimin Ding, Erdong Hu, Christopher Jermaine, Swarat Chaudhuri
In *International Conference on Machine Learning* (ICML), 2024. [PDF]

[2] GraphQ IR: Unifying the Semantic Parsing of Graph Query Languages with One Intermediate Representation.
Lunyiu Nie, Shulin Cao, Jiaxin Shi, Qi Tian, Lei Hou, Juanzi Li, Jidong Zhai.
In *Conference on Empirical Methods in Natural Language Processing* (EMNLP), 2022. [PDF]

[3] Unveiling the Black Box of PLMs with Semantic Anchors: Towards Interpretable Neural Semantic Parsing.
Lunyiu Nie*, Jiuding Sun*, Yanling Wang, Lun Du, Lei Hou, Juanzi Li, Han Shi, Dongmei Zhang, Jidong Zhai.
In *AAAI Conference on Artificial Intelligence* (AAAI), 2023. [PDF]

[4] KQA Pro: A Dataset with Explicit Compositional Programs for Complex Question Answering over Knowledge Base.
Shulin Cao, Jiaxin Shi, Liangming Pan, **Lunyiu Nie**, Yutong Xiang, Lei Hou, Juanzi Li, Hanwang Zhang, Bin He.
In *Annual Meeting of the Association for Computational Linguistics* (ACL), 2022. [PDF]

[5] CoreGen: Contextualized Code Representation Learning for Commit Message Generation.
Lunyiu Nie, Cuiyun Gao, Zhicong Zhong, Wai Lam, Yang Liu, Zenglin Xu.
In *Neurocomputing*, 2021. [PDF]

[6] Code Structure Guided Transformer for Source Code Summarization.
Shuzheng Gao, Cuiyun Gao, Yulan He, Jichuan Zeng, **Lunyiu Nie**, Xin Xia, Michael Lyu.
In *ACM Transactions on Software Engineering and Methodology* (TOSEM), 2022. [PDF]

[7] ATOM: Commit Message Generation Based on Abstract Syntax Tree and Hybrid Reranking.
Shangqing Liu, Cuiyun Gao, Sen Chen, **Lunyiu Nie**, Yang Liu.
In *IEEE Transactions on Software Engineering* (TSE), 2020. [PDF]

SCHOLARSHIPS & AWARDS

- 2023, Outstanding Graduate & Outstanding Thesis Awards, Tsinghua University (Top **1** in CS Department)
- 2022, Grand-Prize Scholarship for Hong Kong, Macao, Taiwan, and Overseas Chinese Students, Ministry of Education (**Highest National Scholarship for Post-graduates**, **∼0.0002%**)
- 2022, Stars of Tomorrow (Internship Award of Excellence), Microsoft Research Asia
- 2018 ~ 2020, Faculty Dean's List, Chinese University of Hong Kong (Top **10%** in Faculty)
- 2018 ~ 2019, College Master's List, Chinese University of Hong Kong (Top **1** per major/class)

SELECTED RESEARCH PROJECTS

Cost-Efficient Large Model Programming & Optimization Framework 2023 – Present

- Proposed *Online Cascade Learning*, a no-regret online learning framework that addresses the cost-efficiency of LLM inference over streaming data, achieving up to 90% cost savings without sacrificing accuracy *(ICML 2024)*.
- Proposed *Online Program Synthesis*, a cost-aware large model programming and optimization framework that allows the coordination among vision models, large language models, and vision-language models for complex multi-modal reasoning, ensuring the computational resources are optimized for each subtask and user query *(Ongoing)*.

Unified Intermediate Representation for Graph Query Languages 2021 – 2023

- Proposed *GraphQ IR*, an intermediate representation that aims to bridge the semantic gap between natural language and graph query languages *(EMNLP 2022)*.
- Developed *GraphQ Trans*, the first source-to-source compiler in the field that supports translation among multiple graph query languages (SPARQL, Cypher, Gremlin, Lambda-DCS, KoPL, etc.).
- Achieved state-of-the-art performance on several semantic parsing benchmarks *Overnight*, *KQA Pro*, *GrailQA, and MetaQA-Cypher,* with promising generalizability (max. 11% accuracy improvement) under non-IID and low-resource (compositional generalization, cross-domain, and zero-shot) settings.

Automatic Code Performance Optimization with Deep Learning 2020 – 2021

- Built up a Python performance bug dataset by collecting the bug-fixing commits from open-source communities and developed a system for automatic performance downgrading detection with a 94.3% recall rate.
- Developed a code pretrained model that predicts compiler optimization heuristics and achieved state-of-the-art performance with at most 3.54× performance speedup on heterogeneous compute device mapping task.

Code Representation Learning for Commit Message Generation & Code Summarization 2018 – 2020

- Proposed *CoreGen*, a code representation learning method that exploits the contextual information behind the code changes with >28.18% BLEU-4 improvement on downstream commit message generation *(Neurocomputing)*.
- Developed *ATOM*, a commit message generation framework that explicitly incorporates the abstract syntax trees for representing the code structures and ensembles retrieval & generative methods via hybrid reranking *(TSE)*.
- Developed *SG-Trans*, a code summarization model that utilizes hierarchical code structures, i.e., local symbolic information and global syntactic structure (e.g., data flow), as the inductive biases in self-attention *(TOSEM)*.

Task-oriented Dialog Modelling with Unstructured Knowledge Access 2020

- Developed a task-oriented dialogue system that incorporates external unstructured knowledge for answering out-of-API-coverage user queries through knowledge retrieval and ensembled response generation.
- Outperformed the baseline with a 58.77% BLEU-4 score improvement, ranked as the **global finalist** in *the Ninth Dialog System Technology Challenge* (DSTC9) Track 1 with a paper presented in the *AAAI-21* workshop.

RESEARCH & INDUSTRY EXPERIENCE

Alibaba Cloud, Elastic High-performance Computing Group 2021 - 2022
Part-time Research Intern, Host: Dr. Xiangzheng Sun

Microsoft Research Asia, Software Analytics Group Jun–Sep 2022
Research Intern, Host: Dr. Yanlin Wang and Lun Du

CUHK, Text Mining Group & Human-Computer Communications Lab 2019 – 2020
Undergraduate Research Assistant, with Professor Wai Lam and Professor Helen M. Meng

Chinese Academy of Sciences, Institute of Computing Technology Jul–Aug 2018
Research Intern, with Professor Yungang Bao

ACADEMIC SERVICES

Served as reviewer for NeurIPS, ICLR, AISTATS, ACL Rolling Review (ACL & EMNLP), TOSEM.