Lin Long

kylin0long@gmail.com · (+86) 134-171-34341 · llong-cs.github.io/llongme

Education

Zhejiang University, Zhejiang, China 2020 – 2024

B.Eng. in Software Engineering, College of Computer Science and Technology.
Overall GPA: **3.95/4.00**; Third-year GPA: **4.00/4.00**; Ranking: **3/92**.

> Selected coursework: Probability Theory & Mathematical Statistics (95), Advanced Data Structure & Algorithm Analysis (94), Operating System (97), Object-Oriented Programming (96).

Advisors: Junbo (Jake) Zhao, Haobo Wang.

Internship

ByteDance Research, Shanghai, China 2024 – Present

Research intern at AI-Lab. Working on memory-enhanced **multimodal agents**.
Advisors: Yuan Lin, Hang Li.

Large Language Model · Multimodal Learning · Agent

Research Interest

- My research primarily focuses on **representation learning** and **natural language processing**, exploring how to improve model architectures and learning algorithms for enhanced data understanding.
- I am currently deeply engaged in **multimodal learning**, with a particular emphasis on **world understanding** through native sensory capabilities such as vision. My aspiration is to develop next-generation models that achieve seamless integration of knowledge across diverse modalities.

Preprints & Publications (with * denoting equal contribution)

Bridging the Semantic Gap Between Text and Table: A Case Study on NL2SQL

Lin Long*, *Xijun Gu**, *Xinjie Sun, Wentao Ye, Haobo Wang, Sai Wu, Gang Chen, Junbo Zhao*

ICLR 2025

Large Language Model · Multimodal Learning

tl;dr: A novel table-language multimodal framework that empowers LLMs with the ability to effectively and efficiently extract and reason over structure-enriched semantics from tabular data.

TableGPT2: A Large Multimodal Model with Tabular Data Integration

*TableGPT Team (as **directional lead of Table Encoder**)*

[report] [model]

Large Language Model · Multimodal Learning

tl;dr: An open-sourced advanced table expert designed to integrate and process tabular data directly and efficiently, overcoming the inherent limitations of current LLMs, especially towards production-level deployment.

On LLMs-Driven Synthetic Data Generation, Curation and Evaluation: A Survey

Lin Long, *Rui Wang, Ruixuan Xiao, Junbo Zhao, Xiao Ding, Gang Chen, Haobo Wang*

ACL 2024 (Findings)

Large Language Model · Data Synthesis · Weakly-supervised Learning

tl;dr: A survey on LLMs-driven synthetic data generation, curation and evaluation.

Positive-Unlabeled Learning by Latent Group-Aware Meta Disambiguation

Lin Long, Haobo Wang*, Zhijie Jiang, Lei Feng, Chang Yao, Gang Chen, Junbo Zhao*

CVPR 2024

Machine Learning | Weakly-supervised Learning

tl;dr: A novel Positive-Unlabeled (PU) learning framework that incorporates a hierarchical contrastive learning module to extract the underlying grouping semantics within PU data and iteratively distills the true labels of unlabeled data through meta-learning.

Property Existence Inference against Generative Models

*Lijin Wang, Jingjing Wang, Jie Wan, **Lin Long**, Ziqi Yang, Zhan Qin*

USENIX Security '24

Machine Learning | Security

♡ Honors and Awards

Outstanding Undergraduates of Zhejiang University	2024
Outstanding Bachelor's Thesis, Zhejiang University	2024
Xiaomi Scholarship	2023
Tencent Scholarship	2022

ℹ Miscellaneous

- Languages: English - Fluent, Mandarin - Native
- Programming Skills: Python, PyTorch, C, C++, LaTex, Java