

LIN LONG

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

🎓 EDUCATION

Zhejiang University, Zhejiang, China 2024 – Present

Master student in Computer Science, College of Computer Science and Technology.

Expected to withdraw.

Zhejiang University, Zhejiang, China 2020 – 2024

B.E. 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**.

Advisors: Junbo (Jake) Zhao, Haobo Wang.

🔍 RESEARCH EXPERIENCE

ByteDance Research, Shanghai, China 2024 – Present

Research scientist 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.

📄 TECHNICAL REPORT

TableGPT2: A Large Multimodal Model with Tabular Data Integration

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

[report] [7-B 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.

📄 PUBLICATION (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.

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

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

❖ MISCELLANEOUS

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