LIN LONG

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

🎓 EDUCATION

Zhejiang University (ZJU), Zhejiang, China   2024 – Present
Master student in Computer Science and Technology at the College of Computer Science and Technology.

Zhejiang University (ZJU), Zhejiang, China   2020 – 2024
B.E. in Software Engineering at the College of Computer Science and Technology.
Overall GPA: **3.95/4.00**, the third year GPA: **4.00/4.00**, rank: **3/92**.

👁 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 multimodal foundation models that achieve seamless integration and synthesis of knowledge across diverse modalities.

📄 PREPRINT

TableGPT2: A Large Multimodal Model with Tabular Data Integration

TableGPT Team (as directional lead of table encoder)

arXiv [cs.LG]

tl;dr: Introduction to TableGPT2, a model designed to integrate and process tabular data directly and efficiently, overcoming the inherent limitations of current LLMs, especially towards production-level deployment.

LLM Table Reading: Bridging the Semantic Gap Between Text and Table

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

ICLR 2025 (under review)

tl;dr: To bridge the gap between tabular and textual information, we propose TNT, a table-language multimodal model that empowers LLMs with the ability to effectively and efficiently extract and reason over structure-enriched semantics from tabular data.

📄 PUBLICATION (WITH * DENOTING EQUAL CONTRIBUTION)

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)

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

tl;dr: We propose LaGAM, a novel 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

♡ 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 speaker
- Programming Skills: Python, PyTorch, C, C++, LaTex, Java