Liming Lu

Email: liming_lu@sjtu.edu.cn | Tel: (86) 15868256485 | GitHub: Elemmire1.github.io

Education

Shanghai Jiao Tong University, Shanghai, China Sept. 2023 - Jun. 2027 (Expected)

B.S. in Computer Science, **John Hopcroft Class**

GPA: 90.2/100 (Major), 90.3/100 (Overall), **Zhiyuan Honorary Scholarship**

Zhiyuan Honors Program (Top 10%)

Publications

- **One Size Does Not Fit All: Token-Wise Adaptive Compression for KV Cache**
  Liming Lu, Kaixi Qiu, Jiayu Zhou, Jushi Kai, Haoyan Zhang, Jingzhi Wang, Bo Jiang, Jingwen Leng, Ziwei He, Zhouhan Lin.
  EMNLP 2026. `https://arxiv.org/abs/2603.04411`
- **Physis-Lang: Self-Evolving Language as a Physical Representation for Video World Model**
  Liming Lu†, Xianzheng Ma†, Wenkun He†, Guanqi Zhan†, Yilin Zhao, Junyu Chen, Mengyao Xu, Jiaojiao Fan, Wenhang Ge, Yuchao Gu, Yunze Liu, Boyi Li, Zhen Dong, Victor Prisacariu, Ming-Yu Liu, Song Han, Han Cai
  `https://physis-intelligence.github.io/physis-lang-web/`
- **Towards Compressive and Scalable Recurrent Memory**
  Yunchong Song, Jushi Kai, **Liming Lu**, Kaixi Qiu, Zhouhan Lin.
  `https://arxiv.org/abs/2602.11212`

Research Experiences

HAN Lab, MIT

Physis-Lang: Self-Evolving Language as a Physical Representation for Video World Model

Advisor: Prof. Song Han, Prof. Han Cai

- Developed a **self-evolving agent system** with video-to-text and text-to-video critics for post-training and inference-time guidance; consistently improves Wan and Cosmos across four physics benchmarks and outperforms **Veo 3.1**.
- Proposed a **language-centered physical world representation** providing scalable physical supervision for video world models.

LUMIA Lab, SJTU

One Size Does Not Fit All: Token-Wise Adaptive Compression for KV Cache

EMNLP 2026, Advisor: Prof. Zhouhan Lin

- Proposed **DynaKV**, a token-wise adaptive KV cache compression framework that dynamically allocates compression rates based on token importance, substantially outperforming PALU and MatryoshkaKV on long- and short-context benchmarks under aggressive compression.

Towards Compressive and Scalable Recurrent Memory

Advisor: Prof. Zhouhan Lin

- Proposed **Elastic Memory**, a HiPPO-grounded recurrent memory architecture that compresses history into a fixed-size state via optimal online approximation, achieving up to **16**× memory efficiency and consistently outperforming existing recurrent-memory baselines on 32K+ contexts.

One-for-all Intermittent Renewable Energy Forecasting with a Physics-guided Foundation Model

Submitted to ***Nature Energy***, Advisor: Prof. Zhouhan Lin

- Developed **IresFM**, a physics-guided foundation model with large-scale virtual-site pretraining and lightweight real-world adaptation, achieving near-state-of-the-art zero-shot wind/solar forecasting and a further **10%** error reduction after fine-tuning.

Skills

Programming: C++, Python, PyTorch, Transformers, Git, Linux, Hugging Face, LaTeX.

Mathematics: Linear Algebra, Mathematical Logic, Information Theory, Probability, Mathemetical Analysis.

Language: English - TOEFL: 106.