

Chen Yanxin | Curriculum Vitae

 <https://github.com/SSSSuperC>  yanxinchen26@m.fudan.edu.cn

 (+86) 13363979511  (+852) 61587613

Research Directions

Multimodal Large Models, Vision-Language Models (VLMs), Long Video Understanding, Retrieval-Augmented Generation (RAG), Efficient Training and Inference, Model Training Infrastructure

Education

Fudan University <i>Ph.D. in Artificial Intelligence</i>	Shanghai, China 2026.09 - Present
City University of Hong Kong <i>Master's in Data Science</i> GPA: 3.76/4.0	Hong Kong, China 2024.09 - 2025.10
Northwest University <i>Bachelor's in Software Engineering</i> GPA: 3.35/4.0 (top 20%)	Xi'an, Shaanxi, China 2020.09 - 2024.06

Awards & Honors

- **Outstanding Performance Award and Best Project Award**, City University of Hong Kong, 2025
- **University First-Class Scholarship (twice)**, Northwest University, 2022, 2023
- **University Second-Class Scholarship**, Northwest University, 2021

Research and Engineering Experience

Fudan University NLP Group: MOSS-VL Training and Infrastructure Optimization	<i>Internship</i>
○ 1K-GPU-scale training experience: Participated in large-scale MOSS-VL training, taking responsibility for four pretraining stages and the full SFT workflow, including ablation experiments and formal training runs. Tracked training stability, throughput, data anomalies, and framework issues in 1K-GPU-scale cluster environments, with hands-on experience in large-scale training debugging and execution. ○ VLM architecture exploration: Adopted cross-attention based vision-text interaction to reduce the computational burden of the original self-attention interaction, making the model more suitable for streaming video understanding and inference. Introduced a cross-gate mechanism to improve training stability. ○ Training infrastructure optimization: Adapted parallel strategies and activation recomputation for cross-attention VLMs in Megatron, and modified FA3 to support cross-attention masks, reducing memory usage and accelerating training. These optimizations helped extend the maximum number of trainable video frames and improve the efficiency of large-scale training.	
City University of Hong Kong AML Lab: Multimodal RAG and Long Video Understanding	<i>Research Experience</i>

- **Long-video RAG:** Studied probe-based retrieval and memory augmentation methods for long-video understanding, targeting context-length limitations and shallow retrieval. Designed probe generation, structured annotation, and re-ranking strategies to improve the localization of deep interaction semantics and cross-modal context.
- **Evaluation and analysis:** Designed LLM-as-Judge and human-evaluation comparison experiments for open-ended long-video understanding, and extended evaluation to benchmark settings such as LVBench and LongerBench to validate improvements in comprehension depth, information density, and trustworthiness.

Academic Publications

- **Co-first Author**, “ProEchoMem: Enhancing Long Video Understanding via Multi-Trace Probe-Echo Memory” *SIGIR '26*. CCF-A

Summary: This work addresses context limitations in long-video understanding and the difficulty of shallow similarity-based retrieval in capturing deep interaction semantics. It proposes a retrieval and memory augmentation framework based on multi-trace probe-echo memory, improving semantic localization and reasoning in ultra-long video scenarios through probe generation, structured memory organization, and re-ranking.

Technical Skills

- **Multimodal large models:** Familiar with VLM training workflows, video-text cross-modal interaction, cross-attention architectures, long-video context modeling, and streaming video understanding scenarios.
- **Training systems and infrastructure:** Experienced with PyTorch, Megatron-LM, and DeepSpeed, including 1K-GPU-scale multi-node training, hybrid parallelism, activation recomputation, memory optimization, throughput analysis, and distributed training debugging.
- **Efficient attention and kernel adaptation:** Experienced with cross-attention masks and FA3/FlashAttention adaptation, with engineering practice in extending trainable video frames and optimizing training bottlenecks.
- **Long-video RAG and evaluation:** Familiar with probe-based retrieval, structured memory, re-ranking, LLM-as-Judge, multi-model evaluation, and alignment with human evaluation, with the ability to complete the research loop from experimental design to paper writing.
- **Engineering implementation:** Python, C/C++, Linux; familiar with distributed training scripts, experiment orchestration, log analysis, and large-scale training failure diagnosis.