# Yuxuan Wang

(+86)18157865833 (WeChat) • yxwang1215@gmail.com

## Education

**Peking University** **Beijing, China**
*Incoming Ph.D. in School of Computer Science* *Expected Enrollment: Sep. 2027*
- **Advisor:** *Associate Professor Tong Yang*
- **Homepage:** https://yxwang1215.github.io/
- **Research Interest:** *Efficient Multimodal Large Language Models, focusing on inference acceleration, KV cache compression, and resource-efficient deep learning.*

**Xidian University** **Xi'an, China**
*B.E. in School of Computer Science and Technology* *Sep. 2023 – Jun. 2027*
- **GPA: 4.0/4.0 CET-6: 591/710 IELTS: 7.0/9.0**
- **Rank: 1/1342** *(Computer Category, 2023–2024)* **1/335** *(Software Engineering)*

**Zhejiang Zhenhai High School** **Ningbo, China**
*Top 10 High School in China, Innovation Class* *Sep. 2020 – Jun. 2023*
- *Admitted to the highly selective Innovation Class at Zhejiang's top high school, where roughly 15 members and 70 total students matriculate to Peking University or Tsinghua University each year.*

## Internships

**LEAP Lab, Department of Automation, Tsinghua University**
*Research Assistant* *Mar. 2025 – Sep. 2025*
- *Engineered robust algorithms for internal carotid artery ultrasound autonomous navigation, collaborating onsite at the Central Main Building.*
- *Co-developed UltraHiT and VA-Adapter architectures, directly leading to accepted papers at* ICRA 2026 *and* MICCAI 2026*.*

**EPIC Lab, School of Artificial Intelligence, Shanghai Jiao Tong University**
*Research Assistant* *Aug. 2025 – May 2026*
- *Spearheaded the design of efficient algorithms for large language diffusion and audio models, leveraging bidirectional attention and novel KV cache eviction techniques.*
- *Research efforts resulted in top-tier conference acceptances and submissions, including* ACM MM 2026, ICLR 2026, *and* NeurIPS 2026*.*

## Publications

1. **AudioKV: KV Cache Eviction in Efficient Large Audio Language Models**
   **Y. Wang**, P. He, X. Gui, X. Liu, J. He, X. Liu, X. Hu, L. Zhang
   - **First Author**; **Accepted to ACM MM 2026 (CCF-A)**.
   - *Proposed AudioKV, an audio-aware KV cache eviction framework tailored to long-context large audio language models, addressing the mismatch between conventional text-oriented compression and temporally continuous acoustic signals.*
   - *Designed semantic-acoustic alignment to identify audio-critical attention heads and perform head-aware KV retention, preserving semantically important acoustic evidence while removing redundant cached states.*
   - *Introduced Spectral Score Smoothing (SSS), which applies FFT-based filtering to stabilize noisy token-importance scores and retain local temporal continuity during cache eviction.*
   - *Reduced KV-cache memory overhead by 60% with minimal accuracy loss, substantially improving the efficiency of long-audio inference.*
2. **Accelerating Diffusion Large Language Models with SlowFast: The Three Golden Principles**
   Q. Wei, Y. Zhang, Z. Liu, P. Zeng, **Y. Wang**, D. Liu, L. Zhang
   - **Accepted** to **ICLR 2026 (CCF-A)**.
   - *Proposed SlowFast Sampling, a dynamic strategy that adaptively alternates between exploratory and accelerated stages based on token certainty, convergence, and positional principles.*

3. **UltraHiT: A Hierarchical Transformer Architecture for Generalizable Internal Carotid Artery Robotic Ultrasonography**
   T. Wang*, H. Jiang*, **Y. Wang***, Z. Sun, X. Yan, X. Li, G. Huang
   - **Co-First Author**; **Accepted** to **ICRA 2026 (CCF-B)** .
   - *Proposed* UltraHiT*, a hierarchical Transformer for autonomous robotic ultrasonography, significantly improving generalization across diverse internal carotid artery (ICA) anatomies.*
4. **VA-Adapter: Adapting Ultrasound Foundation Model to Echocardiography Probe Guidance**
   T. Wang, H. Jiang, **Y. Wang**, Z. Sun, S. Song, G. Huang
   - **Early Accepted** to **MICCAI 2026 (CCF-B, Top 9 %)**
   - *Developed parameter-efficient VA-Adapter, enabling ultrasound foundation models to infer 3D cardiac anatomy from vision-action sequences for precise probe guidance.*

   Yuxuan

## Preprints & Under Review

1. **AudioMarathon: A Comprehensive Benchmark for Long-Context Audio Understanding and Efficiency in Audio LLMs**
   P. He*, Z. Wen*, Y. Wang*, **Y. Wang**, X. Liu, J. Huang, Z. Lei, Z. Gu, X. Jin, J. Yang, *et al.*
   - *Introduced AudioMarathon, a long-form audio benchmark (up to 300s) evaluating LALMs across speech, music, and sound via complex multi-hop reasoning tasks.*
2. **AnchorKV: Anchor-Based Selection for KV Cache Compression on Multimodal Large Language Models**
   X. Gui, X. Liu, P. He, **Y. Wang**, X. Lin, X. Bu, X. Tu, J. Ren, X. Zhang, Y. Wang, L. Zhang
   - *Proposed AnchorKV, a plug-and-play KV compression method using anchor-based rescoring to enhance token selection and long-context MLLM performance on image/video tasks.*
3. **Thinking inside the Mask: In-place Prompting in Diffusion LLMs**
   X. Jin, **Y. Wang**, Y. Gao, Z. Wen, B. Qi, D. Liu, L. Zhang
   - *Proposed In-Place Prompting, a novel paradigm that integrates reasoning chains into the mask denoising process of Diffusion LLMs (dLLMs).*
4. **Self Speculative Decoding for Diffusion Large Language Models**
   Y. Gao*, Z. Ji*, **Y. Wang**, B. Qi, H. Xu, L. Zhang
   - *Developed Self Speculative Decoding (SSD) for Diffusion LLMs, accelerating inference via internal self-drafting and parallel verification without external draft models.*

## Selected Awards

**Huawei Scholarship** (less than 0.1%) *Dec. 2025*
*Issued by Huawei Xi'an Research Institute*
**National Scholarship** (Twice) *2023 – 2024, 2024 – 2025*
**National First Prize**, National English Competition (NECCS) *Apr. 2025*
**First Prize**, 16th National Undergraduate Mathematics Competition *Nov. 2024*
**First Prize**, National Undergraduate Mathematical Modeling Contest *Dec. 2024*
**National First Prize**, *Vocabulary Star* National English Vocabulary Competition *Jan. 2024*
**Honorable Mention** (Twice), MCM 2024, ICM 2026 *Jan. 2024, Jan. 2026*

## Activities

**Xidian Inspur Club**
*President* *2025 – 2026*
- **Club & Lab Management:** Directed daily operations and spearheaded a recruitment campaign onboarding 50+ new members. Organized 10+ academic workshops and orientations to expand and strengthen the undergraduate research community.