Minghui Guo

21 West Coast Crescent Singapore 128045   ✉ mguo77@u.nus.edu   ☎ +65 9807 3183

⌨ github.com/GuoMinghui07

EDUCATION

National University of Singapore   Aug 2026 - Jul 2030 (expected)
PhD in Computer Science (Incoming)

National University of Singapore   Aug 2025 - Jul 2026
MSc AI for Science
- GPA: 4.9/5
- **Core Courses:** Deep Learning, Machine Learning. Numerical Recipes. Complex Systems Analysis and Modelling.

Beijing Jiaotong University   Sep 2021 - Jul 2025
BEng Mechatronics Engineering
- GPA: 86.8/100
- **Core Courses:** Calculus (I) (98), Calculus (II) (98), Linear Algebra (94), Probability Theory and Mathematical Statistics (98), Engineering Physics (100), Circuits and Systems (98), Electronics (95), Programming for Engineers (92).

RESEARCH EXPERIENCE

NExT-GPT v2: A Unified Any-to-Any Multimodal LLM   Dec 2025 - Present
First Author, Supervisor: Dr. Hao Fei and Dr. Shengqiong Wu
- Led research on NExT-GPT v2, a unified any-to-any multimodal LLM that shares one backbone and modality-specific representation spaces jointly used for both understanding and generation.
- Built a hybrid AR + Flow-Matching architecture enabling arbitrary interleaved multimodal inputs/outputs (text, image, video, audio, 3D voxels) and modality translation.
- Enabled cross-modal reasoning over mixed-modal contexts, producing coherent and logically grounded multimodal outputs with seamless modality transitions and competitive generation quality.

Unified Video Understanding and Generation in a Shared Representation Space   Dec 2025 - Present
First Author, Supervisor: Dr. Hao Fei and Dr. Shengqiong Wu
- Proposed Video-RAE, a representation autoencoder framework with a temporally consistent video decoder that reconstructs videos directly from Qwen3VL semantic visual latents, mapping multimodal representations to pixel space while preserving motion continuity and fine-grained spatial details.
- Attached a Flow-Matching generation head to the MLLM backbone to enable video synthesis, leveraging shared representation spaces where perception signals guide generative dynamics.
- Proposed a roadmap for latent-space video reasoning, where the model performs multi-step inference by generating intermediate spatiotemporal latents for iterative understanding and refinement.

Benchmarking Unified Any-to-Any Interleaved Multimodal Learning   Aug 2025 - Nov 2025
Core Author, Supervisor: Dr. Hao Fei and Dr. Shengqiong Wu
- Built UNIM, a unified any-to-any interleaved multimodal benchmark (31,026 instances; 30 domains; 7 modalities incl. text/image/audio/video/document/code/3D) for interleaved multimodal reasoning and generation in open-form QA.
- Proposed the UNIM Evaluation Suite assessing Semantic Correctness & Generation Quality, Response Structure Integrity, and Interleaved Coherence for flexible multimodal I/O.
- Led development of UNIMA, an agentic any-to-any multimodal framework for unified understanding and generation over interleaved modalities, enabling structured cross-modal reasoning and semantically consistent outputs.
- Paper under submission to CVPR 2026.

Extraction and Reconstruction of Variable-Coefficient Governing Equations using Res-KAN Integrating Sparse Regression   Apr 2024 - Mar 2025   *reskan-master* ⌨
First Author, Supervisor: Prof. Xing Lü

- Developed a Res-KAN framework integrating Kolmogorov-Arnold Networks (KANs) with physics-informed neural networks (PINNs) to extract variable-coefficient governing equations from sparse, noisy data for AI4Science modeling.
- Designed an alternating optimization pipeline combining gradient-based training with group sparse regression to jointly identify PDE structures and variable-coefficient functions without prior knowledge.
- Validated on 14 nonlinear PDE benchmarks (e.g., KdV, Burgers, NLS) under sparse data and 10% noise, achieving accurate coefficient recovery and interpretable symbolic parameter extraction via spline-to-function mapping.

PUBLICATIONS

- *Video-RAE: Unified Video Understanding and Generation from Pretrained Semantic Representations.* Planned for submission to **ECCV 2026**.
  Minghui Guo, Hao Fei*, Shengqiong Wu.

- *Benchmarking Unified Any-to-Any Interleaved Multimodal Learning.* **CVPR 2026**, Under Review.
  Yanlin Li, Minghui Guo, Kaiwen Zhang, Shize Zhang, Yiran Zhao, Haodong Li, Congyue Zhou, Weijie Zheng, Yushen Yan, Shengqiong Wu, Wei Ji, Lei Cui, Furu Wei, Hao Fei*, Mong-Li Lee, Wynne Hsu.

- *Extraction and Reconstruction of Variable-Coefficient Governing Equations using Res-KAN Integrating Sparse Regression.* **Physica D: Nonlinear Phenomena**, **2025**. [Paper] ↗
  Minghui Guo, Xing Lü, Yuxi Jin.

INTERNSHIP EXPERIENCE

The Centre for Trusted Internet Community, NUS *Singapore*
Research Intern Aug 2025 – Present

- Conducted research on unified multimodal foundation models, developing benchmarks, any-to-any agents, and large multimodal models for interleaved understanding and generation.
- Explored unified video understanding, generation, and reasoning within a shared representation space; see Research Experience for selected projects and technical details.

The Future Laboratory, Tsinghua University *Beijing, China*
iSpace-Algorithm Development Intern Jun 2024 – Jun 2025

- Developed a script-driven interactive game agent powered by LLaMA 3 8B, enabling context-aware dialogue, narrative guidance, and character-consistent multi-turn interactions.
- Fine-tuned model with LoRA on 2×RTX 4090 (HuggingFace + PEFT) and integrated multilingual TTS (ElevenLabs) for real-time voice synthesis, enhancing storyline understanding and immersive gameplay.

The Institute of Computing Technology, Chinese Academy of Sciences *Beijing, China*
DSAI Intern Jan 2024 – Feb 2024

- Studied the Maximal Information Coefficient (MIC) and MINE framework, reproducing the pipeline with `python-minepy` and applying it to large-scale biomedical datasets ($\approx$90k samples) to identify nonlinear associations and benchmark noise robustness.

AWARDS

USA Mathematical Contest in Modeling | Meritorious Winner (Leader)	May 2024
National College Students Mathematical Contest in Modeling | Second Prize (Leader)	Oct 2023
Mathematical Contest in Modeling of Beijing Jiaotong University | First Prize (Leader)	Jun 2023
The 14th Chinese Mathematics Competition | First Prize	Jan 2023
The 33rd Beijing Mathematics Competition | First Prize	Jan 2023

TECHNOLOGIES

Languages: English (Proficient), Chinese (Native)

Technologies: Python, C, C++, Pytorch, Numpy, Pandas, MATLAB, LaTeX