Che Liu

che.liu21@imperial.ac.uk Github LinkedIn Google Scholar

Research Interest: My research focuses on multimodal learning, integrating diverse modalities such as 2D and 3D visual data, video, audio, natural language, and time-series signals. I am particularly interested in multimodal understanding and the development of unified models that support both perception and generation. My work combines large-scale pretraining with reinforcement learning to enable structured cross-modal reasoning, aiming to build scalable and generalizable systems for complex real-world tasks.

Education

Feb 2022 – 2026 (expected) **Ph.D. Imperial College London**
Research Topic: Multimodal Learning.
Supervisors: Dr Rossella Arcucci, Dr Wenjia Bai

Apr 2024 – Jun 2024 **Visiting PhD. Technical University of Munich**
Research Topic: Multimodal Learning.
Supervisor: Prof. Dr. Daniel Rückert

Sep 2019 – Sep 2021 **M.Sc. with Distinction, Swansea University**.
Major in Computational Mechanics.
(Formerly Erasmus Mundus Program)

Sep 2016 – Jul 2018 **B.Sc. Shanghai University of Engineering and Science**
Major in Automotive Engineering.

Experience

April 2025 – Present **Visiting Researcher – Embodied VLM,** X-humanoid, Remote.
· Leading the VLM post-training team, leveraging SFT and self-evolving RL across 3B to 72B parameter models on large-scale video datasets (30k+ hours) to advance reasoning capabilities for natural and embodied tasks.

May 2025 – Present **Visiting Researcher – OmniLLM,** StepFun, Remote.
· Advancing OmniLLM post-training via joint audio-visual-language SFT and RL, developing specialized data curation pipelines and omni-capability benchmarks to drive state-of-the-art Omni reasoning.

Nov 2024 – Apr 2025 **Research Intern – Unified Multimodal Vision Pretraining,** DAMO Academy, Beijing, China.
· Developed a unified vision pretraining framework using a single framework capable of processing 2D images, 3D volumetric data, and video across modalities.

Jul 2024 – Sep 2024 **Research Intern – Vision-Language Models,** AstraZeneca, Cambridge, UK.
· Investigated multimodal learning using purely synthetic data (image-text pairs) generated by diffusion models and large language models.
· Demonstrated that synthetic pretraining outperforms state-of-the-art models trained on real-world data.

Selected Publications

· Pelican-VL 1.0: A Foundation Brain Model for Embodied Intelligence
C. Liu, X-Humanoid, Technical Report
· Nexus-O: An Omni-Perceptive And -Interactive Model for Language, Audio, And Vision
C. Liu, *et al.* ACM Multimedia 2025
· Unified Visual Self-Supervised Pre-training Across Video, 2D, and 3D Vision
C. Liu, Alibaba Inc, DAMO Academy. Technical Report
· Step-Audio 2 Technical Report
StepFun Audio Team. Technical Report
· Beyond Distillation: Pushing the Limits of Medical LLM Reasoning with Minimalist Rule-Based RL
C. Liu, *et al.* Technical Report
· Investigating Vision-Language Model Architectures for 3D Volume Understanding
C. Liu, *et al.* ACL 2025 Findings
· Succeeding in Vision-Language Pretraining with Purely Synthetic Data
C. Liu, *et al.* ACL 2025 Findings
· Multimodal Time-Series Learning with Test-Time Enhancement
C. Liu, *et al.* ICML 2024
· Vision-Language Pretraining for Dense Visual Representation Learning
C. Liu, *et al.* NeurIPS 2024
· Learning Textual Hierarchies through Vision-Language Pretraining
C. Liu, *et al.* IEEE Transactions on Medical Imaging 2024

Collaboration and Supervision († - Equal contribution, ‡ - Supervision)

· Ariadne: A Controllable Framework for Probing and Extending VLM Reasoning Boundaries
Minghe Shen, Zhengzhong Tu, **Che Liu**‡, *et al.* Under Review
· Emergent Hierarchical Reasoning in LLMs through Reinforcement Learning
Haozhe Wang, **Che Liu**, Wenhu Chen, *et al.* Under Review
SuPreME: A Supervised Pre-training Framework for Multimodal Representation Learning
Mingsheng Cai, **Che Liu**‡, *et al.* EMNLP 2025 Findings
· NOVA: Benchmark for Anomaly Localization and Clinical Reasoning
Cosmin I. Bercea, **Che Liu**, *et al.* NeurIPS Dataset and Benchmark Track 2025 (**Oral**)
· SRPO: Enhancing Multimodal LLM Reasoning via Reflection-Aware Reinforcement Learning
Zhongwei Wan, **Che Liu**, Shen Yan, *et al.* NeurIPS 2025
· Unifying Cross-Lingual Vision-Language Pre-Training by Diminishing Bias
Zhongwei Wan†, **Che Liu**†, *et al.* NeurIPS 2023

Invited Talks

2025 **Incentivizing Medical LLM Reasoning Without Distillation** King's College London.
2024 **ECG Multimodal Learning** Peking University.
Synthetic Data for Medical Multimodal Learning AstraZeneca, Cambridge.
Multimodal Medical AI Stanford MedAI.
2023 **Language Bias in Medical Vision-Language Pretraining** Imperial College London.
Cross-lingual Medical Vision-Language Pretraining Peking University.
Latent Geometry Optimization in Medical Vision-Language Pretraining King's College London.

Awards and Achievements

2024 **Top Reviewer Award (Top 10%).** NeurIPS 2024.
Turing Scheme Funding. Imperial College London.
2022 **Google Research Grant.** Google Cloud Platform.
Engineering School Dean's Fund. Imperial College London.