

# SHUNIAN CHEN

+86 13588489297 [◇ 117010022@link.cuhk.edu.cn](mailto:117010022@link.cuhk.edu.cn) [◇ LinkedIn](#) [◇ GitHub](#)

## EDUCATION

### The Chinese University of Hong Kong, Shenzhen

Ph.D. in Computer Science

Shenzhen, CN

Aug 2024 - Now

GPA: 3.68

Courses: Large Language Models, Machine Learning, Algorithm Analysis, Software Engineering, Natural Language Processing

### Duke University

Durham, NC

Master of Engineering Management, Data Science and Machine Learning Track

Aug 2021 - May 2023

CGPA: 3.8/4.0, MGPA: 4.0/4.0

Courses: AI in Practice, Building Products Using Deep Learning, Probabilistic Machine Learning

### The Chinese University of Hong Kong, Shenzhen

Shenzhen, CN

Bachelor of Engineering; Major in Computer Science and Engineering; Minor in Statistics

Aug 2017 - May 2021

CGPA: 3.6/4.0, MGPA: 3.6/4.0, top 15%

Courses: Fundamentals of Machine Learning, Techniques for Data Mining, Time Series, Stochastic Process, Database System, Operating Systems, Computer Architecture, Data Structures, Optimization, Probability and Statistics

## RESEARCH EXPERIENCE

### The Chinese University of Hong Kong, Shenzhen (Advisor: Benyou Wang)

Shenzhen, CN

Bridging the Knowledge Modality Gap for Large Vision-Language Model

Under review

Co-first Author

- Defined and quantified the “knowledge modality gap” in LVLMs with new metrics and a benchmark, enabling cross-model diagnostics of vision-to-knowledge failures; designed probing experiments to reveal underlying mechanisms.
- Curated a cross-modal, knowledge-intensive dataset via distillation using Wikipedia as reference; introduced a contrastive loss for stronger knowledge–modality alignment.
- Applied GRPO to enforce a “look-then-think” strategy that grounds on the image first and reasons second; full method (distillation + contrastive + GRPO) achieved ~30% overall gain on our benchmark.

### The Chinese University of Hong Kong, Shenzhen (Advisor: Benyou Wang)

Shenzhen, CN

EvA: An Evidence-First Audio Understanding Paradigm for LALMs

Under review

Co-first Author

- Reached open-source SOTA on MMAU, MMAR, and MMSU with a dual-encoder “Evidence-First Audio (EvA)” architecture, yielding **+14 points** on MMAU Perception.
- Built and released EvA-Perception, a high-temporal-precision corpus that grounds reasoning in acoustic evidence and drives the largest perception-heavy gains.
- Engineered efficient training at scale: ~380 hours of audio fine-tuning; ~12h per stage on 8×A100 with frozen encoders + LoRA.

### The Chinese University of Hong Kong, Shenzhen (Advisor: Benyou Wang)

Shenzhen, CN

LongLLaVA: Scaling Multi-modal LLMs Efficiently via a Hybrid Architecture

**EMNLP 2025 Findings**

Second Author

- Developed a hybrid MLLM (LongLLaVA) combining Mamba efficiency with Transformer accuracy, improving multi-image performance while reducing GPU memory usage by > 80%.
- Optimized spatial-temporal data pipelines and progressive training to let a single A100 process **1000+ images**—~5× typical long-context vision baselines.

- Ranked top-3 across 12 video/image benchmarks while sustaining ~220 tokens/sec throughput; enabled scalable use cases in medical imaging and autonomy.

**The Chinese University of Hong Kong, Shenzhen (Advisor: Benyou Wang)**

Shenzhen, CN

ALLaVA: Harnessing GPT4V-synthesized Data for A Lite Vision-Language Model

Under review

Second Author

- Designed a “caption-first, then Q&A” distillation pipeline to improve synthesized data quality; open-sourced ALLaVA, then the largest synthetic multimodal dataset.
- Trained 3B-scale models on ALLaVA that matched or surpassed larger baselines.

**The Chinese University of Hong Kong, Shenzhen (Advisor: Benyou Wang)**

Shenzhen, CN

MLLM-Bench: Evaluating Multimodal LLMs with Per-sample Criteria

**NAACL 2025**

Co-first Author

- Proposed a capability taxonomy with six cognitive levels; released a high-quality VQA benchmark for the community.
- Evaluated 12 leading vision-language models under four settings; among the first to use GPT-4V as an automatic evaluator.

**The Chinese University of Hong Kong, Shenzhen (Advisor: Benyou Wang)**

Shenzhen, CN

MileBench: Benchmarking MLLMs in Long Context

**COLM2024**

Second Author

- Built a benchmark for multi-image capabilities of MLLMs, systematically measuring temporal and spatial understanding across multi-image tasks.
- Benchmarked open/closed-source models and identified key failure modes, informing future model design.

**The Chinese University of Hong Kong, Shenzhen (Advisor: Benyou Wang)**

Shenzhen, CN

Humans or LLMs as the Judge? A Study on Judgement Bias

**EMNLP 2024**

Co-first Author

- Proposed an innovative metric to measure the evaluators’ performance without ground truth; indicated the different **potential biases** of using humans and LLMs as judges.
- Validated the observed bias through manipulating the model, causing erroneous judgments by the models.

## PROJECT EXPERIENCE

**2023 Microsoft Imagine Cup**

Durham, NC

Autonomous Data Science Platform

Jan 2023 - Mar 2023

- Won the **5<sup>th</sup>** position out of 16 teams in the Americas World Finals, **2<sup>nd</sup>** position in Education Track.
- Designed and implemented a no-code platform using Python for data cleaning and ML model building.
- Incorporated Open AI APIs to allow users to control the tools in the platform with **natural language** and voice commands.

## TECHNICAL SKILLS

- Python (Numpy, Pandas, sklearn, Pytorch, Transformers, Accelerate), R, SQL, C/C++, MATLAB, Java, Linux
- Docker, AWS (Code Pipeline, ECS, etc.), Azure (Compute Engine, Container Registry, Databricks, etc.), GCP
- Jupyter, Markdown,  $\text{\LaTeX}$ , MS Office (Word, PowerPoint, Excel, Project, Visio)

## HONORS AND AWARDS

- 5<sup>th</sup> position in the Americas World Finals & 2<sup>nd</sup> position in Education Category, Microsoft Imagine Cup, 2023
- Bo Wen Scholarship, The Chinese University of Hong Kong, Shenzhen, for 4 consecutive years, 2017-2021
- Dean’s List Award, The Chinese University of Hong Kong, Shenzhen, for 4 consecutive years, 2017-2021