

Haochen Shi

Multimodal AI Research · Full-Stack Engineering · LLM Systems

PROFILE

Undergraduate CS student at The Hong Kong Polytechnic University (GPA 3.85, Dean's List) with research experience in **multimodal reasoning**, **vision-language systems**, and **LLM evaluation**. Currently leading an undergraduate research project on multimodal interaction and contributing to benchmark-oriented research in multimodal reasoning and academic survey generation.

EDUCATION

The Hong Kong Polytechnic University

Sep 2024 – Jun 2028

B.Sc. in Computer Science

GPA 3.85

Dean's List

IELTS 7.0

Year 2 of 4

SKILLS

Languages Python, TypeScript, Java, C++, Go, SQL

AI / ML

PyTorch, YOLOv8, LangChain, Qwen2.5-VL, OpenCV, LoRA

Frontend React, Next.js, Vue.js, React Native, Tailwind CSS

Backend

FastAPI, Flask, Spring Boot, Node.js, Docker

Databases PostgreSQL, MySQL, Redis, Supabase, ChromaDB

RESEARCH INTERESTS

Vision-Language Models · Multimodal Evaluation · Benchmark Design · Trustworthy LLM Systems · Human-Centered AI

RESEARCH

Multimodal LLM Interaction System for Home Environments | PolyU URIS Grant · Sole Investigator

Sep 2025 – Present

★ **URIS (Undergraduate Research and Innovation Scheme)**: University-wide competitive grant — only ~100 students selected across all departments; among these, only **5–6 CS students** were matched with a CS faculty supervisor. [Official approved project list ↗](#)

- Built a multimodal interaction prototype for ambiguous object reference resolution in home environments
- Implemented a lightweight-heavyweight pipeline: YOLOv8 + ByteTrack for continuous perception; Qwen2.5-VL invoked on-demand, reducing VLM invocation frequency through on-demand reasoning
- Developed an object registry with temporal memory to support cross-frame disambiguation
- Conducted preliminary case-based evaluation on ambiguous spatial queries; manuscript in preparation

SurveyLens · Discipline-Aware Benchmark for ASG Evaluation | Co-author (data pipeline, evaluation) · Under review

Sep 2025 – Present

review at KDD 2026 · Preprint available upon request

- Built SurveyLens-1k dataset from Semantic Scholar (keyword filtering → LLM classification → citation ranking), **1,000 from 3,000+** across 10 disciplines
- 3-stage cleaning pipeline (MinerU PDF parsing → rule filtering → LLM citation normalization), usability **62% → 94%**
- 8-dimension cross-disciplinary analysis across 10 fields; participated in evaluation of 11 SOTA ASG systems

CCD-VQA · Causal Reasoning Benchmark for Autonomous Driving | Co-author (annotation, QA generation) · Under review

Dec 2025 – Present

Under review at ACM MM 2026 · Preprint available upon request

- Built causal reasoning benchmark CCD-VQA from real accident videos; contributed to scene annotation and QA generation (199 scenarios / 1,194 QA pairs)
- Observed a gap between perception accuracy and causal reasoning performance across leading MLLMs in preliminary benchmark evaluation

★ FEATURED PROJECT

KEY PROJECT PolyInterview · AI Mock Interview Platform | University-Level Initiative · Supervised by Prof. Jiannong Cao (PolyU VP for Research & Innovation) · Student Helper

Jul 2025 – Sep 2025

Qwen3 (self-hosted) Azure Neural TTS Wav2Lip WebRTC Vue.js Flask Docker

- **University-endorsed project** directly supervised by the Vice President of Research & Innovation — among the highest-level endorsements available to an undergraduate student at PolyU
- Four-service AI interview platform (Flask API + Express auth + Vue frontend + LiveTalking avatar), deployed for PolyU use
- **Self-hosted Qwen3** on local GPU (no cloud LLM dependency); engineered prompts for multiple interview styles and KSA/STAR-based adaptive follow-ups
- Replaced offline EdgeTTS with **Azure Neural TTS (real-time streaming)**, cutting speech latency and enabling expressive prosody for a more natural interviewer persona
- Integrated LiveTalking + Wav2Lip lip-sync; built JWT auth, user management, and database modules
- Designed 3-tier evaluation (13 Features → 10 Aspects → Proficiency) with per-question feedback

OTHER PROJECTS

Teaching-Learning Assistant · AI Slide Explainer | Live Demo: [learn.shc66.com](#)

Dec 2025 – Jan 2026

Next.js Supabase Claude Vision API TypeScript

- PPT screenshots → Claude Vision API multimodal understanding → auto-generated explanations, with ROI selection and auto-quiz
- Next.js + Supabase full-stack, split-screen learning interface, used in course-related learning scenarios