

PROFILE

Undergraduate CS student at The Hong Kong Polytechnic University (GPA 3.81, Dean's List) with research experience in **vision-language grounding**, **LLM-based multi-agent systems**, and **benchmark design**. Two co-authored papers under review at ACL ARR; Kaggle Bronze on a Cornell Lab bioacoustic benchmark; sole student PI on a PolyU URIS-funded multimodal interaction project.

EDUCATION

The Hong Kong Polytechnic University Sep 2024 – Jun 2028
B.Sc. in Computer Science GPA 3.81 Dean's List IELTS 7.0 Year 2 of 4

SKILLS

Languages	Python, TypeScript, Java, C++, Go, SQL	AI / ML	PyTorch, Hugging Face, Qwen-VL, Qwen3, LoRA, DoRA, GRPO, CrewAI, LangChain
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

Embodied AI · Vision-Language Models · Computer Vision · LLM-Based Multi-Agent Systems · Benchmark Design

AWARDS & HONORS

Kaggle Bronze Medal — BirdCLEF+ 2026 | Cornell Lab of Ornithology · Google DeepMind · TU Chemnitz · Solo submission Jun 2026
• Ranked **263 / 4,094 teams (Top 6.4%)** on a featured-tier Kaggle competition on Brazilian Pantanal bioacoustic soundscape recognition (234 classes, macro-AUC, CPU-only inference); audited Google Perch V2 as a frozen backbone — documented as a CLEF 2026 Working Note ([code](#))

RESEARCH

Multimodal LLM-Based Interaction System for Home Robots with Real-Time Adaptation | PolyU URIS Sep 2025 – Present
Grant · Sole Student PI · Supervised by Prof. Wei Lou

★ **URIS** — PolyU's competitive university-wide undergraduate research grant. [Approved project list ↗](#)

- Designing a multimodal interaction system that fuses voice, gesture, and visual context for natural human-robot interaction in home environments, with LLM-driven dialogue and real-time adaptation
- Fine-tuned **Qwen2.5-VL-7B + DoRA** on RefCOCO grounding (6×Ascend 910B2, 8.1h): acc@0.5 **78.8% → 91.2%**, mIoU 80.5, 1.62 s/query — building the vision-language backbone for grounded reference resolution

SurveyLens · A Discipline-Aware Benchmark for Automatic Survey Generation | Co-author (dataset Sep 2025 – Present
construction & cleaning pipeline) · Under review at ACL ARR

- Built **SurveyLens-1k**: **1,000 human-written literature surveys evenly distributed across 10 disciplines**, curated from Semantic Scholar via keyword filtering → LLM discipline classification → citation ranking
- Designed a 3-stage cleaning pipeline (MinerU PDF parsing → rule filtering → LLM citation normalization) that lifted valid-sample retention **62% → 94%**

PersonalPlan · Planning Multi-Agent Systems for Personalized Programming Learning | Co-author Mar 2026 – Present
(evaluation runtime & reward-training validation) · Under review at ACL ARR

- Built the plan-to-CrewAI runtime mapper and a **4-check execution-verification evaluator** (instruction fidelity / workflow completeness / interaction quality / content-correctness & guidance-trace) used to gate plans into the MAP-PPL training set (3,043 query-profile-plan instances) via a 100-plan batch eval harness over a multi-language Docker sandbox
- Validated **GRPO Phase-1 reward training** on Qwen3-8B + LoRA (RTX 5090, 156 steps): Tier-1 mean reward **0.41 → 0.90**; diagnosed the Qwen3 /no_think completion-truncation failure mode

★ FEATURED PROJECT

DEPLOYED

PolyInterview · AI Mock Interview Platform | University-Level Initiative · Led Jul 2025 – Sep 2025 · In production
by **Prof. Jiannong Cao** (PolyU VP for Research & Innovation) · *Student Developer*
[Qwen3 \(self-hosted\)](#) [Azure Neural TTS](#) [Wav2Lip](#) [WebRTC](#) [Vue.js](#) [Flask](#) [Docker](#)

- University-endorsed initiative**; now formally deployed and in use by PolyU students
- Four-service stack (Flask API + Express auth + Vue frontend + LiveTalking digital-human avatar) with **self-hosted Qwen3** on local GPU; KSA / STAR-based adaptive follow-ups
- Integrated **Azure Neural TTS (streaming)** + LiveTalking / Wav2Lip lip-sync; built JWT auth, user and database modules, resume-refinement, and VLM-based video evaluation; designed 3-tier feedback rubric (Features → Aspects → Proficiency)