

# 石昊宸

多模态 AI 研究 · 全栈工程开发 · 大语言模型系统

✉ haochen8.shi@connect.polyu.hk

☎ +86 183-6718-2986

🌐 github.com/HaochenSHI66

🌐 shc66.com

📍 香港

## 个人简介

香港理工大学计算机科学本科生 (GPA 3.85, 院长嘉许名单), 研究方向为多模态推理、视觉语言系统与大模型评估。现作为负责人独立承担 URIS 本科生科研项目, 研究多模态人机交互; 同时以共同作者身份参与多模态推理基准及学术综述自动生成方向的论文工作。

## 教育背景

### 香港理工大学

计算机科学理学士

GPA 3.85

院长嘉许名单

雅思 7.0

大二在读, 四年制

2024年9月 - 2028年6月

## 技术技能

编程语言	Python、TypeScript、Java、C++、Go、SQL	AI / ML	PyTorch、YOLOv8、LangChain、Qwen2.5-VL、OpenCV、LoRA
前端	React、Next.js、Vue.js、React Native、Tailwind CSS	后端	FastAPI、Flask、Spring Boot、Node.js、Docker
数据库	PostgreSQL、MySQL、Redis、Supabase、ChromaDB		

## 研究方向

视觉语言模型 · 多模态评估 · 基准设计 · 可信 LLM 系统 · 以人为本的 AI

## 科研经历

### 家居环境多模态大模型交互系统 | 理大 URIS 资助项目 · 独立负责人

2025年9月 - 至今

★ URIS (本科生研究与创新计划): 全校竞争性资助, 全校仅约 100 名学生入选; 其中计算机系仅有 5-6 名学生获配 CS 导师。官方项目公示 ↗

- 构建家居环境中模糊物体指代消歧的多模态交互原型系统
- 采用两级感知流水线: YOLOv8 + ByteTrack 持续追踪目标, Qwen2.5-VL 按需触发, 降低重量级模型的调用开销
- 设计具有时序记忆的物体注册机制, 支持跨帧指代消解
- 针对模糊空间指代查询进行初步案例评估, 论文撰写中

### SurveyLens · 学科感知型学术综述自动生成评估基准 | 共同作者 (数据流水线、评估) · 投稿 KDD 2026 在审 · 预印本可 2025年9月 - 至今

#### 联系索取

- 基于 Semantic Scholar 构建 SurveyLens-1k 数据集, 经关键词筛选 → LLM 分类 → 引用排序三阶段流程, 从 3,000+ 篇论文中精选 1,000 篇, 覆盖 10 个学科
- 构建三阶段清洗流水线 (MinerU PDF 解析 → 规则过滤 → LLM 引用归一化), 数据可用率由 62% 提升至 94%
- 完成涵盖 10 个领域的 8 维跨学科分析; 参与 11 个前沿 ASG 系统的基准评测

### CCD-VQA · 自动驾驶因果推理视觉问答基准 | 共同作者 (标注、QA 生成) · 投稿 ACM MM 2026 在审 · 预印本可联系 2025年12月 - 至今

#### 索取

- 基于真实事故视频构建因果推理基准 CCD-VQA, 负责场景标注与 QA 生成 (199 场景 / 1,194 组问答)
- 在初步基准评测中, 发现主流多模态大模型在感知准确率与因果推理能力上存在显著差距

## ★ 重点项目

**核心项目** PolyInterview · AI 模拟面试平台 | 校级项目 · 指导教师: 曹建农教授 (理大科研及创新副校长) · 2025年7月 - 2025年9月

#### 学生助理

Qwen3 (自托管) Azure Neural TTS Wav2Lip WebRTC Vue.js Flask Docker

- 由理大科研及创新副校长曹建农教授督导, 项目已全校部署上线
- 基于四服务架构 (Flask API + Express 鉴权 + Vue 前端 + LiveTalking 虚拟形象) 构建 AI 面试平台, 已面向全校部署
- 本地 GPU 自托管 Qwen3, 不依赖云端 LLM 服务; 面向多种面试风格与 KSA/STAR 框架设计自适应追问提示词
- 以 Azure Neural TTS 替换离线 EdgeTTS, 实现实时流式语音合成, 降低延迟, 提升表达力与自然度
- 集成 LiveTalking + Wav2Lip 实现唇形同步; 完成 JWT 鉴权、用户管理与数据库模块开发
- 设计三层评估体系 (13 项特征 → 10 个维度 → 能力等级), 支持逐题反馈

## 其他项目

### Teaching-Learning Assistant · AI 课件讲解助手 | 在线演示: learn.shc66.com

2025年12月 - 2026年1月

Next.js Supabase Claude Vision API TypeScript

- 以 PPT 截图为输入, 经 Claude Vision API 多模态理解自动生成讲解文本, 支持 ROI 区域框选与自动出题
- Next.js + Supabase 全栈架构, 分屏学习界面, 已用于课程相关学习场景