

# Diyuan Deng

B.Sc. candidate, Data Science & Big Data Technology | Embodied AI, LLM inference, applied ML

diyuan Deng@link.cuhk.edu.cn | diyuan Deng.com

## Education

**The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen)** Shenzhen, China  
B.Sc., Data Science and Big Data Technology Sept. 2023 – Jun. 2027 (expected)

- GPA: 3.43 / 4.0.
- Coursework: Database Systems, Algorithm Design and Analysis, Stochastic Processes, Optimization, Machine Learning, Statistical Computing.

## Experience

**Data Engineering Intern** **Jun. – Aug. 2025**  
*Shanghai Wuzhi Evolution Technology Co., Ltd., Embodied Data R&D* Shanghai, China

- Collaborated with Prof. Bo Zhao's group (Shanghai Jiao Tong University) on embodied AI data and VLA models; built the xArm data pipeline and optimized VLA (10,000+ trajectories).
- LeRobot stack with multimodal LLMs and natural-language command interfaces; grasping and multi-layer stacking tasks at >90% success.

**Core Developer, Enterprise LLM Inference (SFGlang) [Capstone Project]** **Jan. 2026 – Present**  
*SF Technology Co., Ltd.*

- Improved throughput and communication for Qwen3-class LLMs and multimodal models (4B / 30B) under enterprise load using SGLang and dual-container isolation; on RTX 4090 / H20 clusters, tuned PD/EPD, TP/DP/EP, and FP8/INT8 quantization; built a load-test web platform; two patent applications in progress from core optimizations.
- Representative results: Qwen-30B with TP4+DP2 — TTFT -48.9%, +664% throughput; Qwen-4B with TP1+DP2 — +750% throughput.

**Research Assistant, NSC-Advisor (Ischemic Stroke Prognosis)** **Sept. – Dec. 2025**

- Interpretable favorable / unfavorable prognosis from admission data; related manuscript submitted to ISMRM (under review).
- Few-shot learning, prompting, and RAG for clinical evidence; dual-input UI prototype; 93.75% accuracy on 16 held-out real cases.

## Projects

**Library Management System | [GitHub Link](#)** **Sept. – Nov. 2024**  
*Developer*

- Full-stack Python (tkinter) + SQLite; four normalized tables, composite indexes, 15+ features (SHA-256 authentication), admin dashboard; tests covered 100% of requirements.

**2D RPG Combat (UE5 + C++) | [GitHub Link](#)** **Sept. – Dec. 2024**  
*Developer*

- C++ combat with three enemy AI types; Qt launcher with SHA-256 messaging; PaperZD animation state machines; team core codebase 1,000+ lines; best course project.

## Campus Roles & Honors

**Undergraduate Teaching Fellow (USTF), School of Data Science** Sept. 2024 – Present  
Data Structures, Databases, and one additional core course: 10+ tutorials for 800+ students; co-designed problems and grading; SDS Excellent USTF Award (2024–2025).

**Student Assistant, University Library** Nov. 2023 – Present  
Printers (Bambu Lab X1E, Ultimaker S5, Formlabs Form 2), 30+ jobs; Final Cut Pro (FCP) and Adobe Photoshop edits; 3D-printing workshops.

**Founder & Vice President, 2Tired Cycling Club** Jul. 2024 – Dec. 2025  
Core team of 20 across five functions; Songshan Lake and Tour de France meet-ups; first year 80+ members, ~95% satisfaction.

**Honors:** CUMCM Guangdong Third Prize (2025).  
Bronze, 1500 m (University Games).

Bronze, 100 m breaststroke (Sports Festival).

## **Skills & Interests**

---

**Languages:** Chinese (native); English (professional proficiency for coursework and work).

**Technical:** Python (advanced), SQL; C++, R; Vim; Linux; LLM serving (SGLang, parallelism, quantization); data engineering.

**Interests:** Cycling, photography, hiking, amateur radio, Model United Nations.