YITANG GAO

Hong Kong SAR | +852 5449-7757 | ygaocb@connect.ust.hk

EDUCATION

The Hong Kong University of Science and Technology (HKUST) — Hong Kong SAR
BEng in Computer Science — Sep 2022–Jun 2026 (expected)
CGA: **4.126/4.30**; Rank: **2/145**; Dean's List every term.
TOEFL iBT: **108**.
Selected Coursework: Artificial Intelligence; Machine Learning; Large-Scale ML for Foundation Models;
Big Data Mining/Management; Data Visualization; OOP & Data Structures; Algorithms; Operating Systems;
Software Engineering; Computer Organization; Linear Algebra; Probability and Statistics.

Cornell University — Ithaca, NY
Visiting Research Student, Database Systems Lab — Jun–Aug 2025

RESEARCH EXPERIENCE

HKUST LONG Group — Hong Kong SAR
Final Year Thesis (advisor: Prof. Long Chen) — Sep 2025–Present

- Investigate inference-time classifier-free guidance for autoregressive LMs.
- Design decoding/steering techniques to reduce hallucinations.
- Build long-context eval harnesses; plan guidance-scaling ablations.

Cornell University — Database Systems Lab — Ithaca, NY
Research Intern (advisor: Prof. Sainyam Galhotra) — Jun–Nov 2025

- *Auto Pruned Retriever* (VLDB submission, under review).
- Co-develop token/latency-efficient RAG via a persistent *reasoning subgraph*.
- Engineer edge-space retrieval + top-k frontier sentence encodings with top-m re-rank; *per abstract*: similar accuracy with $\sim 1/20$ LightRAG tokens and $\sim 1/3$ generation latency.
- Implement scalable causal discovery algorithm for the SEEKER project.

HKUST Accounting Department — Hong Kong SAR
Undergraduate Research Assistant (advisor: Prof. Yue Zheng) — Aug–Oct 2024

- Engineer OCR→DB pipeline for 657 noisy/low-res scanned complaints.
- Integrate Tesseract with heuristic post-processing to extract `[file]`-`[keyword]`-`[text]` triplets; reduce complexity from $O(n^4)$ to $O(n^2)$; speedup $\sim 20\times$ (days→hours).

HKUST NLP Group — Hong Kong SAR
Undergraduate Research Assistant (advisor: Prof. Junxian He) — Jan–May 2024

- *Long as a Tale: Unlocking Long-Video Context in VLMs with Hour-long Data.*
- Co-design and implement hour-long data pipeline: world-knowledge taxonomy (6 categories/37 sub-categories) → YouTube retrieval → three-stage filtering (metadata: 20–60 min, $\geq$100 likes/$\geq$1,000 views; subtitle density: $\geq$100 chars/min, English; motion: CLIP cosine non-similarity $\geq$0.8 for $\geq$50% frames); curate $\sim$4.5K long videos.
- Script subtitle→description conversion: chunking + metadata-aware LLM rephrasing; standardize prompts; produce $\sim$2.7K-token descriptions/video; ensure reproducibility.
- Set up long-context training/eval: text-only warm-up to 224K context (RoPE $\theta = 10^9$), 400-frame video training; automate V-NIAH, Video-MME, MLVU, LVBench runs; *paper outcomes*: +6.0 (Video-MME long), +5.8 (medium), +0.7 (short), +3.4 (MLVU), +4.6 (LVBench); robust retrieval up to 1,200 frames.

TEACHING EXPERIENCE

HKUST — COMP2211 *Exploring Artificial Intelligence* — Hong Kong SAR
Undergraduate Teaching Assistant — Jul–Aug 2024 & Jul–Aug 2025
- **Designed** labs with background notes, step-by-step instructions, reference solutions, and **autograders**; **held** office hours (∼200 students).

HKUST — COMP1023 *Introduction to Python* — Hong Kong SAR
Undergraduate Teaching Assistant — Sep–Dec 2024
- **Created** beginner-friendly labs with interactive UIs and **autograders**; **converted** coding tasks into small game-like projects (∼800 students/term).

HKUST — COMP2012 *OOP & Data Structures* — Hong Kong SAR
Undergraduate Teaching Assistant — Jan–May 2024
- **Led** lab sections (∼90 students per session), **demoed** data-structure implementations, and **answered** live Q&A.

PROJECTS & TRAINING

Independent Quantitative Research — Hong Kong SAR
Personal project — Jul 2024–Sep 2025
- **Built** an equity stack: ingest daily/hourly data (e.g., `yfinance`) into **SQLite** then **ClickHouse**; **implemented** factor mining and backtesting with Python+*PyO3*/Rust.
- **Researched** and **combined** published factors; **tested** equal-weight, market-neutral portfolios; **standardized** cleaning and universe rules.

LSE Summer School — London, UK
Fixed Income Securities & Debt Markets — Jul–Aug 2023
Fudan University Winter School — Shanghai, China
Probability, Statistics & Risk Management — Dec 2022–Jan 2023

HONORS & AWARDS

HKSAR Government Scholarship	2024–2026
Asia-Pacific Economic Cooperation Scholarship	2024–2026
Cornell IRIS Summer Scholarship	2025
HKUST SENG International Summer Research Sponsorship	2025
HKUST UROP Support Grant	2024
Extra Mile Charity Foundation Scholarship	2023–2024
University Scholarship Scheme for Continuing UG Students	2023–2024
HKUST Study Abroad Grant	2023
HKSAR Government Scholarship Fund — Talent Development Scholarship	2023–2024
Dean's List (every semester)	2022–2025

SKILLS & INTERESTS

Programming: Python, Java, C/C++, SQL, Bash, HTML, LaTeX
Libraries/Tools: PyTorch, Hugging Face, Pandas, OpenCV, Docker, Git, Linux
Research interests: Inference-time control of LMs; efficient & multimodal AI; long-context modeling; data mining
Interests: Piano (**Level 10** certificate), music, hiking/outdoors, cooking