

Tsz-To Wong

✉ tsztowong@gmail.com ☎ +886-918-046-668 🌐 aidenw.me 📞 aiden1020 📄 tsz-to-wong

EDUCATION

- **National Yang Ming Chiao Tung University** Hsinchu, Taiwan
M.S. in Computer Science GPA: 3.86/4.3 Academic Achievement Award (Fall 2024) Sep 2024 – Present
 - Thesis Research: Self-Generated Text Effects in Closed-Loop Streaming Video LLMs
Advisor: Prof. Ching-Chun Huang
 - Relevant Coursework: Compilers, Real-Time Operating Systems, Deep Learning, Parallel Programming
- **National Sun Yat-sen University** Kaohsiung, Taiwan
B.S. in Computer Science and Engineering Sep 2020 – Jun 2024

EXPERIENCE

- **National Yang Ming Chiao Tung University** Hsinchu, Taiwan
Research Assistant, NSTC Badminton Video Intelligence 2024 – 2025
 - **Reduced end-to-end processing time from 220s to 73s** by replacing static multi-GPU partitioning with a shared-queue dynamic scheduler that mitigated workload imbalance.
 - **Reduced time-to-first-output from 11s to under 1s** by implementing an asynchronous inference and output-streaming pipeline that decoupled model execution from downstream response handling.
 - **Enabled multiple project teams to reuse a common video-processing pipeline** by building shared infrastructure for highlight generation and tactical video QA.
 - **Improved multi-task QA accuracy from 73% to 85%** by integrating supervised fine-tuning with teacher-generated training data into the end-to-end training and inference pipeline.
 - Worked with professional coaches and partner teams to define video-QA requirements and presented the resulting system to NSTC reviewers and faculty.

PUBLICATIONS & MANUSCRIPTS

- **Tsz-To Wong**, Tao-Ting Chen, Shyang-En Weng, Hong-Han Shuai, and Ching-Chun Huang. “Annotation-Grounded Evaluation of Self-Generated Text Effects in Closed-Loop Streaming Video LLMs.” *Under review, ICLR 2027*.
- Yu-Heng Shih, Bing-Chen Wu, **Tsz-To Wong**, et al. “From Global Alignment to Local Grounding: Zero-Shot Chinese Character Recognition with Radical Verification.” *ACCV 2026 Oral*.
- **Tsz-To Wong**, Ching-Chun Huang, and Hong-Han Shuai. “COACH: Collaborative Agents for Contextual Highlighting — A Multi-Agent Framework for Sports Video Analysis.” *AAAI 2026 Workshop LaMAS*. 📄 arXiv:2512.01853

PROJECTS

- **LocalSubs — On-Device Real-Time Subtitle Translation** 📄 Web Store 📄 Tech Report
Go, Python, llama.cpp 2026
 - Built and shipped a fully on-device subtitle translation system with browser-side GPU inference and a Go native backend for CPU/GPU execution.
 - Improved translation quality to a **59.2% tie-excluded win rate** against Google Translate, within **5.4 points** of GPT-4o mini, by fine-tuning a context-aware **0.6B** model on a cleaned subtitle corpus.
 - Achieved a preliminary **1.63×** batch-size-one decode speedup on Apple M2 Metal by identifying vocabulary projection as the bottleneck and reducing the tokenizer vocabulary from 151k to 64k tokens.
- **Parallelizing Discrete PSO for the Travelling Salesman Problem** 📄 GitHub
C++, MPI, Slurm course project 2025
 - Built sequential and MPI-based dPSO implementations with a benchmarking harness for runtime, communication overhead, scaling efficiency, and solution quality.
 - Achieved **9.9×** speedup over sequential dPSO on a 2,319-city TSP instance by replacing globally synchronized execution with an asynchronous island-model migration scheme.

SKILLS

Programming: Python, C++, Go **Spoken:** Chinese, Cantonese, English

ML: PyTorch, Hugging Face Transformers, llama.cpp

Parallel & Systems: MPI, Docker