Sheng Wang

(+86) 158 2710 8517 | forence1999@outlook.com | Forence1999

Education

The University of Hong Kong PhD.Eng. • Computer Science *Hong Kong, China 2022.09 - 2025.11*

- Primary Supervisor: Prof. Chuan Wu | Co-Supervisor: Prof. Lingpeng Kong

Huazhong University of Science and Technology B.Eng. • AI & Automation *Wuhan, China 2017.09 - 2021.07*

- Grade: 91.5/100 | GPA: 3.95/4 | Rank: 2/28 (special class) | English: IELTS (7.0) | Monitor

Skills

Programming Language Python, Pytorch, Vue3, HTML, CSS, TypeScript, React, Supabase, Scikit-Learn, Tensorflow, Keras, C++, R, Matlab

Professional Software Claude Code, Antigravity, Cursor, VS Code, HBuildX, Perplexity, Gemini, Docker, Git, Anaconda, Stich, Figma

Research

Interests Agent, LLM Super-Alignment, Data Synthesis

Selected Publications *2023 - 2026*

- **[ICML2026 Submission - Co-Corresponding Author.]** Jingqi Zhou, Sheng Wang*, Dezhao Deng, Junwen Lu, Qintong Li, Jiahui Gao, Junwei Su, Hao Wu, Jiyue Jiang, Lingpeng Kong, and Chuan Wu*. **ToolSelf: Runtime Self-Reconfiguration for Robust Long-Horizon Agentic Tasks.**
- **[ICML2026 Submission.]** Fangrui Huang, Souhad Chbeir, Arpandeep Khatua, Sheng Wang, Sijun Tan, Kenan Ye, Lillian Annabelle Bailey, Merryn Daniel, Ryan Louie, Sanmi Koyejo, and Ehsan Adeli. **TherapyGym: Evaluating and Aligning Clinical Fidelity and Safety in Therapy Chatbots.**
- **[NIPS2025 Spotlight.]** Sheng Wang*, Pengan Chen*, Jingqi Zhou*, Qintong Li, Jingwei Dong, Jiahui Gao, Boyang Xue, Jiyue Jiang, Lingpeng Kong, and Chuan Wu. **TreeSynth: Synthesizing Diverse Data from Scratch via Tree-Guided Subspace Partitioning.** arXiv preprint arXiv:2503.17195, 2025.
- **[ICLR2025.]** Sheng Wang*, Liheng Chen*, Pengan CHEN, Jingwei Dong, Boyang XUE, Jiyue Jiang, Lingpeng Kong, and Chuan Wu. **MoS: Unleashing parameter efficiency of low-rank adaptation with mixture of shards.** In The Thirteenth International Conference on Learning Representations, 2025.
- **[EMNLP2025 Main.]** Juntao Zhao*, Wenhao Lu*, Sheng Wang*, Lingpeng Kong, and Chuan Wu. **QSpec: Speculative decoding with complementary quantization schemes.** arXiv preprint arXiv:2410.11305, 2024.
- **[EMNLP2025 Main - Co-Corresponding Author.]**Jingqi Zhou, Sheng Wang*, Jingwei Dong, Lei Li, Jiahui Gao, Lingpeng Kong, and Chuan Wu*. **ProReason: Multi-modal proactive reasoning with decoupled eyesight and wisdom.** arXiv preprint arXiv:2410.14138, 2024.
- **[ACL2024 Main.]** Sheng Wang, Boyang Xue, Jiacheng Ye, Jiyue Jiang, Liheng Chen, Lingpeng Kong, and Chuan Wu. **PRoLoRA: Partial rotation empowers more parameter-efficient LoRA.** In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 2829–2841.
- **[ACL2024 Findings.]** Sheng Wang*, Liheng Chen*, Jiyue Jiang, Boyang Xue, Lingpeng Kong, and Chuan Wu. **LoRA meets dropout under a unified framework.** In Findings of the Association for Computational Linguistics ACL 2024, pp. 1995–2008.
- **[ACL2023 Main.]** Jiyue Jiang, Sheng Wang, Qintong Li, Lingpeng Kong, and Chuan Wu. **A cognitive stimulation dialogue system with multi-source knowledge fusion for elders with cognitive impairment.** In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 10628– 10640.
- **[ICLR2025.]** Qintong Li, Jiahui Gao, Sheng Wang, Renjie Pi, Xueliang Zhao, Chuan Wu, Xin Jiang, Zhenguo Li, and Lingpeng Kong. **Forewarned is forearmed: Harnessing LLMs for data synthesis via failure-induced exploration.** In The Thirteenth International Conference on Learning Representations, 2025.
- **[ACL2025 Main.]** Boyang Xue, Fei Mi, Qi Zhu, Hongru Wang, Rui Wang, Sheng Wang, Erxin Yu, Xuming Hu, and Kam-Fai Wong. **UAlign: Leveraging Uncertainty Estimations for Factuality Alignment on Large Language Models.** In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 6002–6024.
- **[ACL2025 Findings.]** Boyang Xue, Hongru Wang, Rui Wang, Sheng Wang, Zezhong Wang, Yiming Du, Bin Liang, Wenxuan Zhang, and Kam-Fai Wong. **MlingConf: A Comprehensive Study of Multilingual Confidence Estimation on Large Language Models.** In Findings of the Association for Computational Linguistics: ACL 2025, pp. 2535–2556.
- **[EMNLP2025 Findings.]** Jiyue Jiang, Alfred Kar Yin Truong, Yanyu Chen, Qinghang Bao, Sheng Wang, Pengan Chen, Jiuming Wang, Lingpeng Kong, Yu Li, and Chuan Wu. **Developing and Utilizing a Large-Scale Cantonese Dataset for Multi-Tasking in Large Language Models.** In The 2025 Conference on Empirical Methods in Natural Language Processing.
- **[NAACL2025 Findings.]**Jiyue Jiang, Pengan Chen, Liheng Chen, Sheng Wang, Qinghang Bao, Lingpeng Kong, Yu Li, and Chuan Wu. **How well do llms handle cantonese? benchmarking cantonese capabilities of large language models.** The 2025 Annual Conference of the Nations of the Americas Chapter of the ACL.

Experience

Family Education LEADER *2025.02 - 2026.01*

- Extensively read books (spanning psychology, family education, team management, project management, product design, biographies, content operations, user operations, etc.), establishing a comprehensive business perspective.
- Assembled a cross-functional team covering full-stack development, algorithm research, product design, user operations, and content operations, further polishing my leadership qualities.
- Iteratively optimized multi-facet AI-native SOPs, covering market research, user modeling, product iteration, development collaboration, content production, user interviews, and community operations, leveraging cutting-edge AI products to multiply team collaboration and project management efficiency.
- Continuously iterated on user personas, journeys, pain points, needs, and solutions, successfully determined the product's market positioning and MVP, enhancing rapid iteration and optimization capabilities.
- Comprehensively enhanced full-stack development capabilities, deepened understanding and proficient application of future programming paradigms (*i.e.*, vibe coding).

Smart Speaker MAIN CONTRIBUTOR *2022.09 - 2024.09*

- Targeting the high resource consumption associated with multi-user personalization scenarios, introduce MoS and PRoLoRA algorithms, reducing the trainable parameters by times compared to vanilla LoRA.
- Targeting the under-performance of multi-modal tasks, propose the ProReason framework to decouple visual and reasoning capabilities and leverage LLMs to enhance reasoning, improving multi-modal performance.
- Propose the QSpec technique to share quantized weights and KV Cache in a speculative decoding framework, achieving lossless inference acceleration in low-resource scenarios (e.g., edge devices).
- Construct Cantonese evaluation datasets for common knowledge, factual generation and complex reasoning tasks, and compare the performance of 35 open-source and closed-source LLMs.
- Extract audio from 1939 hours of dysarthric video, convert it into transcripts with offline automatic speech recognition (ASR) services, and finalize it through manual revision.
- Implement MoChA and Whisper models for end-to-end ASR, design enhancement modules for dysarthric speech, expand vocabulary, and post-train models for improved Cantonese ASR performance.
- Train a Transformer-based acoustic model and HiFi-GAN vocoder for text-to-speech (TTS) based on the "aidatatang_200zh" corpus, and inject customized voice characteristics into it with a speaker encoder.

Smart Robotic Walker - Sound Source Localization ONLY MAJOR CONTRIBUTOR *2021.09 - 2022.12*

- Set up 14 sound source positions at the 4-th floor of Chow Yei Ching Building in the HKU, and collect 1984 pieces of multi-channel microphone data, totaling about 500 minutes. Process it with voice signal framing, windowing, filtering, normalization, and other preprocessing techniques.
- Investigate and test traditional and learning-based noise suppression and speech separation algorithms, and finally select the NSNet2 model for noise reduction of the preprocessed speech signals.
- Extract GCC-PHAT and MFCC features, and innovatively design an ensemble learning-inspired parallel module to enhance RD3Net for sound source localization (SSL), improving the accuracy from 88.2% to 93.6%. In real-world scenarios, combined with the above noise suppression module, RD3Net performs satisfactorily, facilitating basic indoor voice navigation.
- Utilize A2C and D3QN reinforcement learning algorithms to fine-tune the model online, and obtain the expected effect in both the simulation and real environment.
- Design a low-computation traditional SSL algorithm based on TDOA, accurately detecting near-field sound source positions for user fall alerts.

Mathematical Contest in Modeling TEAM LEADER *2020.03 - 2020.04*

- Learn data mining and mathematical modeling methods, deepening the understanding of data processing and scenario modeling.
- Analyze the rating and review data of products on Amazon, identify key patterns and relationships, design indicators, explore potential function designs, and formulate e-commerce sales strategies to promote product reputation and sales.
- Organize our team for efficient preparation, assign tasks reasonably, collect literature, design the modeling scheme, and write the final paper.

AI Summer Experience in National University of Singapore MAIN CONTRIBUTER *2019.07 - 2019.07*

- Collaborate to complete the "Heaven's Scrutiny" project, a basic criminal arrest system based on face recognition and skeleton detection.
- Design feature extraction module, optimize AlphaPose algorithm, and integrate all the modules of team members.
- Rank the first in the course evaluation system, be recognized by the judge panel, and win the double titles of "Best Team" and "Best Individual".
- Complete high-intensity learning tasks and projects, cultivating the team spirit and the ability to work under high pressure.

Monitor *2017.09 - 2020.07*

- Organize various activities of our class, boosting the ability of organization and coordination.
- Communicate with teachers and classmates proactively, improving communication skills.
- Handle multiple affairs of class and individual at the same time, cultivating self-management ability.

Honours

Honours Degree (9/300) & Outstanding Graduate, Huazhong University of Science and Technology	*Wuhan*	2021
National Scholarship & Xiaomi Scholarship & Merit Student	*Wuhan*	2020
Meritorious Winner, Mathematical Contest in Modeling	*Wuhan*	2020
Best Team & Best Individual, AI Summer Experience in National University of Singapore	*Singapore*	2019
Samsung Scholarship & Merit Student	*Wuhan*	2019
National Encouragement Scholarship & Model Freshman of Academic Records	*Wuhan*	2018