YURUN SONG

PhD ~ University of California, Irvine

rain9876.github.io
songyr888@gmail.com
(949) 527 8385
github.com/Rain9876
Irvine, USA
/in/yurunsong9876

SUMMARY

Yurun Song is a 4th year PhD student at the University of California, Irvine, focusing on developing efficient AI-driven products for LLM training and inference. He holds an MSc in Advanced Computing from Imperial College London. His research interests include Parameter Efficient Finetuning, Distributed LLM Deployment, and Multi-turn Agentic Reinforcement Learning. He is supervised by Ian G. Harris at UC, Irvine.

SKILLS

Languages: Python, Java, C++, Matlab, SQL, C.

Technologies: Pytorch, Deepspeed, PEFT, ONNX, RAG, Android, Flask

EXPERIENCE

06/2025 - 09/2025 **Graphical User Interface Agent (GUI Agent) For Mobile App** — **Hithink Financial Inc.**
Research on the application of GUI Agent based on multimodal models in mobile app automated testing. By applying GRPO training, the model can understand screenshots, actions, and expected outcomes, it enables automated app interaction and builds an end-to-end testing pipeline, including automatically generating test cases, executing tests, and providing feedback for refinement.

05/2024 - 09/2024 **Portrait Video Generation** — **UC, Irvine**
Guided by Prof. Simon Penny. Research related to utilize LLMs to produce stylish educational presentations with AI-generated portrait videos of specific individuals, combining both video and voice synthesis with NLP for educational content creation. Accepted for exhibition at the IJCAI 2025 Art Gallery.
`https://simonpenny.net/works/AISelfPortrait.html`

02/2024 - 07/2024 **QuArch Dataset for Computer Architecture** — **Harvard**
Guided by Prof. Vijay Janapa Reddi. Developed the QuArch dataset and utilized the Axolotl Docker environment for training LLMs on hardware knowledge. Conducted comprehensive performance evaluations using the QuArch benchmark to optimize model accuracy and efficiency
`https://quarch.ai/`

12/2022 - 06/2023 **5G CoreNetwork with ChatGPT** — **Blackberry**
Research related to utilize LLM for the purpose of generating code, managing and executing queries in the 5G Core network.
`https://rain9876.github.io/courses/5G-Simulation-with-ChatGPT/5G_Simulation.pdf`

12/2021 - 04/2022 **Multi-Intents-Detection for SMS Message** — **Aktify**
Design a Multi-intents Detection Model for Aktify Company. Using their conversation data and classify customs' intents behind. The project is accepted by EMNLP conference in 2023.
`https://aclanthology.org/2023.findings-emnlp.636.pdf`

12/2020 - 11/2021 **Python Automatic Code Generation** — **UC, Irvine**
Supervised by Prof. Ian. G. Harris, my team members are developing a new AI model to translate Natural Language to Python Code. And the work is accepted by EACL workshop in 2022.
`https://aclanthology.org/2023.eacl-srw.4.pdf`

12/2020 - 03/2021 **Document Summarization Evalution** — **Google**
Guided by Dr. Kishore Papineni, I and my partner research on a new topic about document level MT evaluation metric. We focus on the combination of semantic and cross-lingual document metric.

02/2021 - 05/2021 **Sentence-level Semantic Evalution** — **Imperial College, UK**
Mentored by Prof. Lucia Specia, we worked on her summer proposal about developing an evaluation metric in Machine Translation. The project is accepted by NACCL-HIT conference in 2023.
`https://aclanthology.org/2021.naacl-main.252.pdf`

EDUCATION

09/2022- 06/2026 **Computer Science (PhD.)** — **University of California, Irvine**
Focus on research: Parameter Efficient Finetuning, Multimodal Training, Multi-intent detection
GPA: 3.95

09/2019 - 11/2020 **Advanced Computing (MSc.)** — **Imperial College London, UK**
Focus on Courses: Machine learning for Imaging, Natural Language Processing, Computer Vision
GPA: 3.7 (Upper Second Class Honours)

09/2016 - 06/2019 **Computer Science (BSc.)** **King's College London, UK**
Focus on Courses: Software Engineering Group Project, Distributed system, Artificial Intelligence
GPA: 4.0 (First Class Honours)

PUBLICATION

Aug 2025 **AMAQ: Adaptive Mixed-bit Activation Quantization for Collaborative Parameter Efficient Fine-tuning**
- Yurun Song, Zhouyi Yang, Ian G. Harris, Sangeetha Abdu Jyothi
- `https://arxiv.org/pdf/2510.05468`

Jun 2024 **ShareLoRA: Parameter Efficient and Robust Large Language Model Fine-tuning via Shared Low-Rank Adaptation**
- Yurun Song*, Junchen Zhao* Ian G. Harris, Sangeetha Abdu Jyothi
- *Under review 2025 AAAI*
- `https://arxiv.org/pdf/2406.10785`

Dec 2023 **LinguaLinked: A Distributed Large Language Model Inference System for Mobile Devices**
- Junchen Zhao*, Yurun Song*, Simeng Liu*, Ian G. Harris, Sangeetha Abdu Jyothi
- *2024 ACL System Demonstration Track*
- `https://aclanthology.org/2024.acl-demos.16.pdf`

Jun 2023 **PCMID: Multi-Intent Detection through Supervised Prototypical Contrastive Learning**
- Yurun Song*, Junchen Zhao*, Ian G. Harris, Spencer Koehler, Amirali Abdullah
- *Findings of the 2023 Conference on Empirical Methods in Natural Language Processing*
- `https://aclanthology.org/2023.findings-emnlp.636.pdf`

Jan 2022 **GAP-Gen: Guided Automatic Python Code Generation**
- Junchen Zhao*,Yurun Song*, Ian. G. Harris
- *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*
- `https://aclanthology.org/2023.eacl-srw.4.pdf`

Jun 2021 **SentSim: Crosslingual Semantic Evaluation of Machine Translation**
- Yurun Song*, Junchen Zhao*, Lucia Specia
- *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics*
- `https://aclanthology.org/2021.naacl-main.252.pdf`

TEACHING EXPERIENCES

Fall 2022 **CS 268: Introducation to Optimization**

Fall 2023 **ICS 53: Principles in System Design**

Winter 2024 **CS 145: Embedded Software**

Fall 2024 **CS 171: Introduction to Artificial Intelligence**

AWARDS AND HONORS

2023 **UCI Dean's Award Scholarship** **UC, Irvine**
Awarded a one-year $10,000 Dean's Award in recognition of exceptional research potential.