

Qinchen Wu

📍 Shanghai, China ✉ qinchen.wu62@gmail.com ☎ 18964321758 🌐 Personal Website 🐙 GitHub

Education

National University of Singapore <i>MS in Computer Engineering</i>	<i>Aug 2023 – Jan 2025</i>
◦ GPA: 4.6/5.0	
Nanjing University of Aeronautics and Astronautics <i>BS in Electrical Engineering and Automation</i>	<i>Aug 2018 – Jun 2022</i>
◦ GPA: 3.8/5.0	

Research Interest

Vision and Language, LLM Agents (GUI agents), Video understanding, Code Generation

Publications

[1] GUI-Narrator: Detecting and Captioning Computer GUI Actions Qinchen Wu, Difei Gao, Lin Qinghong, Zhuoyu Wu, Mike Zheng Shou.	ACMMM 2025
[2] VideoGUI: A Benchmark for GUI Automation from Instructional Videos Lin Kevin Qinghong, Linjie Li, Difei Gao, Qinchen Wu , Mingyi Yan, Zhengyuan Yang, Lijuan Wang, and Mike Zheng Shou. VideoGUI	NeurIPS 2024 DB
[3] Assistgui: Task-oriented desktop graphical user interface automation Gao D, Ji L, Bai Z, Ouyang M, Li P, Mao D, Wu Q , Zhang W, Wang P, Guo X, Wang H., Mike Z. AssistGUI	CVPR 2024
[4] Harmonizing Unets: Attention Fusion module in cascaded-Unets for low-quality OCT image fluid segmentation Zhuoyu Wu, Qinchen Wu , Wenqi Fang, Wenhui Ou, Qianjun Wang, Linde Zhang, Chao Chen, Zheng Wang Harmonizing Unets	Computers in Biology and Medicine

Research Experience

Detecting and Captioning GUI actions in Videos <i>Showlab, National University of Singapore</i>	<i>Jan 2024 – Apr 2025</i>
◦ Designing pipelines for collecting users' actions from GUI environments	
◦ Integrating temporal grounding and ROI-aware mechanisms to enhance VLMs' action understanding and computational efficiency.	
◦ Develop a benchmark standard for GUI-centric-action evaluation based on LLMs. (GPT-4o).	
◦ Accepted by ACM MM 2025.	
Developing AI Assistant For GUI Interface Automation <i>Showlab, National University of Singapore</i>	<i>Aug 2023 – Mar 2024</i>
◦ Propose a new baseline for GUI-workflow automation. Including Agent planning, Agent acting, UI-Parsing, and Agent-Critic.	
◦ Excavate the In Context Learning ability by Prompt engineering on LLMs (GPT, GLM, LLaMA) to generate concise and fine-grained plans from human demonstration videos.	
◦ Empowering the Model with RAG to alleviate LLMs' hallucinations of Task planning, build a more robust planning method by langchain and LLMs.	
◦ Accepted as poster for CVPR 2024.	

Attention Fusion-based Attention U-Net for OCT Image Segmentation.

Jan 2023 – March 2024

Durham Univeristy

- Proposed a cascaded U-Net framework for low-resolution image segmentation, achieving improved Dice scores on three datasets.
- Introduced and experimentally validated the effectiveness of Adaptive Attention Fusion (Channel Fusion, Spatial Fusion).
- Accepted by Computers in Biology and Medicine.

Professional Experience

LLM algorithm Engineer

Apr 2025 – Present

PDD, TEMU

- Pretrained a 1.5B-parameter LLM from scratch using the Megatron-LM framework.
- Curated high-quality data focused on mathematics and chain-of-thought reasoning to enhance performance on math-related tasks.
- Optimized the model architecture via Multi-Token-Prediction (MTP) and partial RoPE (0.75) and Multi-latent attention for improved performance while maintaining high efficiency during inference.

Multimodal Algorithm Engineer

July 2024 – Oct 2024

Tencent WeChat, Intern

- Improve the online performance of Matching model based on Sentence Bert. Includes optimization in Data, Feature, Model Config and Loss function. Improve Precision from 75% to 81%.
- Upgraded the model to a multimodal Mixture of Experts (MOE) framework, boosting offline precision performance by leveraging visual modalities. enhancing the model's ability to understand and describe dynamic visual content.
- SFT LLMs (Qwen2) as Baseline. Enabling the model to extract key elements from long-context inputs. Boosting understanding performance in live-stream environments.

Awards

Third Prize, Jiangsu Aircraft Control Innovation Competition

Jun 2021

Third Place Scholarship Award, Nanjing University of Aeronautics and Astronautics

Sep 2020

Second Place Scholarship Award, Nanjing University of Aeronautics and Astronautics

Mar 2019

Skills

Languages & Tools: Python, C, Matlab, Spark, SQL

Computer Vision: OCT Segmentation, SAM, SIFT, Opencv

NLP & LLM: BERT, multimodal (llaVa, Qwen-VL, MMF), Retrieval Augmentation Generation

Machine Learning: K-means, Markov random field

Framework: Megatron LM, Deepspeed