Xueren Ge

•*Tel: +1 470-662-2854* •*Email: zar8jw@virginia.edu* •*Web: https://xueren-ge.github.io/*

EDUCATION BACKGROUND

University of Virginia *May. 2022 – Present*

Major: Computer Engineering Degree: Doctor of Philosophy **GPA**: 3.88/4.0

Courses: Reinforcement Learning, Benefits and risks of Large Language Models, Learning for Interactive Robotics

Georgia Institute of Technology *Aug. 2020 – May. 2022*

Major: Electrical and Computer Engineering Degree: Master of Science **GPA**: 3.91/4.0

Courses: Statistical Machine Learning, Natural Language Processing, Deep Learning, Convex Optimization

Chongqing University *Sep. 2016 – Jun. 2020*

Major: Electrical Engineering and Automation Degree: Bachelor of Engineering **GPA:** 3.58/4.0

Courses: Computer Communication Networks, Intro to Database System, Advanced Programming Techniques

Scholarships: 2020 Georgia Tech Shenzhen Campus Level A "Merit-Based Scholarship" (5%)
2018 Yangtze Power Scholarship (2/324)
2016, 2018, 2019 Excellent Second Undergraduate Comprehensive Scholarship (6%)

Awards: 2019 Chongqing University Excellent Undergraduate (10%)
2018 Chongqing University Excellent Student (5%)
2018 Chongqing University Science and Technology Innovation Advanced Individual (5%)

TA: ECE Statistical Machine Learning (2022 Fall)
Dependable Computing System (2025 Fall)

Reviewer: ICRA 2024, NAACL 2025, ACL 2025, IJCAI 2025, AAAI 2026
ACM Transactions on Computing for Healthcare

Volunteer: EMNLP 2024

PUBLICATIONS

Ge, Xueren, et al. Expert-Guided Prompting and Retrieval-Augmented Generation for Emergency Medical Service Question Answering. Submitted to *AAAI-26*.

Weerasinghe, K., **Ge, Xueren.,** Heick, T., Wijayasingha, L. N., Cortez, A., Satpathy, A., Stankovic, J. A., & Alemzadeh, H. EgoEMS: A High-Fidelity Multimodal Egocentric Dataset for Cognitive Assistance in Emergency Medical Services. Submitted to *AAAI-26*.

Ge, Xueren, et al. DKEC: Domain Knowledge Enhanced Multi-Label Classification for Diagnosis Prediction. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2024.

Weerasinghe, K., Janapati, S., **Ge, Xueren**., Kim, S., Iyer, S., Stankovic, J.A., & Alemzadeh, H. Real-Time Multimodal Cognitive Assistant for Emergency Medical Services. *2024 IEEE/ACM Ninth International Conference on Internet-of-Things Design and Implementation (IoTDI)*, 85-96.

Fang Xinxin, Wang Bingkai, Kong Hang, **Ge Xueren**, Yang Zhifang, Yu Juan, & Li Wenyuan. (2023). Human posture feature recognition method for neuropsychological comprehension test. *Journal of Chongqing University*, 46(4), 108-119.

Ma, Yuqing, & **Ge, Xueren**. An Effective Method for Defect Detection of Copper Coated Iron Wire Based on Machine Vision. *IOP Conference Series: Materials Science and Engineering*. Vol. 631. No. 2. IOP Publishing, 2019.

PATENTS

Li, R., **Ge, Xueren.**, Li, Q., Chen, C., Ma, J. *Heart rate measurement method and apparatus, and electronic device and storage medium*. WIPO Patent **WO2023061042A1**, filed Aug 19, 2022, published Apr 20, 2023. Assignee: Shanghai Sensetime Intelligent Technology Co Ltd.

Li, R., **Ge, Xueren.**, Li, Q., Zhao, M., Bao, M., Ma, J. *Respiratory rate detection method and device, electronic equipment, and storage medium*. China Patent **CN113887474B**, filed Oct 15, 2021, granted. Assignee: Shenzhen Sensetime Technology Co Ltd.

Yu, H., Kong, H., **Ge, Xueren.**, Wang, B., Wang, Z., Li, W., Yang, Z., Yu, W. *Automatic gesture recognition system for AD*

Xueren Ge

•Tel: +1 470-662-2854 •Email: zar8jw@virginia.edu •Web: https://xueren-ge.github.io/

(analog-to-digital) meter understanding capability test. China Patent **CN111652076B**, filed May 11, 2020, granted. Assignee: Chongqing Zhiyixing Technology Development Co Ltd.

Yu, H., Wang, B., Kong, H., He, W., **Ge, Xueren.**, Yang, W., Li, W., Yang, Z., Lü, Y. *An action recognition method based on Tensorflow target detection*. China Patent **CN111860103B**, filed May 11, 2020, granted July 15, 2025. Assignee: Chongqing Zhiyixing Technology Development Co Ltd.

Ge, Xueren. *Highway electronic information direction board*. China Utility Model Patent **CN209118652U**, filed Oct 9, 2018, granted July 16, 2019. Assignee: Individual.

SKILLS

Programming Languages: Python, C/C++, Java, MySQL, HTML/XML, MATLAB
Machine Learning Frameworks: PyTorch, Tensorflow, Wandb, Huggingface, FSDP, DDP
Modeling & Algorithms: RAG, FAISS, LangChain, LoRA, PPO, DPO, GRPO
Tools & Infrastructure: Slurm, Git, Linux, Docker, HPC Systems

EXPERIENCES

University of Virginia **Graduate Research Assistant** Summer 2022 – Present

DKEC: Domain Knowledge Enhanced Multi-Label Classification for Diagnosis Prediction

- Automated **heterogeneous knowledge graph** construction from 3,000+ medical webpages (Wikipedia, MayoClinic, MedlinePlus) using BeautifulSoup, Web APIs, and LLM prompting (chain-of-thought), extracting 5,000+ normalized medical entities via **UMLS API**.
- Designed a **label-wise attention mechanism** that incorporates heterogeneous knowledge graphs to train language models, to improve multi-label classification by **5% in micro f1** compared with SOTAs on MIMIC-III datasets

Expert-Guided Prompting and Retrieval-Augmented Generation for EMS Question Answering

- Created **EMSQA**, the first EMS MCQA dataset of 24.3K questions, curated based on public and private sources, covering 10 subject areas and 4 certification levels, and accompanied by a structured, subject area aligned EMS knowledge base (KB) with 40K documents and 4M real-world patient care reports.
- Developed two techniques to inject domain expertise into Large Language Models (LLM): 1) an expertise-guided prompting (**Expert-CoT**) that encourages step-by-step reasoning from a domain-specific perspective. 2) an expertise-guided RAG (**ExpertRAG**) that retrieves expertise-aligned knowledge from curated EMS KBs and patient records
- Benchmarked multiple LLMs on EMSQA, evaluating performance across certification levels and subject areas, and compare our framework against SOTA RAG methods. Experimental results show that combining Expert-CoT and ExpertRAG yields up to a **4.67% improvement** in accuracy. Notably, the **32B expertise-augmented models** pass all the EMS certification simulation exams.

EgoEMS: A High-Fidelity Multimodal Egocentric Dataset for Cognitive Assistance in EMS

- Designed and executed benchmarking pipeline for **zero-shot audio models** (Whisper, Whisper-Timestamped, Google Speech, Gemini-2.5-Pro), evaluating ASR accuracy, latency, and word-level timestamp precision in EgoEMS dataset.
- Conducted comprehensive comparative analysis of zero-shot **vision-language models** (Qwen-2.5-VLM, VideoLLaMA-3, Gemini-2.5-Pro) for **video understanding**, measuring **action classification** accuracy, **temporal segmentation** quality, and inference efficiency to guide model selection for downstream applications.

SenseTime Incorporated **Algorithm Developer Internship** *Dec. 2020 – Jun. 2021*

Contactless Vital Signs Estimation from Thermal Imaging using Computer Vision

- Designed and implemented a contactless physiological monitoring algorithm for the SenseThunder Air product, integrating **landmark detection** and **homography transformations** to extract Regions of Interest from thermal video streams for real-time heart rate and body temperature estimation.
- Applied advanced signal processing methods—including **FFT**, smoothing, and **bandpass filtering**—to recover weak respiratory and cardiac signals from continuous thermal frames.

Xueren Ge

•Tel: +1 470-662-2854 •Email: zar8jw@virginia.edu •Web: https://xueren-ge.github.io/

- Developed a complete **end-to-end Python framework** for data preprocessing, algorithm implementation, performance evaluation, and error analysis, enabling robust real-world deployment.

Chongqing University **Undergrad Research Assistant** *Jul. 2019 – Nov. 2019*

Intelligent Diagnosis of Alzheimer's disease Based on AI Image Technology

- Designed and implemented an AI-based system to automatically assess the severity of Alzheimer's disease using video and image data, enabling objective and scalable cognitive evaluation.
- Developed a human eye state recognition GUI using **landmark detection Dlib** for automated cognitive assessment.
- Designed a video analysis pipeline integrating **OpenPose-based posture estimation**, **image morphology processing**, and **Fast R-CNN detection** to detect patient's action recognition for Alzheimer's severity evaluation.
- Trained a **CNN with a ResNet-50 backbone** in PyTorch to classify patient hand-drawn geometric figures, improving diagnostic accuracy and enabling early detection of cognitive impairment.

COURSE PROJECTS

LLMs for Diagnosis Prediction *Aug. 2024 – Dec.2024*

Intro: Explored prompting and finetuning strategies for diagnosis prediction using Electronic Health Records (EHRs).

- Developed a "chain-of-diagnosis" prompt template, enabling effective finetuning of Llama-3.1-8B with LORA for enhanced diagnostic accuracy.
- Conducted comprehensive comparisons between our finetuned model and state-of-the-art Medical LLMs using chain-of-thought prompting to evaluate improvements in diagnosis prediction.

N-Version Programming on LLMs *Aug. 2023 – Dec.2023*

Intro: Improve multiple-choice question answering accuracy by merging multiple LLMs

- Designed a majority voter and weighted majority voter to combine responses from GPT-3.5, PALM-1 and Llama-2 and Improved multiple-choice question answering by 5% in accuracy

BERT Visualization and Interpretation *Aug. 2022 – Dec. 2022*

Intro: Analysis of whether BERT create reasonable embeddings in each layer

- Visualized anisotropy problem by randomly sampled two words and calculate cosine similarity in 12 layers
- Conducted biased analysis by removing sentences related to "female" concepts in SNLI training datasets
- Discussed the redundancy of BERT model by freezing each layer for classification problems

Phishing Websites detection based on Machine Learning *Jan. 2021 – May. 2021*

Intro.: Designed phishing websites detection algorithms and a phishing web inquiry application

- Extracted and encoded URL features (domain, IP address, DNS record, redirection, protocol and etc.) by Python
- Trained SVM, RF, GDBT and etc. models separately, then built ensemble model to boost overall performances
- Developed a web application based on ensemble model interface and PhishTank API

Research on Bus Station Waiting People Statistics Based on Wireless Interactive System *Jun. 2017 – Jul.2018*

Intro.: Developed a wireless interactive terminal to get around the problem of difficulty in dispatching campus sightseeing buses in Chongqing University and random number of waiting people.

- Implemented wireless communication, LCD display and other logical C programming based on STM32 core
- Designed and soldered circuit board with STM32 as the core and ZigBee wireless transmitter, matrix keyboard module, power module and LCD display module as the peripheral circuit
- Modeled campus bus dispatching as TSP problem and derived optimal solution by ant colony algorithm on MATLAB

ACADEMIC COMPETITIONS

2018 Mathematical Contest in Modeling (MCM) Meritorious Winner *Feb. 2018*

- Data Cleaning: used the grey predictive modeling method to supplement the missing data, exclude outliers
- Used NARX autoregressive neural network to predict energy development and discussed the model Robustness
- Converted energy development strategy into nonlinear optimization problem and solved with Genetic Algorithm