

Jiamin Yao

Location: Greater St. Louis, open to relocation
618-631-5638, jamiyao0508@gmail.com, [LinkedIn](#), [Portfolio](#)

SKILLS

Languages: Python, JavaScript, TypeScript, C#, C/C++, R, Java, Bash, SQL

Frameworks: PyTorch, TensorFlow, scikit-learn, React, SupaBase, SvelteKit, Next.js, Transformers, OpenCV

Tools: Docker, Git, Agile, SQLite, NumPy, Pandas, Matplotlib, MATLAB, Power Bi

EDUCATION

Master of Science in Computer Science, Southern Illinois University Edwardsville Jan 2024 – Dec 2025

Coursework: Algorithms, Machine Learning, Deep Learning, Artificial Intelligence, Digital Image Processing

Bachelor of Science in Marine Technology, Xiamen University

Sep 2015 – Jun 2019

Coursework: Fundamental Computer Application, C Programming

WORK EXPERIENCE

Research Assistant, *SIUE*, Edwardsville, IL, USA

Feb 2026 – Present

- Researched and applied uniform and quadratic quantization techniques to compress large language models and diffusion models, reducing model size while preserving generation quality.
- Implemented model retraining and fine-tuning pipelines using PyTorch and Bash to optimize quantized models, ensuring high performance and accuracy across various downstream benchmarks.

Software Developer, *Anrib LLC*, St. Louis, MO, USA

May 2025 – Aug 2025

- Architected an autonomous AI agent using C#, Semantic Kernel and LangGraph-based state machines to orchestrate multi-step task generation and execution.
- Engineered a GUI automation engine with atomic input operations like keystroke injection and programmatic clicks; managed environment isolation using Docker.
- Collaborated in an Agile team environment, using Git for version control and CI/CD workflows.

Advisor Global Buyer, *Dell Technologies*, Xiamen, Fujian, China

Jul 2019 – Sep 2023

- Delivered Power BI dashboards using data storytelling to provide leadership with actionable supplier risk insights.
- Optimized SAP procurement workflows, improving supplier allocation accuracy by 15% through SQL data analysis.
- Leveraged Oracle Data Hub to analyze real-time data, prioritizing shipments to reduce costs and prevent shortages.

ACADEMIC CASES/ PROJECTS

- **Gongshami.com** - Developed a responsive web app using SvelteKit and Next.js for the frontend, with SupaBase and SQLite for real-time data storage and user authentication. Implemented custom AI personalities and social reactions via OpenAI API, deployed using Docker containers.
- **Chatbot Preference Prediction** - Built a pipeline comparing LightGBM, CNN, BERT, and Gemma to predict user preferences from 57K samples; processed large-scale data using Pandas and NumPy. Conducted feature engineering and model training within a TensorFlow/PyTorch environment to achieve high prediction accuracy.
- **Breast Cancer Prediction** - Developed a diagnostic classifier using PCA dimensionality reduction and scikit-learn's Grid Search CV to optimize 7 ML models (SVM, Random Forest, etc.). Visualized model performance (ROC-AUC) and feature importance using Matplotlib and Pandas.
- **Image Enhancement Pipeline** - Built a 3-stage pipeline for pet photography using MATLAB and OpenCV (Python) for CLAHE contrast enhancement, unsharp masking, and distance-based background blur. Evaluated image quality using SSIM, PSNR, and edge strength metrics.
- **Travel Route Planner** - Implemented TSP optimization using C++, comparing Greedy $O(n^2)$, Divide & Conquer $O(n \log n)$, and Dynamic Programming $O(n^2 \cdot n)$ across 18 European cities. Analyzed algorithmic efficiency for multi-modal transport analysis.

SELECTED PUBLICATIONS

- **SPQ: An Ensemble Technique for Large Language Model Compression.** (LREC 2026 Oral) - Achieved 75% memory reduction and 1.9× faster inference while maintaining LLM performance on perplexity and accuracy on downstream benchmarks such as C4, TruthfulQA, and GSM8K.