Last updated in Jul 2025

Xiaobo Xing

Brisbane, Australia xiaobo.xing@student.uq.edu.au +61 0493712603

Personal Details

I am Xiaobo Xing, a Master's graduate in Computer Science (Management) from the School of Electrical Engineering and Computer Science at The University of Queensland (UQ), where my research focused on large language models (LLMs) for tabular data. Before joining UQ, I gained one year of experience as a software engineer and obtained a Bachelor's degree in Information Engineering from Jilin University. I am a citizen of the People's Republic of China, a native Mandarin speaker, and proficient in English.

Educational Qualifications and Academic Awards

The University of Queensland | Brisbane, Australia *Jul 2023 – Jul 2025*
Master of Computer Science (Management)

- School of Electrical Engineering and Computer Science (EECS)
- **GPA: 6.625/7**
- **Research Thesis:** LLM-based E-tutors **(GPA 7, HD)**
  - This thesis research project focuses on developing a chatbot powered by Large Language Models (LLMs) to serve as a virtual tutor for courses at the University of Queensland. The e-tutor addresses the limitations of traditional lecture and tutorial hours, which often prevent students from receiving real-time academic assistance outside of scheduled sessions. By leveraging the advanced capabilities of LLMs, the chatbot is designed to answer course-related questions, provide detailed explanations, and support student learning through personalised and effective interactions. The project's primary objective is to enhance the student learning experience and improve academic performance by filling the gap in real-time, reliable academic support. The research covers the full development life-cycle, including data collection, experimentation, and evaluation, with an emphasis on building a scalable and high-performing solution applicable to real-world academic settings.
  - **Supervisor:** Prof. Hongzhi Yin
  - **Dissertation Length:** 60-70 pages (year-long 6 units)
  - **Technical Workflow:**
    - **Data Collection:** Skilled in manual data collection and developing customised processing methods tailored to diverse resources, ensuring high data quality and relevance for downstream tasks.
    - **Training and Fine-tuning:** Applied tailored fine-tuning approaches based on resource characteristics, including continued pre-training (CPT) to enrich domain-specific knowledge and instruction-tuning for task-specific applications.
    - **Retrieval-Augmented Generation (RAG):** Expertise in optimising traditional embedding similarity (vector store) methods by integrating BM25 for keyword exact match and using Reciprocal Rank Fusion (RRF) techniques to combine multiple ranking strategies. Enhanced retriever accuracy significantly and reduced hallucination in LLMs by leveraging LLMs to summarise retrieved contexts effectively.
    - **Prompt Engineering:** Focused on optimising small-scale LLMs through advanced prompting strategies, such as Chain-of-Thought (CoT) reasoning and in-context learning (e.g., one-shot and few-shot), to enhance performance on complex tasks.
    - **Agent Development:** Experienced in building AI agents to manage chatbot behaviour and

RAG workflows, supporting multi-turn conversations, role definitions, and context-aware task execution. This ensures a seamless, dynamic, and interactive user experience.

- **Academic Achievement Awards:**
  - Dean's Commendation for Academic Excellence 2025 Semester 1, Faculty of EAIT
  - Dean's Commendation for Academic Excellence 2024 Semester 2, Faculty of EAIT
  - Dean's Commendation for Academic Excellence 2024 Semester 1, Faculty of EAIT

Jilin University (Project 985 University of China) | **Changchun, China** *Aug 2018 – Jun 2022*
Bachelor of Information Engineering

- College of Communication Engineering
- **Grade:** 79.33/100
- **Research Thesis:** A Survey of Cyber-Physical Attacks and Deep Learning-Based Detection Methods in Smart Distribution Systems **(GPA 88/100, HD)**
  - This graduation thesis focuses on designing a deep learning-based anomaly detection algorithm to identify cyber-physical attacks in water distribution systems. With the transition of traditional water supply systems to cyber-physical systems (CPS), efficiency and water quality have significantly improved; however, these advancements also increase vulnerability to cyber-physical attacks. Traditional detection methods often fail to predict unknown attacks, highlighting the need for innovative solutions. Leveraging a publicly available dataset for water network security research, the project utilised TensorFlow framework and developed an autoencoder (AE) model to detect anomalies, classify attack scenarios, and evaluate detection performance using advanced metrics like accuracy and ROC curves. The research covers key processes such as dataset analysis, model development, and parameter optimisation, aiming to enhance the practical applicability of anomaly detection in real-world CPS environments.
  - **Supervisor:** Prof. Juan Li
  - **Dissertation Length:** 18000 words, 38 pages
  - **Technical Workflow:**
    - **Data Preparation:** Collected and prepared datasets for anomaly detection, including training, validation, and testing sets. The training set contains normal operational data for water distribution systems, while validation and testing sets include simulated cyber-physical attack scenarios. Leveraged public datasets like BATADAL for realistic simulations.
    - **Algorithm Design:** Developed a detection algorithm based on autoencoder (AE) models for unsupervised learning. Implemented encoding-decoding structures to analyse input-output reconstruction errors for detecting anomalies effectively.
    - **Model Training:** Trained the autoencoder using TensorFlow and optimised with Adam optimisation, adjusting parameters like learning rate and batch size for optimal performance. Fine-tuned hyperparameters based on experimental results to improve model accuracy and recall.
    - **Threshold Selection:** Defined reconstruction error thresholds based on performance metrics (e.g., accuracy, recall). Evaluated thresholds by analysing precision-recall trade-offs and optimising for low false positive/negative rates.
    - **Parameter Tuning:** Investigated the effects of different learning rates and thresholds on model performance. Demonstrated that a learning rate of 0.01 and a reconstruction error threshold of 950 achieved the best results, maximising AUC and minimising classification errors.
    - **Evaluation and Testing:** Assessed model performance using metrics like confusion matrices and Receiver Operating Characteristic (ROC) curves. High Area Under Curve (AUC) scores validated the model's classification capability and robustness in anomaly detection.

- **Academic Achievement Awards:**
  - Certificate of Participation in the Electronic Design Competition, College of Communication Engineering

Employment History

Casual Academic (Casual) *Brisbane, Australia*
School of EECS, The University of Queensland *Feb 2025 – Jun 2025*

- **Tutor of Social Media Analytics (INFS7450)**
  - Conducted weekly 2-hour tutorial sessions, guiding **over 50** students through course content and practical coding demos.
  - Assisted students with course-related questions, explaining and clarifying complex topics.
  - Evaluated and graded assignments and exams.

Software Engineer (Full-time) *Changchun, China*
CRRC CHANGCHUN RAILWAY VEHICLES CO., LTD. *Jul 2022 – Jul 2023*

- **Spring Boot Micro-Services Backend Develop (Java)**
  - Designed and implemented scalable and efficient server-side miscro-service systems using Spring Cloud, including developing RESTful APIs and managing databases.
  - Optimised backend infrastructure to support data-intensive applications in manufacturing enterprises.
  - Contributed to the digital transformation of manufacturing enterprises by digitising internal workflows and optimising manufacturing and material workflows.
- **Kingdee Fullstack Web Develop (Java)**
  - **Independently** led the design and development of solutions for digitising material and part workflows, integrating multiple external systems to establish a unified data flow.
  - Resolved legacy issues such as duplicate codes for a single item and mismatched codes for multiple items, significantly improving data accuracy and system efficiency.
  - Optimised internal workflows, including financial profit and loss reports, travel reimbursement systems, and company performance tracking platforms.

Other Research Experience

2019 Vancouver Summer Program *Vancouver, Canada*
School of Electrical Engineering and Computer Science *Jul 2019 – Aug 2019*
The University of British Columbia (UBC)

- **Completed Two Full-Credit Courses in the Field of Communication and Digital Systems**
  - Communication Systems: Technology Embedded in Daily Life
  - Introduction to Digital Systems Design with FPGAs
- **Research Training**
  - Participated in structured research training modules integrated into the coursework, focusing on problem formulation, experiment design, and data analysis.
  - Conducted a mini-project in digital systems design, which involved:
    - Designing and simulating digital circuits using FPGA tools to solve a predefined engineering problem.
    - Presenting findings in a formal report, highlighting innovative solutions and design efficiency.

- Developed foundational research skills, including systematic literature review and technical report writing, relevant to the field of communication systems.
- Received a certificate for successfully completing the program, underscoring a strong academic foundation for future research endeavours in computer science and communication engineering.

Research Outputs

None to date

Skills

Programming and Development

- Proficient in Java, Python, and SQL; experienced in developing applications using HTML, CSS, PHP, and JavaScript. Solid debugging skills.
- Proficient in Algorithms and Data Structures, including Graphs, Stacks, Heaps, Search Trees, and Huffman Coding.
- Skilled in development on Linux servers, with expertise in leveraging High-Performance Computing (HPC) systems for advanced applications.
- Skilled in containerisation and orchestration technologies: Docker, Dockerfile, Docker Compose, and Kubernetes.
- Proficient in using PyTorch for development and research in machine learning and deep learning applications; familiar with TensorFlow.
- Intermediate knowledge of RDD programming and Apache Spark for distributed computing.

Data Science

- **Data Mining & Machine Learning:** Expertise in Classification, Clustering, Anomaly Detection, Association Rule Mining, Text Mining, and Linear Regression.
- **Deep Learning:** Skilled in Logistic Regression, Neural Networks, Hyperparameter Tuning, Mini-batch Gradient Descent, Regularisation, Normalisation, Momentum, RMSprop, Adam, Learning Rate Decay, Batch Normalisation, and Softmax Regression.
- **Large Language Models:** Proficient in the full life-cycle of LLM deployment, including research, development, and implementation. Experienced in adapting LLMs for real-world applications through advanced techniques such as data collection, continuous pre-training, fine-tuning, Retrieval-Augmented Generation (RAG), prompt engineering, and AI agent development. Familiar with multimodal large language models (MLLMs) and their integration with visual, tabular, and structured data inputs.

Tools & Services

- Version Control & Database Management: Git, Navicat, MongoDB, Data Warehouses.
- API Development & Testing: Postman, SOAP API, REST API.
- Remote Access & Messaging Systems: MobaXterm SSH Client, Web Services, Messaging Queue (MQ).

Awards

- Dean’s Commendation for Academic Excellence 2025 Semester 1 EAIT, The University of Queensland
- Dean’s Commendation for Academic Excellence 2024 Semester 2 EAIT, The University of Queensland
- Dean’s Commendation for Academic Excellence 2024 Semester 1 EAIT, The University of Queensland
- Best New Employee Award 2022, CRRC CHANGCHUN RAILWAY VEHICLES CO., LTD.
- The Star of Events 2022, CRRC CHANGCHUN RAILWAY VEHICLES CO., LTD.