

MingJia Yang

+86 18940200957 | mingjiay2026@yeah.net

EDUCATION

Xiamen University Malaysia, Sepang, Malaysia Sep. 2022 - Aug. 2026 (Expected)

- **Degree:** B.Eng. (Hons) in Computer Science and Technology | **GPA:** 3.72/4.0
- **Relevant coursework:** Data Structures, Algorithm Design, Compiler Principles, UNIX Programming, Object-Oriented Programming-C++, Probability & Statistics, Linear Algebra, Discrete Mathematics
- **Honours & Awards:** Distinction Award (Academic Year 2022, 2023, 2024) – Top 30% academic performer in B.Eng.

PROFESSIONAL EXPERIENCE

LLM Intern, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China Sep.2025 – Present

- Conduct a literature review on applying large language models to remote sensing tasks (RS-LLM), examining recent techniques for integrating advanced LLM models with geospatial data analysis.
- Co-author a comprehensive survey of RS-LLM research, summarizing state-of-the-art methods and identifying key challenges and opportunities in this emerging field.
- Fine-tune Vicuna v1.5 (7B) with Hugging Face/DeepSpeed, applying LoRA on domain-specific remote sensing corpora and tracking metrics via Weights & Biases.
- Configure multi-GPU A100 training with dataset sharding and ZeRO optimization; build preprocessing pipelines and evaluate models using perplexity, BLEU, and F1.
- Test LoRA vs. prefix-tuning and curriculum training strategies, improving domain adaptation for remote sensing text and reducing misclassification in geospatial tasks.

ERP Implementation Intern, Shenyang Yongyou Software Co., Liaoning, China Jan.2025 - Feb.2025

- Participated in a structured ERP training program and gained functional experience across core data-driven modules such as Inventory, Procurement, and Manufacturing.
- Configured backend workflows and explored SQL-based data schemas across multi-entity business records, enabling hands-on exposure to structured tabular data and ERP logic mapping.
- Extracted transactional logs and operational records from test environments to perform basic metrics tracking and error analysis, identifying configuration bottlenecks and input inconsistencies.
- Designed and delivered Excel-based analytical dashboards using pivot tables and formulas to summarize module coverage, data integrity status, and operation flags.
- Presented workflow validation findings to internal supervisors with a focus on system data flow and process consistency, receiving positive feedback for clarity and analytical depth.

ML Engineer Intern, Shenyang Diyue Network Technology Co., Ltd., Liaoning, China

Jul. 2024 - Sep. 2024

- Supported internal analytics initiatives by building a Python-based preprocessing pipeline for customer feedback datasets (~800k records); applied text normalization, missing value imputation, and category standardization.
- Assisted in feature extraction from unstructured log data using regex-based parsers and basic NLP techniques (e.g., keyword flagging, sentiment token matching), feeding into subsequent modeling tasks.
- Collaborated with the data science team to experiment with simple logistic regression and tree-based models to explore potential indicators of product drop-off or churn.
- Built interactive diagnostic plots and data summaries, helping stakeholders understand feature distributions, error cases, and temporal usage patterns.
- Participated in weekly model review meetings, using Git and Jupyter notebooks to share progress and maintain code clarity across the team.

RESEARCH EXPERIENCE

Multi-Strategy Ensemble-Enhanced BERT-LSTM Model for Sentiment Analysis Feb. 2025 - Jun. 2025

Conference: 2025 International Conference on Modeling, Simulation and Computing Science (EI Indexed)

Advisor: Prof. Tingwei Zhang

Co-First Author

- Developed an ensemble NLP model combining BERT and BiLSTM for sentiment analysis of hotel reviews. Achieved ~87% accuracy (F1-score ~0.87), outperforming individual baseline models (BERT, BiLSTM, TextCNN).
- Implemented and fine-tuned multiple model variants and integrated their predictions via a multi-strategy ensemble. Responsible for data collection & preprocessing (cleaned 10,000+ review samples) and establishing the training/evaluation pipeline with hyperparameter tuning.
- Drafted and submitted a research manuscript (under peer review), including writing key methodology sections and creating model architecture diagrams.

PROJECT EXPERIENCE

Food Order System

Jun. 2025

- Led a team of 3 to develop a full-featured C++ food-ordering application for a course project, meeting requirements for object-oriented design (classes, inheritance, polymorphism), STL usage, file I/O, and exception handling.
- Designed core modules (LoginSystem, FoodOrderSystem, Order, Restaurant, and a hierarchy of Food subclasses) with clear separation of interface and implementation across multiple files.
- Implemented features including user authentication, dynamic menu generation, customizable food preferences, multiple delivery options, discounts, and reorder functionality, using STL containers (vector, pair, string) and CSV files for persistent storage, all orchestrated through a looping CLI.
- Integrated teammates' code and resolved critical bugs: replaced raw pointers with smart pointers to fix memory leaks, added boundary checks and exception guards, and refactored the CSV parsing logic to eliminate segmentation faults. Oversaw comprehensive testing and documentation (including UML diagrams), resulting in a stable system that earned full marks for both functionality and documentation quality.

Interactive Matrix Expression Calculator

Nov. 2024

- Designed and implemented an interactive matrix calculator in C++ using Flex for lexical analysis and Bison for parsing.
- Developed a custom grammar supporting variable assignments, numeric constants, and matrix expressions.
- Implemented syntax-directed semantic actions to perform matrix operations (addition, subtraction, scalar and matrix multiplication, element-wise operations, transpose) and advanced routines (via Gauss-Jordan elimination). Included runtime checks for dimension compatibility and undefined variables, providing clear error diagnostics without crashing the session.
- Designed a lightweight REPL interface for continuous user input with results displayed in aligned tabular format and automatic memory management (RAII). Developed 40+ unit tests for edge cases (e.g. non-square or singular matrices).

AutoManager Pro

Jun. 2024

- Developed AutoManager Pro, a C++ command-line vehicle information and billing system, showcasing algorithm selection and complexity optimization in data management tasks. Encapsulated car, billing, and user records in STL containers and provided features such as date-range billing queries, inventory sorting, and license-plate search via a menu-driven interface.
- Implemented core operations with customized algorithms: used merge sort, quick sort, or selection sort for different sorting needs and binary search for lookups, all wrapped in reusable modules for performance benchmarking.
- Optimized billing queries by converting date strings into sortable timestamps and applying quick sort followed by binary search, improving query performance from linear to logarithmic time complexity.
- Ensured robust operation through extensive input validation and boundary checks, reinforced by 30+ unit tests.
- Documented algorithm trade-offs on 10k-record datasets for feature integration and future development.

TECHNICAL SKILLS

- **Languages:** C++, Java, Python (Proficient)
- **Frameworks & Tools:** Git, Linux/Bash, GDB, OpenGL, Flex/Bison
- **Machine Learning:** PyTorch, TensorFlow, HuggingFace, NumPy, Pandas, Matplotlib, Scikit-learn, OpenCV
- **Additional:** LaTeX, Markdown, Microsoft Office, Adobe Photoshop