Andy (Xiangyu) Cui

GitHub | LinkedIn | Portfolio

xiangyucui@outlook.com | (402)-853-3000

SKILLS

- Programing languages: Java, Python, JavaScript, Django, C/C++, Assembly
- Database: MySQL, PostgreSQL, MongoDB, Redis, SQLite
- Machine Learning: TF-IDF, Naïve Bayes, GPT-2, BERT, CNN, Transformer, XGBoost, Random Forest
- AI Tools: Cursor, Bubble, OpenCV, UiPath, Google Colab, Deepseek R1

WORK EXPERENCE

CAC Auto Group Inc • Full-Time Southborough, MA, USA

Data Engineer • predictive maintenance system for e-commerce platform Feb. 2024-Present

- Developed and maintained a predictive pricing system for vehicles on CarGurus using **AWS serverless** architecture, enhancing market compatibility and operational efficiency. This system utilizes AWS services such as **S3**, **Lambda**, **DynamoDB**, **SNS**, **CloudWatch**, and **Kinesis** for real-time data processing, with **Python** and **AWS CloudFormation** for infrastructure.
- Implemented real-time monitoring with **AWS Kinesis** to track market data fluctuations, enabling automated detection and response to price deviations. This improved the accuracy of vehicle pricing on the company's website, resulting in significant enhancements in operational efficiency and real-time price adjustments.
- Designed and managed a structured AWS data lake to facilitate efficient data access for BI teams, improving data processing efficiency and maintaining high data quality through continuous monitoring with AWS Step Functions, Glue Data Catalog, and Athena.

AlpaLifeBio Inc • Internship Woburn, MA, USA

Software Engineer of Data Engineer • Analysis of Potential Consumer Demand Dec. 2022-June. 2023

- Developed a high-capacity AWS streaming data pipeline to automate the ingestion of biomedical data from multiple public databases into a Kinesis Data Stream. This robust system, processing over **500,000 entries** daily, utilized Lambda for real-time data transfer and leveraged **S3** and **DynamoDB** for efficient, scalable storage and retrieval, reducing latency and operational costs.
- Configured and optimized a structured **SQL** database to integrate and streamline data from diverse biomedical sources. Implemented an efficient tag-processing system that reduced data retrieval times by 80%, significantly enhancing the accessibility and analytical utility of the data.
- Applied advanced data matching algorithms, including **TF-IDF** techniques, to perform detailed data comparisons, achieving a **95% accuracy** in identifying potential client profiles. This approach effectively supported targeted marketing strategies by precisely pinpointing high-value clients in the biomedical field.

SELECTED PROJECT

Job Recommendation System Design Nov. 2024-Jan. 2025

- Modeling (TF-IDF, Collaborative Filtering): Content-based filtering using TF-IDF and cosine similarity, achieving 82% precision in matching user skills to job descriptions; Collaborative filtering with implicit user feedback, improving recommendation diversity by 18% via matrix factorization.
- Deepseek RL Integration: Leveraged deepseek API to dynamically adjust recommendations based on real-time user feedback.
- Hybrid Matching: Reduced cold-start bias by 30% through RL-driven exploration of niche roles.

Auto QA Chat Agent for Customer via NLP Sep. 2023-Dec. 2023

- Data preprocessing (Python, pandas): Collected and cleaned large-scale Amazon customer Q&A datasets to train and fine-tune the AI agent, ensuring robust data quality for model training.
- Modeling (PyTorch, NLP): Developed a conversational AI agent leveraging state-of-the-art NLP models, including BERT and GPT2, fine-tuned for question answering (QA) tasks. Optimized hyperparameters for both models (BERT: 25 epochs, LR=2e-5, BS=8; GPT2: 25 epochs, LR=5e-5, BS=8) to enhance contextual understanding and response generation.
- Reduced operational costs by automating 70% of routine inquiries, allowing human agents to focus on complex cases. Integrated continuous learning to adapt to evolving customer queries and improve long-term efficiency.
- Evaluation (sklearn): Conducted 10-fold cross-validation for BERT and 5-fold for GPT2, achieving average BLEU scores of 0.9264 (BERT) and 0.7695 (GPT2).

EDUCATION | Relevant Coursework | Location & Date

Northeastern University Boston, MA, USA

M.S. in Artificial Intelligence of Khoury College Sept. 2021-Dec.2023

University of Nebraska-Lincoln Lincoln, NE, USA

B.S. in Computer Science of Arts Science College Sept. 2016-May. 2020