

Andy (Xiangyu) Cui

xiangyucui@outlook.com | (402)-853-3000 | [Linkedin](#) | [Portfolio](#)

EDUCATION

Northeastern University

M.S. in Artificial Intelligence of Khoury College

Boston, MA

Dec 2023

University of Nebraska-Lincoln

B.S. in Computer Science of Arts Science College

Lincoln, NE

May 2020

PROFESSIONAL EXPERIENCE

CAC Auto Group LLC

Feb. 2024-Present

Data Engineer

Southborough, MA

- Developed and maintained a predictive pricing system for vehicles on CarGurus using **AWS serverless architecture**, optimizing market compatibility and operational efficiency. Utilized AWS services including S3, Lambda, DynamoDB, SNS, CloudWatch, and Kinesis for real-time data processing, 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. Improved the accuracy of vehicle pricing on the company's website, resulting in an 80% boost in daily operational efficiency and over 50% improvement in real-time price adjustments compared to industry standards.
- Designed and deployed a **CRM** system on **Docker** using **WordPress**, ensuring the system's scalability and simplifying ongoing maintenance; Customized the CRM with essential features such as source tracking and lead status monitoring to enhance sales processes and improve visibility, saving significant time in monitoring sales conversion rates and successfully recovered 80% of potential sales leads, resulting in a notable increase in overall sales performance.

AlpaLifeBio Inc

Dec. 2022-June. 2023

Software Engineer Intern

Woburn, MA

- Developed and deployed a **Python**-based web scraper to collect 500K+ public biomedical records daily; Designed and implemented a high-capacity AWS streaming data pipeline to automate data ingestion from multiple public databases into Kinesis Data Stream; Utilized AWS Lambda for real-time data transfer and **S3/DynamoDB** for scalable storage, reducing latency and operational costs by 30%.
- Applied advanced data matching algorithms and **TF-IDF** techniques to perform relevancy analysis on the data collected, achieving 95% accuracy in identifying potential client profiles with high relevancy of the company's core business.
- Utilized **T-SNE** for dimensionality reduction and presented visualized insights to stakeholders, demonstrating correlations with GPCR drug support; Consolidated high-potential collaboration leads, increasing lead conversion rates by 20% for the business development team.

PROJECTS

Job Recommendation System Design

Jan 2025

- Developed a user interface for job searching using **Axure RP 10**; Applied content-based filtering using **TF-IDF** and cosine similarity, achieving 82% precision in matching user skills to job descriptions; Conducted **collaborative filtering** in **Python** with implicit user feedback, improving recommendation diversity by 18% via matrix factorization.
- Leveraged **deepseek API** to dynamically adjust recommendations based on real-time user feedback; Reduced cold-start bias by 30% through RL-driven exploration of niche roles.

Stock Price Prediction with Deep Learning

Oct 2024

- Collected the historical stock price and other financial assets data on the company of interest; Conducted data preprocessing by applying min-max scaling in **Sklearn** to normalize stock price values, ensuring consistency across the dataset.
- Implemented **LSTM**, **GRU**, and **Transformer models** in **PyTorch**, optimizing hyperparameters (e.g., number of layers, optimization methods) through grid search, increasing model accuracy by 20%; Visualized opening and closing price trends to assess model performance in Python.

Auto QA Chat Agent for Customer via NLP

Dec 2023

- Collected and cleaned large-scale Amazon customer Q&A datasets, ensuring high-quality data for effective AI model training; Developed a conversational AI agent in **PyTorch** leveraging state-of-the-art NLP models, including BERT and GPT2, fine-tuned for question answering (QA) tasks.
- Optimized hyperparameters for **BERT** and **GPT-2** to improve contextual understanding and response generation; Conducted 10-fold cross-validation for BERT and 5-fold for GPT-2, achieving average BLEU scores of 0.9 for BERT and 0.8 for GPT-2.
- Integrated continuous learning capabilities to adapt to evolving customer queries, improving long-term operational efficiency; Successfully automated 70% of routine customer inquiries, significantly reducing operational costs.

SKILLS

- Programming & Frameworks:** Java, Python, JavaScript, HTML, CSS, C/C++, Assembly, VB, Spring Boot, React.js, Node.js, jQuery, Maven, Tomcat, Servlet, JUnit, Nginx
- ML/AI:** NLP, Naive Bayes, GPT-2, BERT, CNN, LSTM, Transformer, LLM, Prompt Engineering, PyTorch
- DevOps & Tools:** MySQL, PostgreSQL, MongoDB, Redis, SQLite, GitHub, Azure DevOps, Jenkins, Jira, AWS, Docker, Axure, TFSA