

Xinchi (Chloe) Liu

Madison WI, US

+1 (608) 286-2976 | liuxinchi9@gmail.com | [Portfolio](#) | [LinkedIn](#) | [Github](#)

EDUCATION

University of Wisconsin - Madison

Aug.2024 - Jun.2026 (Expected)

Master of Science in Computer Science

Madison, US

- Machine Learning, Based Image Synthesis and Manipulation (Expected)

South China University of Technology

Aug.2015 - Jul.2019

Bachelor of Accounting, GPA 3.6/4.0

Guangzhou, CN

- Statistic Pinciple, Database Management, Data Modeling, Probability Theory, Calculus, Linear Algebra

WORK EXPERIENCE

SHEIN

Algorithm

Apr.2021 - Aug.2024

- Spearheaded the integration of large multimodal models **GPT4**, **CLIP** to analyze product fashion attributes and categorization, achieving **93%** accuracy
- Developed and maintained a **Hadoop** big data governance framework, overseeing the ingestion and management of billions of competitor data points, ensuring data integrity within **97%**.
- Implemented analytics **OLAP** platforms to monitor user behavior and manage event life cycles, increasing data query efficiency by **30%**.

Evergrande Real Estate Group

Data Product Manager

May.2019 - Apr.2021

- Implemented user analytics dashboards, event tracking framework, and attribution analysis supporting business decisions. Designed 113 metrics based on **AARRR** Pirate Metrics framework and monitored performance consistently.
- Led the development of a customer relationship management (**CRM**) platform encompassing the entire customer journey, integrating data analysis to enhance customer insights and drive **2 million** in revenue.

WORK PROJECTS

Enhanced Product Categorization with RAG

SHEIN

Algorithm Engineer

Dec.2023 - Jul.2024

- Leveraged **Retrieval-Augmented Generation**(RAG) and GPT3.5 to auto-match product titles with category trees, surpassing traditional NLP models by 23% in accuracy.
- Implement **indexing** strategies for effective embedding with text-embedding-3-large.
- Fine-tuned algorithms with cosine distance and SVM clustering to balance cost and precision, achieving **99%** accuracy in top 20 candidate sets.
- Construct prompt** and generate the most matching category with GPT3.5, reaching 93% accuracy.

Fashion Attributes Extraction for Trend Analysis with LMMs and CLIP

SHEIN

Algorithm Engineer

Jan.2023 - Jul.2024

- Leverage **Large Multimodal Models**(LMMs) like GPT4, Gemini, Idefics2 with prompt engineering to recognize and categorize fashion attributes from competitors' products, achieving 96% accuracy across 500 attributes.
- Developed an agent for model coordination and feedback integration with **RLHF** and **fine-tuning**, increasing accuracy by 23% than traditional CNN ways.
- Assist in **CLIP** fine tuning with internal product text image pairs for trend product recall, achieving 2.5% higher exposure rate.

Competitor Intelligence Framework with Hadoop Big Data Governance

SHEIN

Big Data Engineer

Jun.2022 - Feb.2024

- Data Governance**: Managed trillions of web scraping data from three competitor websites, address dirty data from anti-crawling to main the data quality accuracy within 97%.

- **Data Warehouse Architecture:** Designed a scalable Hadoop Hive architecture and implemented backup strategies to improve data repair efficiency by 30%.
- **Sales Forecasting:** Developed an ARMA model for predicting daily sales, maintaining annual Revenue accuracy within 5%.
- **Graph Search Service:** Developed a Kafka CV service with completeness from 63% to 99%. Utilized Milvus vector database to enhance accuracy from 78% to 83% .

Event Tracking and Behavioral Analytics Platform

SHEIN

Big Data Product Manager

May.2021 – Dec.2022

- **Platform Development:** Designed and launched the Event Management platform, managing life cycle of 30,000+ event trackers, increasing test efficiency by 30% and tracker completeness from 75% to 87%.
- **User Journeys Analysis:** Implemented OLAP capabilities for trillion-scale data, utilizing linked lists and directed graphs to enhance query efficiency by 30%.
- **Impact:** Successfully engaged over 1000 active monthly users.

FURTHER EDUCATION

Coursera Certification

- Miracles of Human Language: An Introduction to Linguistics by University Leiden
- Natural Language Processing with Probabilistic Models by DeepLearning
- Applied Data Science with Python Specialization by UMich
- Machine Learning Specialization by DeepLearning
- Data Analytics Certification by Google
- Chatgpt Prompt Engineering for Developers by DeepLearning

English Certification

- TOEFL 106, GRE Q156+V167+W3.5

Kaggle Projects

- House Price Predict with top 3% ranking
 - * Feature engineer one-hot encoding to analyze 114 features.
 - * Use ElasticNet, Ridge, LGBM Regressor and other 4 models to predict.

AWS Training

- Architecting on AWS with AWS Jam
- Machine Learning and GenAI application Development

TECHNICAL SKILLS

Programming: Python, Sql, LaTeX, JavaScript

AI & NLP: LLMs, Gemini, GPT, Claude, Stable Diffusion, RAG

Big Data: Kafka, Hadoop Hive, Spark, ClickHouse, OLAP

Frameworks Tools: Tensorflow, Pytorch, Git, AWS, Scikit-learn.

Data Visualization: Matplotlib, Seaborn, Tableau, Google Analytics

Product Design: Figma, Xmind, Axure, ER Map, Prototype, UML