

SHIJUN JU

jushijun@gmail.com ◊ [linkedin.com/in/shijunju/](https://www.linkedin.com/in/shijunju/) ◊ shijunju.com ◊ 647-649-8130 ◊ Barrie, ON

SUMMARY

Versatile and application-focused Data Scientist and AI Engineer with a Ph.D. in Economics, combining diverse skillsets and extensive experience in AI, machine learning, and data analytics. Proficient in Python, with a strong foundation in ML and advanced AI technologies.

- Keywords: LLM, Speech-to-text, Fine-tuning, LoRA, QLoRA, Economics

RELEVANT EXPERIENCE

KaggleX Fellowship *Finetuning LLMs* [LINK](#) Remote Sep 2024 – Dec 2024

- Selected as one of 81 fellowship recipients from over 3,000 applicants for the fellowship program.
- Project **ReguGuard AI** selected as one of the 16 showcases in the end of program celebration
- Developed an AI chatbot for financial risk compliance by **fine-tuning Gemma-2b-en** and **Gemma-7b** using **LoRA** and **QLoRA** on GPU and **TPU**.
- Achieved an accuracy rate of **78.6%** with the best model, demonstrating expertise in machine learning and model optimization.
- Adapted LLamaIndex's RAFT DatasetPack module and used OpenAI API to generate over **14,000** question-answer pairs for training

Skinopathy: AI Researcher *Fine-tuning Speech-to-text Model* Remote Sep 2024 – Dec 2024

- Fine-tuned the **OpenAI Whisper Large-v3 Turbo** model to improve audio recognition of dermatology terminology using **QLoRA**
- Achieved a **33% reduction** in Word Error Rate, improving accuracy from **14.5% to 9.5%**.
- Cleaned and processed over 100 dermatology documents, which were used to synthesize more than **780** English audio samples with varied accents and genders.

Data Engineering *Pipeline using Microsoft Fabric, Azure Services* Remote July 2024

- Optimized end-to-end data engineering pipeline to ingest, process, and visualize on-premise SQL Server data leveraging **Azure Data Lake**, **Azure Data Factory**, **Azure Databricks**, **Azure Synapse Analytics**, and **Power BI**
- Developed a **sentiment analysis** pipeline leveraging **Microsoft Fabric's Data Factory**, **OneLake**, **Synapse Data Engineering**, **Synapse Data Science**, and **Power BI**

Data Science Projects *AutoML, LLM, NLP, BERTopic* Canada May 2024 – July 2024

- Ranked in the **Top 1%** (15th out of 1,847) in the 2024 **KaggleX Skill Assessment Challenge**
- Ensembled **AutoML** frameworks (**AutoGluon**, **LightAutoML**) to predict used car prices
- Executed comprehensive topic modeling on 3,000+ multiple-choice questions using **BERTopic**, providing insights into educational content trends.

LiuXueWangXiao.com Founder, Programmer, Teacher *JS, Unity, Wordpress, L^AT_EX, Python (Scrapy), MongoDB* China Jun 2018 – Aug 2023

- Founded and led an online learning platform, enhancing user experience through data-driven design and interactive features.
- Spearheaded the development of educational tools (Virtual Reality Classroom, APPs), fostering engagement and improving learning outcomes.

- Implemented daily data scraping pipelines using Python (**Scrapy**) and **MongoDB** to extract, transform, and store blogs and news articles from various university websites efficiently.

COMPUTING AND LANGUAGE SKILLS AND TOOLS

Programming: Python, R, SQL, SAS, L^AT_EX, Azure Cloud, Javascript

AI / Data Analytics: AutoML (AutoGluon, LightAutoML, MLJAR), PyTorch, Keras, LoRA, QLoRA, RAG, BERTopic

Big Data, Cloud Technologies: Azure (AI Services, Data Factory, Data Lake, Databricks, Synapse Analytics, OpenAI, Power BI), AWS Machine Learning

Languages: English (fluent), Mandarin Chinese (native)

PROFESSIONAL CERTIFICATIONS & TRAINING

Passed all three levels of Chartered Financial Analyst (**CFA**) exams

EDUCATION

Graduate Certificate in Artificial Intelligence, Georgian College, Canada	Dec 2024 (expected)
Graduate Certificate in Marketing Analytics, Centennial College, Canada	May 2024
Ph.D., M.A. in Economics, University of Pittsburgh, United States	Nov 2015
B.A. with Honorary M.A. in Economics, University of Edinburgh, Scotland	Jun 2006