Diana Liu 202.739.1368; [lxjiao0805@gmail.com](mailto:lxjiao0805@gmail.com)

Proficient in mathematical statistics, econometrics, machine learning and deep learning. Offering 15 years of extensive project management expertise with a deep involvement in experimental design, data integration and cleansing, feature engineering, as well as mastery in machine learning, optimization, and deep learning implementations.

With extensive experience in leveraging advanced AI technologies, I have successfully utilized prompt engineering, Retrieval-Augmented Generation (RAG), and agents to develop robust applications that address complex business challenges, automate processes, and drive business growth.

**PROFESSIONAL EXPERIENCE**

**AI Data Scientist Freelancer** March 2023 – present

* Enhance product descriptions and recommendations through various recommender systems.
* Automate responses to customer inquiries and FAQ by retrieving relevant information from knowledge bases and historical conversations.
* Provide personalized recommendations based on user preferences and behavior in an agentic workflow.
* Analyze customer data to tailored marketing strategies and engaging with potential customers through personalized email campaigns and chatbots.

**Fidelity Investment** Fairfax, VA (remote)

*Principal Data Scientist* June 2020 – present

* Develop advanced machine learning models and optimization techniques to deliver personalized experiences and recommendations to Workplace Investing customers.
* Lead the design and development of Fidelity’s content inventory and generation product, utilizing ensemble search, prompt engineering, RAG (retrieval augmented generation), and various NLP practices.

**Fannie Mae** Washington D.C.

*Lead Quantitative Modeler* June 2020 – August 2022

* Apply NLP techniques for domain specific word embedding and NER using BERT/GPT-3 transfer learning, enhancing document search and recommendations via text summarization and similarity, and generate scoring algorithm to reflect the quality of review reports.

**The Gallup Organization** Washington D.C.

*Lead Data Scientist*  July 2012 – June 2020

* Lead Data Scientists and Data Engineers to predict customer engagement, conversion, transactions, and retention for Gallup Learning E-Commerce and conduct predictive modeling for clients through various machine learning models like generalized linear model with regularization, random forest, CatBoost/XGBoost, GBM and stacked ensemble models, with GridSearch/Bayesian Optimization to tune hyper-parameters.
* Leverage lead scoring, propensity score model, A/B testing, customer lifetime value prediction and other data science techniques for customer segmentation and digital marketing analytics to drive sustainable user growth.
* Apply NLP to understand markets for consumer-facing product and conduct novel qualitative research and open-ended questions analysis through topic modeling, semantic clustering and verbatim auto-coding.
* Support development and evaluation of internal survey automation pipeline which improves the efficiency of data extract, transform, load (ETL), anomalous detection, weighting and exporting.
* Collaborate with business development consultants to develop innovative winning solutions leveraging data science, conduct technical innovation to ensure statistical rigor in econometrics and modeling projects, and give internal team technical training on analytics topics like recommender system and imbalance data modeling.
* Build a query generator trained on a corpus of historical expert queries with Generative Adversarial Network (GAN) architecture for a government client.

**EDUCATION**

Professional Degree in Artificial Intelligence (Dec 2022); Stanford University

Master of Science in Statistics (May 2012), Overall GPA: 3.89; The George Washington University

Bachelor of Science in Mathematical Statistics (Jul 2010), GPA: 3.93, Rank: 1/120; Capital University of Economics and Business (Beijing, CHINA)

**TECHNICAL SKILLS**

Python, SQL, LangChain, crewAI, AWS, PyTorch, Keras, PySpark, H2O, TensorFlow, Apache Spark, SAS, Tableau