Casey Soh

515-587-1724 | kachun.soh@gmail.com | https://github.com/sohkachun | linkedin.com/in/casey-soh-442098b9/

EDUCATION

Boston College, MA

MS in Applied Analytics | CGPA: 3.76

- Coursework: Data Analytics, Computer Vision, ML Algorithms, Linear Algebra and Vector Calculus, NLP, Product Management, Big Data Econometrics, AI / ML Software Tools and Platforms

Drake University, IA

Bachelor of Science in International Business and Management | CGPA: 3.56

- Coursework: International Finance, Corporate Finance

SKILLS

- **Programming Languages:** Python (Intermediate), SQL (Intermediate), R (Intermediate)
- **Data Analysis & Forecasting Tools:** Power BI, Tableau, Excel, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
- **Deep Learning & Modeling**: PyTorch, TensorFlow, Transformers, CNNs, RNNs, ARIMA, ViT, Time Series Analysis
- **LLMs & Agentic Frameworks:** LangChain, LangGraph, OpenAI API, QLoRA, Prompt Engineering, RAG
- **Database & Infrastructure:** PostgreSQL, MySQL, BigQuery, Docker, Git, Azure, Google Cloud
- **Data Science Skills**: Machine Learning, Neural Networks, NLP, Computer Vision, EDA, Data Storytelling, Model Interpretability, MLflow
- **Soft Skills**: Team Collaboration, Problem Solving, Public Speaking, Time Management, Cross-Functional Communication, Analytical Thinking

Projects

LLM and Agentic Project (2025)

- **Designed and deployed a multi-agent system** using **LangGraph** to **automate customer support workflows**, enabling retrieval, updates, and cancellations of **flight**, **hotel**, and **car rental** information from structured **SQL databases**.
- **Developed an internal Retrieval-Augmented Generation (RAG) assistant** using **local LLaMA models** with **QLoRA adapters** and **OpenAI embeddings**, integrated via **LangChain** and deployed with a **Flask-based web UI** to retrieve **company policies**, **handbooks**, and **onboarding materials**.
- **Performed exploratory data analysis (EDA)** using the **OpenAI API**, applying **few-shot prompt engineering** to **automatically summarize datasets**, identify key **patterns**, and generate **natural language insights and visualizations**.

Skin Lesion Classification for Cancer Detection Using Deep Learning (2024)

- **Built a deep learning pipeline** to classify **skin lesions** using **EfficientNet-B0** and **ResNet-18** architectures with **PyTorch**.
- **Optimized classification accuracy** by integrating **image metadata** into a **multi-input neural network**.
- **Preprocessed images** with **augmentation techniques** and evaluated models using **partial AUC**, **precision**, and **recall**.
- **Tuned model architecture** and compared **bias-variance tradeoffs** to select the **best-performing model**, with all experiments **tracked using MLflow** for reproducibility and performance monitoring.

NLP Model Development for PII Detection (2024)

- Conducted **NER (Named Entity Recognition)** to classify and extract sensitive entities from unstructured text data.
- Designed and fine-tuned an NLP pipeline using **DBERTA** to detect Personally Identifiable Information (PII) in text datasets, incorporating **LoRA (Low-Rank Adaptation)** to improve efficiency and reduce computational cost.
- Achieved a **95% precision and 90% recall** through iterative training, validation, and hyperparameter tuning on a dataset of 10,000 documents, optimizing the model's performance.

Time Series Forecasting for Sales and Product Analysis (2023)

- **Conducted time series analysis** using **ARIMA models** to **forecast sales volume** and **predict product-level performance.**
- **Performed EDA** to uncover **seasonal patterns**, **cyclical behavior**, and **external drivers** influencing sales fluctuations.
- **Identified temporal trends** and anomalies to support **data-driven inventory planning** and **promotional strategies**.
- **Developed interactive dashboards in Power BI** to visualize forecasts and **communicate insights effectively** to **non-technical stakeholders** for strategic decision-making.

Work Experience

DLL, IA

Contract Administrator *Feb 2022- Jun 2022*

- Created a project planner for a team of 5 members to improve efficiency by allocating tasks based on individual experience and skillset, ensuring the timely completion of time-sensitive deliverables.
- Demonstrated leadership by taking on manager's responsibilities during her 2-week leave, collaborating effectively with teammates to ensure priorities were addressed and tasks were completed on time.