CAMILLE TRAN

📞 714-350-1220 ✉️ Cttran2021@gmail.com 🔗 linkedin.com/in/camilletran-datasci ⭘ github.com/Ctt011

Career Objective

With a June 2026 anticipated BS degree in Data Science, I have a solid foundation in machine learning and a passion for building AI applications on core development frameworks, including cloud platforms such as AWS and generative AI platforms. Seeking an entry-level role to utilize skills in Python for industry modeling applications of predictive analytics, diagnostic evaluation, operation optimization, or business-specific support. I bring a tactical approach, positive "can-do" attitude, and great communication skills.

Education

University of California - San Diego, Major GPA: 3.68   La Jolla, CA
Bachelor of Science in Data Science - Halıcıoğlu Data Science Institute   *Expected Graduation: Spring 2026*

Senior Project (Fall 2025 – Winter 2026): Developing an AI Digital Twin using LLM-augmented frameworks to enhance model accuracy, reduce hallucinations, and enable advanced "what-if" simulations for improved predictive tuning.

Technical Skills

Languages: Python, R, Java, C++, SQL, Scala, HTML/CSS, MATLAB

Tools & Frameworks: TensorFlow, PySpark, Databricks, Airflow, Kafka, Hive, NiFi, FAISS, Streamlit, Git, Jupyter, Excel, Tableau, Power BI, Linux

Cloud Platforms: AWS (S3, EC2, Bedrock, Glue, Athena, Aurora PostgreSQL, DMS, SCT, Timestream, Systems Manager, QuickSight), Azure (Logic Apps, SQL Database)

Certifications: AWS Cloud Practitioner (2023), AWS Data Engineer – Associate (In Progress)

Experience

US State Department - US Embassy London   **September 2023 – May 2024**
AI/Data Science Developer Intern
- Designed an AI-driven document retrieval pipeline for policy Q&A, integrating LangChain for text preprocessing and FAISS for fast vector search across government documents.
- Optimized embedding-based retrieval to improve query accuracy and reduce latency, enhancing policy search responsiveness.
- Participated actively in Scrum teams, during weekly and daily meetings collaborating effectively to meet project goals.

Projects

Cloud-Native Multimodal RAG System | *AWS Bedrock, FAISS, Streamlit*   **September 2025 - Present**
- Currently building a multimodal Retrieval-Augmented Generation (RAG) system combining textual and visual data to generate context-aware food recommendations.
- Leveraging AWS Bedrock, S3, and FAISS for cloud-native data storage, vectorization, and scalable similarity search.
- Developing a Streamlit-based chatbot that enables cross-modal search through Claude-Sonnet Multimodal LLM.

Real-Time Big Data Pipeline on AWS | *NiFi, Kafka, PySpark, Hive, Airflow, Tableau, QuickSight*   **July 2025**
- Designed and deployed a real-time data pipeline on AWS to process COVID-19 API data every 5 minutes, leveraging Apache NiFi for ingestion and encryption, and Kafka for distributed message streaming
- Implemented PySpark jobs to transform and clean streaming data, writing outputs to HDFS and exposing curated data through Hive external tables for analytics
- Automated pipeline scheduling using Apache Airflow, ensuring fault-tolerant execution across AWS EC2 nodes
- Delivered interactive KPI dashboards in Tableau and AWS QuickSight, visualizing confirmed, recovered, and fatal case trends by region

Large-Scale Data Processing Pipeline | *Azure SQL, Databricks, Power BI*   **May 2025**
- Built an end-to-end cloud-based data pipeline using the Medallion architecture (Bronze, Silver, Gold layers) in Databricks to process 1M+ rows of European water sensor data, performing scalable transformations, cleansing, and outlier detection with SQL/Scala and Spark.
- Automated ingestion with Azure Logic Apps and designed interactive PowerBI dashboards using DAX measures to monitor KPIs and visualize water quality trends and anomalies.

Shelf of Health: Food Processing Visualization | *Data Visualization, Statistical Analysis, Excel, Python*   **March 2025**
- Awarded Best Project Award, ranking in the top 3% of 180 students, for a creative and tangible visual design praised for its clarity and evocative storytelling.
- Developed an Excel-based visualization mapping grocery store shelves by food processing levels (NOVA classes), allowing users to intuitively see how "healthy" stores still stock mostly ultra-processed foods.