Tony Lei

Seattle, WA | (415) 519-9672 | tonyllei54@gmail.com | www.tonylei.me

EDUCATION

University of Washington, Seattle **Expected Graduation:** *March 2027*
Master of Science in Statistics

University of California, Los Angeles **Graduated:** *June 2024*
Bachelor of Science in Applied Mathematics, Minor in Data Science Engineering Cumulative GPA: 3.80

RESEARCH

Classifying Emotions from Images | Paper **Los Angeles, CA**
Research Assistant *Mar 2023 – Jun 2023*

- Designed an image-preprocessing pipeline that downloaded, filtered, and merged >100K images from 8+ CSV sources
- Built an image classification workflow that labeled all faces in images with one of eight emotions using PyFeat API
- Merged classified emotions with original dataframes to produce cleaned datasets ready for emotion analysis

TECHNICAL PROJECTS

LLM Japan Travel Assistant | Website **San Francisco, CA**
Individual Project *May 2025 – Present*

- Devised a full-stack (Next.js + FastAPI) app providing LLM-powered Japan travel guidance and recommendations
- Processed 83K+ r/JapanTravel Reddit posts and 619K+ comments into structured training data for LLM fine-tuning
- Fine-tuned Granite 8B (via MLX) using 50K Japan specific question/answer pairs to train a specialized travel assistant
- Engineered few-shot prompting strategies to improve conversational summarization and travel advice quality

Matcha Shop Knowledge Retrieval System **San Francisco, CA**
Individual Project *May 2025 – August 2025*

- Built a RAG application (Next.js+Chroma DB) supporting natural language queries across 10+ operational documents
- Enabled users to retrieve exact pages with top-k results using cosine similarity with accurate source attributions
- Designed a ChromaDB pipeline with chunking, HNSW indexing, and metadata tagging for fast and precise retrieval

Customer Inquiry AI **Los Angeles, CA**
Individual Project *May 2024 – Jun 2024*

- Developed a Flask application to handle customer SMS messages via Twilio API with automated response generation
- Implemented a vector database in Cassandra to store embeddings and retrieve relevant responses for customer inquiry
- Reduced customer response time from 48+ hours (manual) to <1 second for common inquiries (hours, menu, etc.)

Plane Ticket Price Prediction **Los Angeles, CA**
Project Lead *Jan 2023 – Mar 2023*

- Led a 7-person team to build a Random Forest based fare prediction model, achieving $100 average margin of error
- Automated daily web scraping pipeline using GitHub Actions, collecting hundreds of flight prices per run
- Built ETL pipeline processing 31GB dataset (6M rows, 20+ features) with Pandas for multi-source data integration

EXTRACURRICULARS

2022/2023 ASA DataFest at UCLA **Los Angeles, CA**
Finalist (Top 15/75), Team Lead *Apr 2022, Apr 2023*

- Competed in a 48-hour data-thon against 5 universities (300+ students), presenting findings to a panel of judges

SKILLS

Languages: Python, SQL, JavaScript, TypeScript, HTML/CSS
Frameworks/Libraries: Pandas, NumPy, PyTorch, React, Next.js, Flask, FastAPI, LangChain, MLX, Tailwind CSS
Tools/Databases: Git, Docker, ChromaDB, Cassandra