

JAEKANG LEE

+1 778-929-6940 · tonylee015@gmail.com · [leejaeka github](#)
[jaekangai_blog](#) · [jaekang-lee linkedin](#)

DATA SCIENTIST & AI LEAD

PROFESSIONAL EXPERIENCE

Canadian Food Inspection Agency Data Scientist

June 2021 - Present

Accomplishments:

- Built and deployed an image classifier that automatically labels seeds products with 97%+ accuracy, outperforming human experts by 7% and streamlining processes. Identified bottleneck with explainable AI, gradCAM. Solved with YOLO and background-removing algorithms and parameter tuning.
 - AI lead [Finesse project](#), introducing semantic search via embedded vector stores, modernizing document search engines beyond keyword searches for internal documents. Retrieving 130% more accurate results than previous search. Keywords- llama index, langchain, clustering, unsupervised
 - AI lead [Louis project](#), a ChatGPT-like virtual assistant with large language models using Openwebui frontend. Overcame users' unpredictable prompt engineering hurdles by implementing agentic flow using LangGraph, resulting in a 132% model correct output quality boost.
 - Built advanced NLP models for the secure processing of sensitive government documents, featuring a deep learning-based English-to-French translator and zero-shot text classification capabilities.
-

EDUCATION

University of Toronto

Bachelor's Degree, Major in Mathematics, Minor in Computer Science and Statistic

Relevant Coursework: Algorithm, Data Structures, Probability and Statistic, Software Design, Computer Vision, Calculus, Data Analysis, Linear Algebra, Bayesian Statistic

Udacity

Data Scientist Nanodegree

Relevant Projects: ML blog post, Disaster Response App, Recommendation System with IBM, Jane-Street-Competition

Certificate

[MOOC CERTIFICATES - \(COURSERA, KAGGLE\)](#)

SKILLS

Data Science & Machine Learning

MLOps, LLM, Langchain, RAG system, ETL(Extract, Transform, Load), Data Science pipeline (cleaning, wrangling, visualization, interpretation), SQL, Tableau/Looker, Statistical Inference, Spark, Tensorflow, Flask, A/B testing, Deep Learning, Computer Vision, Recommendation System, Prompt Engineering, Agents

Tools

Python, Git, Docker, MS Office, Latex, Web deployment, OOP, APIs, Bootstrap, Jupyter, API, React.js, Spreadsheet, Azure, D3, Plotly, Matplotlib, Seaborn, Numpy, Pandas, SQLAlchemy, BeautifulSoup, NLP, Feature Engineering, Data Augmentation, Database, HADOOP

PROJECT

[Incrementum \(UofTHacksVIII\)](#)

- Winner of the Best use of Google Cloud award at the University of Toronto Hackathon VIII as a team of three
- Financial Productivity Web Application with Generative Deep Learning that rewards users with GAN generated pixel trees
- Implemented WGAN-GP then deployed backend Flask REST-API that generate and return pixelated tree on GET by frontend

[Jane Street Market Prediction \(Google Data Science Competition\)](#)

- Kaggle competition to maximize market return. Worked with 6GB of data to build exploratory data analysis and prediction scripts.
- Scored of 9443.499 (249th place out of 3,616 competitors and 41,274 entries) using Multilayer perception(classic Neural Network)

[Non-Medical Use of Prescription Drugs \(ASA DataFest at UCLA\)](#)

- Successfully answered two research questions and generated suggestions to improve the efficiency of future surveys and focus.