James Liang

james.liang.cje@gmail.com | github.com/pooky1955 | linkedin.com/in/jamesyueliang | youtube.com/c/pookycodes

EDUCATION

Diploma of College Studies *(DCS)* 2021 – 2023
Honours Pure and Applied Sciences Marianopolis College
Extracurriculars: Marianopolis AI Club (founder)

Secondary School Diploma *(DES)* 2016 – 2021
Collège Jean-Eudes

SKILLS

Machine Learning: Python, Scikit Learn, Tensorflow, Keras, OpenCV, NLTK, Pandas, Kaldi, R, Pytorch
Frontend Development: React, HTML, CSS, Javascript
Backend: Node.js, Flask, Go, C++, Java, SQL, Linux
Languages: French, English, Cantonese, Mandarin

EXPERIENCES

Software Developer | University of Montreal – LabTNS 2021 –
Developed OCRx, a platform to standardize drug terminologies (drug ontology)

- Developed and designed a website using **React** and **Material UI** allowing users to search clinical concepts such as substance, form, and route of administration and obtain their standardized information.
- Developed a **Java Spring** backend exposing a **REST API** and integrated it with the C++ library FaCT++ Reasoner.
- Built a search field with auto-complete that communicated with the backend in real-time without divulging internal information by restricting the number of results returned.

Project Developer | University of Toronto Machine Intelligence Student Team 2021 –
Significantly reduce image noise and data usage in video calls using deep learning, image to image translation of spectrograms, and speech recognition [Website].

- Reviewed literature on audio resolution and used Wav2Vec to perform **real-time speech recognition**.
- Built **personalized speech synthesizer** using Tacotron and speaker embeddings generated from 15 second audio clip.
- Preprocessed data using **Pytorch** and audiomentations to acquire Mel Spectrograms and noisy versions of audio.
- Worked on using **MaskCycleGan-VC** to generate robust speech from noisy audio and **wav2vec** embeddings.
- Built a website using React and Flask where users can upload audio and download the denoised version.

Data Science Intern | Boehringer Ingelheim Canada Summer 2021
Optimized marketing email schedule using historical data to increase click-through rates and open rates.

- Preprocessed dataset using **R** scripts and **Pandas** from raw, untabulated data.
- Compared performance of various machine learning models such as **XGBoost**, **Random Forest**.
- Used state of the art **machine learning methods** (multitarget encoding, Linear Discriminant Analysis, SMOTE) to predict optimal send time with a top 3 accuracy of 80% (**2.6 times higher** than baseline model).

Biostatistics and Data Science Summer Intern | Boehringer Ingelheim Canada Summer 2020

- Created an **RShiny** app that allows users to download and filter clinical trials from clinicaltrials.gov. [Website]
- Created Python programs that use **NLP** and **Google's Tesseract** to extract specific features from scanned informed consents and categorize them according to their features.
- Predicted clinical trial success using inclusion and exclusion criteria with **NLP** techniques (Word2Vec, TF-IDF).
- **Contributed to Synthea**, an open-source synthetic patient generator by generating localized (Canadian) demographics data (using Statistics Canada open datasets). [GitHub]

Software Developer | Inspiratiq [GitHub] [Website] Winter 2020

- Developed and deployed a website on **Netlify** for Inspiratiq's artwork using **Gatsby** and **Material UI,** and **GraphQL**.
- Integrated payment within the website using **Stripe** Checkout Sessions using **serverless Netlify Functions**.
- Automated sending of order confirmation emails with Python using the Simple Mail Transfer Protocol (SMTP).

Winner at Pharmahacks 2020 [Devpost] | Hackathon (MILA, Montreal) November 2019

- Experimented with **XGBoost**, **Naive Bayes**, and NLP techniques, such as **Word2vec** and **TF-IDF.**
- **Deployed** ML algorithms to production using a **Flask REST API** to predict the probability of success of clinical trials.

PROJECTS

- Created *Summarize And Ask*, a website that summarizes articles or texts using **MobileBERTSUM** and **BERT** to answer questions about the inputted text [GitHub].
- Created *iNews*, a **deep learning model** to guide the users into detecting fake and real news leveraging stance detection and applying **Word2vec embeddings** and **multi-channel convolutional architecture,** achieving 85% accuracy vs 82% from state-of-the-art [*Montreal Regional Science Fair canceled due to the 2020 pandemic*] [GitHub].