James Wu

jian.wu.22@ucl.ac.uk | jswu18.github.io
+44 7951 633010

RESEARCH EXPERIENCE

INTEGRAL PROBABILITY METRICS — UCL
Research Assistant (Msc.) — June 2022 - Present
JAX

- Explored kernel-based distribution discrepancies such as the **MMD** (Maximum Mean Discrepancy) and **KSD** (Kernel Stein Discrepancy)
- KSD Blog Post
- GitHub Repository

NATURAL LANGUAGE PROCESSING — University of Toronto
Research Assistant (BASc.) — Sept 2018 - Apr 2019
PyTorch

- Achieved state-of-the-art performance for named-entity recognition of biomedical literature using **transfer learning** with **BERT** models
- Research Link

WORK EXPERIENCE

REVOLUT LTD. — London, UK
Machine Learning Engineer — Aug 2019 - Sept 2022

Bayesian Time-Series Forecasting: Gaussian Processes
Pyro, PyTorch, BayesOpt, Airflow, GCP
To earn rebates of $100M+/year and ensure Revolut's profitability:

- Managed global card issuance for Revolut users (>20M)
- Developed a generalised time-series forecasting pipeline with **Gaussian Processes** to predict activity, spending, and growth
- Integrated business risks from newly emerged COVID-19 behaviours
- Published models to internal **PyPi** and used by other projects across the business (i.e. liquidity forecasting for Treasury)

NLP Text Embeddings: BERT Sentence Transformers
PyTorch, MLflow, Elasticsearch, Scikit-Learn, Airflow
To deploy self-serviced customer support and improve live agent workflows:

- Developed semantic search and intent recognition of customer chat messages with **one-shot learning** and **transformer-based embeddings**, integrated into real-time services
- Improved text embedding quality with **multi-tasked learning** to leverage multiple corpora during training

User Behaviour: Recurrent Neural Networks
TensorFlow, PySpark, Dataproc, Airflow, GCP
To personalise user experience and improve retention:

- Developed **LSTM** solution to predict spending behaviours at a user level
- Guided personalised content by visualising model predictions with **t-SNE** to cluster user behaviours

ANALOG DEVICES INC. — Toronto, CA
Data Scientist — May 2017 - Aug 2018

Person Tracker for In-Home Monitoring: Algorithm Analysis
openCV, Scikit-Learn

- Built a GUI to analyse and identify corner cases for **computer vision** algorithms, improving model performance by ~20%

EDUCATION

UNIVERSITY COLLEGE LONDON
MSc. Computational Statistics & Machine learning
In Progress

UNIVERSITY OF TORONTO
BASc. Engineering Science
Graduated with Honours

SOFTWARE SKILLS

Strong Proficiency:
Python • SQL • Git

Intermediate Proficiency:
PySpark • Kubeflow • Airflow
Docker • GCP • MATLAB • Latex

Familiar:
C • PIC Assembly

RELEVANT STUDIES

Probabilistic & Unsupervised Learning

Convex Optimisation

Machine Learning & Data Mining

Computer Vision

Algorithm Design & Analysis

PUBLICATIONS

Power Optimization Using Embedded Automatic Gain Control Algorithm with Photoplethysmography Signal Quality Classification (ICASSP 2020)

Robust Beat-To-Beat Detection Algorithm for Pulse Rate Variability Analysis from Wrist Photoplethysmography Signals (ICASSP 2018)

Development and Validation of a 3D-Printed Neuronavigation Headset for Therapeutic Brain Stimulation (Journal of Neural Engineering 2018)

PERSONAL INFO

Canadian Citizen

Native English Proficiency