Jonathan Siu Chi Lim

+1 5148137956 | jonathan.lim@u.nus.edu | linkedin.com/in/jonathanlimsc | github.com/jonathanlimsc | jonathanlimsc.com | Toronto, Canada

Education

Université de Montréal / Mila Quebec AI Institute — Montreal, Canada
Masters of Science in Computer Science (Machine Learning) – GPA 4.2/4.3 (4.0/4.0) — *Sep. 2022 – May 2024*

- Mila Quebec AI Institute MSc Machine Learning program. Program supervisor Prof. Yoshua Bengio
- Research topic: Molecular Representations in Drug Discovery with Deep Learning
- Awarded Mitacs Accelerate research grant (valued at $30,000)
- Awarded UdeM scholarship (valued at $20,950)

Stanford University — California, United States
Management Science and Engineering (Exchange program) — *Aug. 2015 – Aug. 2016*

- Selected for NUS Overseas College (NOC) Program in Silicon Valley, studied at Stanford University and worked at a tech startup (UpGuard) for 1 year as a software engineer

National University of Singapore — Singapore
Bachelor of Science in Computer Science (Honours, Distinction) — *Aug. 2013 – Dec. 2017*

- Awarded NUS Business Dean's Scholarship (valued at $63,450)

Research

Gotta Be SAFE: A New Framework for Molecular Design
Emmanuel Noutahi, Cristian Gabellini, Michael Craig, **Jonathan Siu Chi Lim**, Prudencio Tossou
Royal Society of Chemistry, Digital Discovery Journal 2023. NeurIPS 2023, AI4D3 Workshop, AI4Mat Workshop

Brain disease classification based on routine MRI using a Convolutional Neural Network Algorithm
Jonathan Lim Siu Chi, Liu Yaou, Joe Wu Zhen Zhou, Yunyun Duan, Peiyi Gao, Yongjun Wang
BioMind, 2018 (Company internal research report)

Experience

Machine Learning Researcher — Feb 2024 – Present
Mila Quebec AI Institute / CERC-AAI Lab — *Remote, Canada*

- Researching into multimodal generalist agents, foundation models and large language models (LLM) with Prof. Irina Rish at the CERC-AAI Lab
- **Skills:** LLMs, SLURM, multi-node-multi-GPU training, HuggingFace, Pytorch

Machine Learning Research Engineer (Co-op) — May 2023 – Apr. 2024
Recursion Pharmaceuticals (Valence Labs) — *Montreal, Canada*

- Co-authored NeurIPS 2023 workshop paper: "Gotta be SAFE: A New Framework for Molecular Design". Trained LLMs for drug molecule generation.
- Designed and built vector DB infrastructure for literature search and molecule hit expansion, enabling retrieval-augmented generation for LOWE (LLM for drug discovery).
- Researched multimodal models fusing text, molecule graph structure and phenomics, improving molecule property prediction to filter drug candidates
- **Skills:** LLMs, Langchain, HuggingFace, PEFT, Pytorch, SLURM, multi-GPU training, Weaviate Vector DB

Data Scientist (Part-time Contract) — Jun. 2023 – Dec. 2023
Binance — *Remote, Canada*

- Researched and trained fraud detection models using GNNs such as GraphSAGE and RGCN, resulting in a 5% improvement of fraud recall rate at 95% precision, saving thousands of customers from potentially fraudulent transactions.
- **Skills:** GNNs, Spark, AWS Sagemaker, Pytorch, Deep Graph Library

Senior Machine Learning Engineer Apr. 2021 – Jun. 2022

StarHub *Singapore*

- Architected and implemented machine learning pipelines on GCP Vertex AI and AWS Sagemaker, enabling continuous integration and deployment (CI/CD) of ML models
- Developed user and website domain interest graph for entity understanding and relationships, providing users recommendations based on their interests
- Built user embedding models based on web browsing behaviour for fast vector similarity search using autoencoders and graph convolutional networks (GCN), to identify similar user groups for ads targeting
- **Skills:** GNNs, Pytorch, PyG, Deep Graph Library, AWS Sagemaker, AWS EMR, GCP Vertex AI, CI/CD, Neo4J, FastAPI, Flask, Spark, Airflow

Senior Machine Learning Engineer Nov 2018 – Mar 2021

Shopee *Singapore*

- Architected and implemented a distributed, scalable machine learning platform for model training to deployment on Kubernetes, enabling the Data Science team to deploy models
- Productionized Machine Translation microservices for multiple languages, including English, Chinese, Bahasa Indonesia, Thai and Vietnamese, based on Fairseq
- Developed NLP Sentiment Analysis models based on BERT Transformer architectures for customer chats with up to 80% precision
- Developed an Anti-Fraud identity card detection model based on YOLO neural network architecture with >90% precision and recall
- Productionized a multi-GPU OCR inference pipeline into a live API microservice with Docker containers, high-availability architecture and automatic failover
- Developed Customer Life-time Value models using user behavior data with XGBoost and model stacking, feature engineering using Spark and Scikit-learn, achieving >90% accuracy
- Developed a centralized logging platform for real-time API monitoring using Elasticsearch, Logstash and Kibana, increased visibility and prevented model drift
- **Skills:** Computer Vision, NLP, LLMs, GCP Kubernetes, Docker, GitLab CI/CD, Tensorflow, Pytorch, HuggingFace, Fairseq, Sklearn, Elasticsearch, Logstash, Kibana, Jenkins, Horovod multi-GPU training, FastAPI, Flask

Machine Learning Research Engineer Jan 2018 – Aug 2018

Biomind *Singapore*

- Trained and developed a CNN model based on DenseNet to classify 3D MRI brain scans into disease categories (Normal, Tumor, Ischemic blood vessel disease and Multiple Sclerosis). Achieved overall >85% sensitivity and >95% specificity on proprietary dataset.
- Co-authored an internal company research paper: "Brain disease classification based on routine MRI using a Convolutional Neural Network Algorithm"
- Built a CNN auto-encoder based on a 2D UNet architecture for style-transfer learning between T1-weighted and T1- Contrast MRI brain scans, solving shortage of labelled T1-Contrast scans through synthetic data generation
- **Skills:** Computer vision, Tensorflow, Docker, Python, DICOM image processing

Technical Skills

Languages: Python, SQL, Javascript, HTML/CSS
Machine Learning: Pytorch, DeepSpeed, HuggingFace, PEFT, Tensorflow, Pytorch-Geometric, Scikit-learn, Pandas, Weaviate Vector DB, Langchain
Software Engineering: FastAPI, Flask, Redis, OpenSearch, Elasticsearch, Logstash, Kibana, PostgreSQL, MongoDB
Data Engineering / ETL: Airflow, Spark, Hadoop
MLOps/DevOps: SLURM, Docker, Kubernetes, MLFlow, Weights&Biases, CometML, AWS Sagemaker, GCP VertexAI, GitLab, Jenkins