CHRISTOPHER WANG
1-613-986-4968 | christopher.jj.wang@gmail.com | https://github.com/Christopher-Wang

EDUCATION

Software Engineering - Bachelor of Engineering | Carleton University | Ottawa, Ontario **09/2016 – 08/2021**
• **Cumulative GPA:** 11.88/12 (A+) | **Chancellor's Medal** (rank 2 of 795 undergraduate students) | Dean's List
• **Student Involvement**: Carleton AI Society (President), Carleton Kaggle Club (President)

RESEARCH EXPERIENCE

Undergraduate Researcher | Carleton University | Ottawa, Ontario (Part Time) **05/2018 – 04/2021**
• Applied unsupervised image segmentation by backpropagation models to multi-channel satellite imagery
• Conducted hierarchical remote sensing on environmental satellite imagery using unsupervised random forests as an embedding algorithm on pixel-wise channel features and agglomerative clustering, achieving an accuracy of ~70%
• Presented a seminar on Git to non-technical collaborators, emphasizing the value of proper development practices

Research Scientist Intern | Nuance Communications | Montreal, Quebec (Remote) **05/2020 – 08/2020**
• Implemented homomorphically encrypted representation search using Face-Net embeddings and Microsoft SEAL
• Enabled efficient encrypted dot product by exposing the sum_vector operation in the TenSEAL open source library
• Experimented with polynomial activation functions for fully encrypted convolutional neural network inference
• Performed literature review on private machine learning inference with a focus on cryptographic solutions

Undergraduate Researcher | Carleton University | Ottawa, Ontario (Part Time) **09/2019 – 04/2020**
• Fine-tuned pretrained Resnet-18 models to orca whale call spectrograms supplied by Fisheries and Oceans Canada and the INTERSPEECH 2019 ComParE Orca Activity Subchallenge achieving an AUC ROC of 0.89
• Developed a gradient-boosted tree model trained on short-window audio features achieving an AUR ROC of 0.72

Undergraduate Researcher | Carleton University | Ottawa, Ontario (Full Time) **05/2018 – 12/2018**
• Created an ensemble of U-Net Architectures of varying feature extractors (inception modules, residual blocks, dense blocks, attention connections, etc) for left atrial MRI segmentation attaining a Dice score of 0.92%
• Developed a framework for rapid prototype testing of image segmentation models with data augmentation (translation, rotation, gaussian blur, elastic transform) for 2D and 3D MRI images built on Keras/TensorFlow

PAPERS & PATENTS

Systems and methods for dynamic demand sensing and forecast adjustment (ongoing patent draft)	**Present**
Systems and methods for dynamic demand sensing (US20210110413A1)	**12/2022**
Systems and methods for features engineering (US20230086226A1)	**09/2022**
Reciprocal perspective as a super learner improves drug-target interaction prediction (Scientific Reports)	**08/2022**
Canadian Jobs amid a Pandemic (IEEE COMPSAC)	**07/2022**
U-Net Architecture Ensembles for Left Atrial Segmentation (SPIE Medical Imaging)	**06/2019**

AWARDS, FUNDING & SCHOLARSHIPS

Chancellor's Medal (Awarded to the undergraduate student with the second-highest overall average)	**10/2021**
Internship-Carleton University Research Experience for Undergraduate Students ($2250)	**09/2020 – 04/2021**
Carleton Academic Scholarship ($4000)	**09/2020**
Internship-Carleton University Research Experience for Undergraduate Students ($2250)	**09/2019 – 04/2020**
Carleton Academic Scholarship ($4000)	**09/2019**
Internship-Carleton University Research Experience for Undergraduate Students ($2250)	**09/2018 – 04/2019**
Carleton Academic Scholarship ($4000)	**09/2018**
NSERC Undergraduate Student Research Award ($6300)	**05/2018 – 08/2018**
Carleton Academic Scholarship ($4000)	**09/2017**
Carleton Academic Scholarship ($4000)	**09/2016**
General Dynamics Scholarship ($5000)	**09/2016**

WORK EXPERIENCE

Machine Learning Engineer | Kinaxis | Ottawa, Ontario (Full Time) **05/2021 - Present**

• Designed and implemented, in a small cross-functional tiger team, a retrieval-augmented generative AI backend system using the OpenAI API, FastAPI, Azure Cognitive Search, and MongoDB, with automatic CI/CD integration using Docker, Helm, and Github Actions, deployed to Azure Kubernetes, within a single sprint (two weeks)
• Crafted a novel New Product Introduction (NPI) algorithm for demand forecasting using the PySpark DataFrame API, increasing the accuracy for new product forecasts by 35% (from 0.85 WMAPE to 0.50 WMAPE)
• Architected a data segmentation module using the Databricks Job API, PySpark, and Argo Workflows, enabling parallel execution of downstream modules like inference and training, allowing for the processing of terabyte-scale data, resulting in a 8.19x end-to-end speed up (from 131 hours to 16 hours), 18.84x reduction in data storage (from 6.8 TB to 361GB), and a 10.11x reduction in yearly cloud costs (from 1,377,012$ to 136,084$)
• Integrated multiple machine learning modules into a single unified Argo Workflow for streamlined deployment
• Led end-to-end ownership of the forecast explainability feature based on SHapely Additive exPlanations (SHAP)

Machine Learning Developer Co-Op | Kinaxis | Ottawa, Ontario (Full Time) **01/2019 – 12/2019**

• Prototyped various architecture designs for an automatic time series forecasting using technologies like Apache Kafka, Spark, Dask, Pandas, LightGBM, PrestoDB, Amazon S3, and Postgresql/PostGIS on a Kubernetes Cluster
• Operationalized the time series AutoML architecture using cloud technologies like Docker, Kubernetes, and Helm, following SOLID principles, agile development, and Python best practices, bringing the prototype to customer beta
• Experimented with neural networks (LSTMs, 1D convolutions), ensemble tree models, SVM regressors and feature selection techniques (genetic algorithms, mutual information, Boruta) for times series forecasting
• Introduced Dask on Kubernetes for parallelizing python jobs, resulting in linear speed ups (up to 16x)

RELEVANT PROJECTS

Projection Efficient Intrinsic Dimensionality | Personal Research **05/2023**

• Experimented with LOw Rank Adaptation (LORA) style down projections prior to learning with intrinsic dimensionality, reducing the dense projection size by over 15x with negligible impact to accuracy on MNIST dataset

Implicit Neural Neural Representation (INNR) Models | Personal Research **01/2023**

• Encoded several layers of a Densenet121 CNN trained on CIFAR10 dataset in an implicit neural representation (INR) network for weight compression, reducing parameter size by over 40% with negligible impact to accuracy

Inebriation Prediction using Keystroke Dynamics | Personal Research **10/2021**

• Developed a web application using Vue.js and flask to collect keystroke dynamic data of inebriated vs sober users
• Applied a neural network to classify inebriation, achieving an accuracy of 95.83% (this was a fun project for me)

Transformers for Protein Function Prediction | Capstone **04/2021**

• Formulated protein ontology prediction as an eXtreme MultiLabel Classification problem, applied probabilistic label trees and fine-tuned ProtBERT models beating the DeepGoPlus model baseline across all ontology domains

SHapely Additive exPlanations for Metric Explainability | Personal Research **04/2020**

• Developed a novel formulation of Kernel SHapely Additive exPlanations for probabilistic and bayesian model explanations using a Kullback–Leibler divergence metric, submitted to CCAI 2020 but was weakly rejected

Deep Learning by Explaining: The Deep Feynman Technique | Personal Research **02/2020**

• Prototyped a U-Net based generative adversarial network to approximate guided-backpropagation explanation outputs of a teacher network for the purpose of fast explanation generation and for a novel knowledge distillation algorithm that conditioned a student network on its teacher's explanations in addition to the teacher's logits

CANSSI National Case Study Competition | National Student Kaggle Competition **09/2019**

• Achieved first place at the local chapter level by applying bayesian hyperparameter optimization, time-series feature engineering, HSIC Lasso feature selection, and ensembles of CatBoost, XGBoost and LGBM models

• Achieved second place at the national level poster competition presented at Simon Fraser University

Jigsaw Unintended Bias in Toxicity Classification | Kaggle Competition **06/2019**
• Achieved Bronze (top 10% of competitors, placing 195th of 3165 competitors) for toxic comment classification with BERT fine tuning, model ensembling, dynamic mini-batching, and proportioned head/tail text truncation

DistriFlow | Software Development **02/2019**
• Created a Node.js module that implements federated learning and asynchronous stochastic gradient descent for distributing neural network training on web browsers, built using Typescript, socket.io and TensorFlow.js

TECHNICAL SKILLS

• **Programming Languages:** Python, Java, SQL, R
• **Analytics/ML Libraries:** Pandas, Pytorch, PySpark, SciKit-Learn, LightGBM, HuggingFace, Keras, TensorFlow
• **Deployment Technologies:** Docker, Kubernetes, Helm, Azure, GCP, Databricks, Argo, GitHub Actions, SLURM