Hillary Ngai

UofT Computer Science Master's Student & Vector Institute Researcher

hngai@cs.toronto.edu | linkedin.com/in/hillary-ngai | github.com/HillaryNgai

- Passionate about researching natural language processing, computer vision, and deep learning

SKILLS

Proficient:

Python	Pandas	SQL
Matplotlib	PyTorch	AllenNLP
TensorFlow	Keras	sci-kit learn
NLTK	spaCy	OpenCV
R	C++	C#
MATLAB	Flask	Git
JSON	JavaScript	HTML/CSS

PROJECTS

Anxiety and Depression Detection on Social Media

March 2020 – April 2020

- Developed various state-of-the-art language representation models—**transfer learning** with **BERT**, **RoBERTa**, and **XLNet**—in **PyTorch** to detect anxiety and depression from Reddit text
- Achieved **92.8% accuracy** and **92.4% F1 score** using **XLNet**

Colorectal Cancer Tissue Image Classification

March 2020 – April 2020

- Built state-of-the-art **CNN**-based model to classify 8 categories of colorectal cancer tissue in histological images using transfer learning with **EfficientNet** in **PyTorch**
- Achieved **96.4% accuracy**

AWARDS

Dean's Honor List

Sept 2019 – Dec 2019

4.0 GPA, **Rank 2** in class

President's Scholarship of Distinction

June 2015

98% entrance average

President's Research Award

2018 & 2019

Machine learning research

INDUSTRY EXPERIENCE

Ultimate Software — Data Scientist

May 2019 – Aug 2019

San Francisco, US

- Built text-based emotion detection model in **Python** using **PyTorch**, **Pandas**, and **AllenNLP**; **improved accuracy** by **28%** using **transfer learning** with state-of-the-art language representation model, **BERT**
- Designed, developed, and tested BERT pre-training **RESTful service** for the data science team in Python and **JSON** using **TensorFlow**, and **Flask**
- Collected, cleaned, analyzed, and presented raw emotion-text data from **MTurk** in Python and **HTML/CSS** using **spaCy**, **NLTK**, Pandas, and **Matplotlib**
- Iteratively **collaborated** with data science team and senior director to evaluate and redesign text-based emotion detection model

Bell Canada — Data Scientist

Jan 2019 – Apr 2019

Toronto, CAN

- Created **model evaluation API** including lift, feature impact, and r-squared analysis; **reduced evaluation time** from **~13 minutes to 30 seconds**
- Designed, developed, and documented **machine learning pipeline** for business optimizations in Python and **SQL** using **scikit-learn** and Pandas

Teledyne DALSA — Computer Vision Developer

May 2018 – Aug 2018

Waterloo, CAN

- **Improved computational efficiency** of **hyperspectral imaging system** by **700%** and **reduced RAM usage** by **50%**
- Designed, developed, documented, and tested hyperspectral imaging **API** in **C++** and Python using **OpenCV**
- **Publicized** hyperspectral imaging system in several **AI/computer vision** meetups to clients and data science professionals

RESEARCH EXPERIENCE

Vector Institute of AI— NLP Researcher

April 2020 – June 2020

Toronto, CAN

- Developed COVID-19 **question-answering system** using transfer learning with **SBERT** in Python using PyTorch and **Transformers** for **Kaggle competition**
- Publicized question-answering system at **Vector Institute's NLP symposium**

University of Waterloo — NLP Researcher

Sept 2019 – Dec 2019

Waterloo, CAN

- Fine-tuned BERT model to predict toxic/irrelevant comments in a social network in Python using PyTorch and AllenNLP; achieved **88% accuracy**

EDUCATION

University of Toronto — MSc in Computer Science

Sept 2020 – April 2022

Toronto, CAN

- Affiliated with **Vector Institute for Artificial Intelligence**

University of Waterloo — Biomedical Engineering

Sept 2015 – April 2020

Waterloo, CAN

- Achieved **4.0 final-year GPA** and **ranked 2nd** in final year