

Evan Pai

(647) 622-4699 | ei2pai@uwaterloo.ca | evanpai.com | linkedin.com/in/evan-pai | github.com/Wubaboo

Experience

Thomson Reuters - Data Science Intern Jan 2021 - Aug 2021

- Built a HTML **content extraction system** for use in a **fake news detection** pipeline by cross-referencing results from multiple Python libraries
- Extended the backend of an **automated curated newsletter**, informing subscribers of relevant industry trends while **logging user engagement** using **AWS DynamoDB, Flask** and **SparkPost**
- Evaluated and documented machine learning methods for **parsing and extracting HTML content**
- Wrote Python scripts to modify metadata and validate strings to assist in a **facial recognition** and **image captioning pipeline**

Statistics Canada - Data Science Intern Sept 2019 - Dec 2019

- Developed Python tools for a standardization of English and French metadata variables using **Natural Language Toolkit** to reform variables from a **many-to-many relationship** to a **one-to-one relationship**
- Created a **PyQT5 GUI** used to **query survey database keywords** based on associated semantics
- Managed a metadata database with **50,000+** entries and added components to automate procedures

Sigma Corporation - Data Analyst Feb 2019 - Apr 2019

- **Predicted manufacturing times** of operations for over **400+** parts using regression, resulting in an optimized production schedule and **projections of labour costs** for all operations

Projects

ytSheetMusic  Mar 2023

AWS EC2+S3, React, and Flask Web App to convert sheet music from YouTube URL to PDF

- Identifies and isolates regions of interest using **OpenCV Contour Detection**, merging results to PDF
- Prevents sheet music duplicates by calculating **structural similarity index** between images
- Reduced result processing time from ~1 minute to instantaneous output by caching images in **AWS S3**

DataBake  Aug 2020

PyQT5 GUI and SQLite Database for baking needs

- Tracks ingredient inventory and costs based on sales and expenses
- Records recipes and computes costs of ingredients to produce batches
- Creates graphical visualizations to summarize sales, profit margins, and bestselling products

Education

University of Waterloo 2018 - 2022

Bachelor of Mathematics in Computational Math

Minor in Computing, Minor in Combinatorics and Optimization

Skills

Languages: Python, JavaScript, C, C++, R

Technologies: SQL, NumPy, pandas, PyTorch, OpenCV, Spark, AWS, React