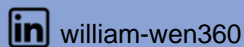


# William Wen



## SKILLS

- **Languages:** Python // C++ // C // Java // SQL // JavaScript // HTML // CSS
- **Technologies & Frameworks:** Flask // Docker // Git // React // GCP // Unittest // Pandas // Beautiful Soup
- **Machine Learning:** PyTorch // TensorFlow // Scikit-Learn // SpaCy // Rasa Bot Framework // NLTK

## EXPERIENCE

### Soapbox Innovations | Machine Learning Engineering Intern

Jan 2020 – Apr 2020

- Wrote **Flask REST APIs** to provide ML services like chatbots for webapp with **100,000+** users across **300+** companies
- Developed a chatbot using **CRFs** to perform **NER**, achieving **95.2%** intent and **85.3%** entity F1-Scores
- Applied transfer learning using **SpaCy's** word-embeddings and CNNs on to-do list items to differentiate between the assigner and assignee of tasks, achieving an **82.8%** F1-Score and later entering production for **2600+ teams**
- Developed backend service to internally crowdsource training data using Flask and **SQLAlchemy**
- Configured Docker to containerize services together including Flask, Unicorn, MySQL and **Unittest**

### Vision and Image Processing Lab | ML Research Assistant

May 2020 – Jun 2020

- Trained and hyper-parameter tuned a **U-Net** in **PyTorch** to perform semantic segmentation on the **PascalVOC** dataset according to Brox et al.'s "U-Net: Convolutional Networks for Biomedical Image Segmentation" paper
- Wrote scripts to evaluate differences in performance for each class when 1 or 2 classes are removed
- Read research papers and experimented with different segmentation models to determine what to use for the project

### Flatten Inc. | Software Engineer – Maps & Data Team | Visit at [Flatten.ca](https://flatten.ca)

Mar 2020 – May 2020

- Created **Python** scripts to extract self-reported COVID-19 cases from an online form filled by **400,000+** participants
- Performed analysis on acquired data to extract insights to display to users or repurpose for ML using **Pandas**
- Assisted in creating choropleth map that displays confirmed cases using **React-Leaflet**, a mapping library
- Wrote backend REST APIs to create cookies and track IPs with **Express.js** to maintain form data integrity

### Hikedu Inc. | Software Engineering Intern

May 2019 – Aug 2019

- Designed an information collection process via online forms to automate the international exchange approval process for **~17,000** Lanzhou University students using **React**
- Created a dashboard for teachers to track their students' academic performance with **Ant Design**, a React Library
- Coordinated with backend developers to create REST APIs to handle all requisite student / university information

## PROJECTS

### Manga Colorizer | [doc](#)

Apr 2020

- Built an encoder-decoder convolutional neural net to automatically color Japanese Comics (Manga) using **PyTorch**
- Trained CNN with **~160,000** images, preprocessed using PyTorch transforms and **OpenCV** image manipulation tools
- Optimized the net using a **discriminator (GAN)**, the CIELAB color space, and correct identification of character tags
- Utilized implementation described by Furusawa et al.'s paper, "Comicolorization, Semi-Automatic Manga Colorization"

### Privacy Policy Index | [doc](#)

Sept 2019

- Used **Scikit-Learn's** Linear Regression model to analyze the words used in privacy policies and score their strength out of 5, achieving a standard deviation of **+/- 0.5**
- Employed the **Porter Stemmer** and a **TF-IDF vectorizer** to create a domain-specific vocabulary
- Preprocessed texts from 50 privacy policies using NLTK methods to remove stop words and punctuation

## EDUCATION

University of Waterloo | Candidate for Bachelor of Software Engineering

Sept 2018 – Present