

# FLAWNSON TONG

I'm a 20 year old Machine Learning Researcher who's passionate about the use of A.I in the discovery and optimization of new molecules for chemical and biological applications.

flawnsontong1[at]gmail.com

+1 (647) 713 5354

10 Harfleur Road, Toronto

www.flawnsong.com

/in/flawnsong



## SKILLS

- **Web Development:** Proficient in Python, Javascript, Java, HTML, CSS. Familiar with JetBrains IDEs, Git, WSL, Vim, JSON. Experienced with Docker, Shellscript and various libraries, CMSs, DBs and APIs (jQuery, Bootstrap, PostGres).
- **Machine Learning and Data Science:** Proficient in Pytorch, Tensorflow, SKLearn, GCP, AWS. Familiar with various notebooks (Colab, Jupyter, Kaggle) and packages (numpy, matplotlib, pandas, RDKit, Openbabel, NLTK, OpenCV).
- **Entrepreneurial and research** Proficient at Photoshop, deck-building, presentations, pitches, networking, and research. Experienced with Microsoft office Suite, Google Suite, Tablaeu, and Zotero.

## EXPERIENCE

**Machine Learning Research Intern** | *Kebotix - Boston, MA (Remote)* Sept. 2019 - Present

- Building and benchmarking repurposed language models on string representations for molecular discovery.

**Machine Learning Developer Intern** | *Relation Therapeutics - London, UK* Sept 2019 - May 2020

- Built data preprocessing pipeline for BioGrid protein interaction graphs and Graph Convolutional model (PyG)
- Built validation pipeline (SciKitLearn), autodocumentation (Sphinx), and live training visualization (Tensorboard)
- Contributor to Sparse Dynamic Distribution Decomposition: Efficient Integration of Trajectory and Snapshot and

**Research Assistant** | *Globus Labs - University of Chicago* May 2019 - Sept 2019

- Filtered data and facilitated testing incoming datasets on benchmark machine learning models.
- Built data pipeline to extract molecular coordinates and energies from open source dataset.
- Researched and built POC for transfer learned NNP featurizer for graph convolution network.

**Front End Developer Intern** | *CIBC Digital Live Labs - Toronto, Canada* Jun. 2018 - Sept 2018

- Designed interface with Initial user experience (with InVision) testing at and 89% completion rate.
- Developed tutorial to Increase customer engagement of travel rewards points by 2x as projected in guerilla tests.
- Calculated to save CIBC approximately \$130 CAD worth of points per user on average.

## PROJECTS

### Generating novel molecules with an LSTM trained on SMILE string representations

- Developed in Python, built with Keras, used Google Collab and Google Cloud GPU Services to reduce training time to generate a subset of candidate molecules, 40% of which were valid.
- Formatted, normalized, and preprocessed unlabelled dataset of 250,000 unique molecules, used non-canonical SMILE strings to augment final dataset, divided into 3 and cross-validated with k-fold.

### Graph convolutional network with edge features for spatio-temporal urban traffic forecasting

- Researched Graph learning techniques for supervised directed edge regression methods and built POC.
- Preprocessed dense temporal (2 years) graph of 1200+ nodes and 1.2M edges, with 20+ features per node and edge