

Dylan Lewis

dylanrl97@gmail.com | +1 210-818-8787 | dyllew.github.io | LinkedIn: [/dyllew](#) | GitHub: [/dyllew](#)
San Antonio, TX

EDUCATION

Massachusetts Institute of Technology

Master of Engineering in Electrical Engineering and Computer Science; GPA: 5.0/5.0

Cambridge, MA

May 2022

Thesis: *Towards Automated Assessment of Crowdsourced Crisis Reporting for Enhanced Crisis Awareness and Response*

Concentration: AI

Bachelor of Science in Electrical Engineering and Computer Science; GPA: 4.3/5.0

May 2020

Relevant Coursework: Software Studio (TA); Software Construction; Data Structures & Algorithms; Computer Systems Engineering; Statistics, Computation, & Applications; Machine Learning; Advanced Natural Language Processing; Machine Learning & Data Science in Politics

SKILLS

- **Programming Languages:** Expert:{ JavaScript, Python }; Proficient:{ Java, R }
- **Technologies:** AWS, Docker, Git
- **Libraries & Frameworks:** NumPy, pandas, Matplotlib, PyTorch, scikit-learn, SciPy, TypeORM, NestJS, Vue.js, React, Redux

EXPERIENCE

MIT Urban Risk Lab

Cambridge, MA

Research Assistant

September 2020 - July 2022

- Designed system and constructed the backend for a web app which assists crisis managers during crisis events by leveraging machine learning models to classify crowdsourced crisis text and image data using TypeORM, PostgreSQL, RDS, NestJS, Docker, Terraform, and AWS namely S3, ECR, Lambda, API Gateway, and SageMaker
- Utilized insights from crisis managers in the US and Fukushima (FC), Japan to develop the Human Risk/No Human Risk classification task and determined that the F2 performance metric strongly aligned with their priorities; achieved a 92.8% F2 score on a test set of past flood event Japanese text data in FC with a SVM model and pretrained BERT embeddings
- Built pipeline and visualization tool for experimenting with various featurizations of Japanese crisis text data, dimensionality reduction techniques, and clustering algorithms, in order to yield human-interpreted labels from the documents found in each cluster; uncovered 9 human-interpreted labels from past flood event data in FC
- Led annotation effort forming a ground-truth test image dataset from past flood events in FC; achieved a 82.5% weighted F1 score on the test dataset using a trained CNN image Flood/Not Flood classification model
- Developed an open-source [Python package](#) for training, testing, and predicting with pretrained CNNs for classifying crowdsourced crisis image data; developed an open-source [Python package](#) for featurizing crisis text data, training and testing with a variety of classification machine learning models, and visualizing clusters of featurized text data
- Defined and supervised 8 undergraduate research projects adjacent to thesis research
- Presented main findings from research to key stakeholders from a wide range of technical backgrounds

Southwest Research Institute

San Antonio, TX

Software Engineering Intern

June 2020 - August 2020

- Developed a full-stack web application with React, Redux, TypeScript, Google Protocol Buffers, CouchDB, and Docker to enable a more intuitive and usable interface for performing experiments, iterating on feedback from peer review

Isobar

Boston, MA

Front-end Development Intern

May 2019 - August 2019

- Implemented business logic and user stories for a popular car rental company website UI using React

PROJECTS

- **[Supervised & Unsupervised Methods for Evidence Synthesis in International Development Gray Literature](#):** Python NLP project that used supervised & unsupervised methods to optimize the manual evidence synthesis process in International Development Gray Literature
 - Evaluated various text inputs and pretrained SpaCy Named Entity Recognition (NER) models against a simple baseline model to extract the Country of Study (CoS) associated with each international development paper in the corpus
 - Achieved 91.0% accuracy on the entire corpus for CoS extraction using the concatenated title and abstract as input to a pretrained Transformer NER model therefore enabling accurate filtering of unlabeled corpora for papers by predicted CoS
- **[Evolution of the U.S. TV News Narrative on Climate Change](#):** Data Science & NLP project in Python that investigated the evolution of climate change coverage frequency & content between U.S. TV News Networks CNN, Fox News, and MSNBC over Jul. 2009-Jan. 2020
 - Constructed TF-IDF embeddings for documents made from climate change news audio transcripts based on network, year, and network & year combinations to extract the most important words to each network, to each year, and to each network in each year
 - Computed cosine similarity between document embeddings to have a measure of content similarity between the documents to understand how climate change coverage content differed between the networks over time
- **[Boomerang](#):** Full-stack web application where users can efficiently and reliably borrow items from others within their communities
 - Drafted UI/UX wireframes and implemented full-stack functionality for login page and sign up flow in Express.js and Vue.js