

EDUCATION

Stanford University <i>MSc in Statistics</i>	Stanford, CA <i>June 2022</i>
Stanford University <i>BSc in Mathematical and Computational Science</i>	Stanford, CA <i>June 2021</i>

EXPERIENCE

Graduate Research Assistant <i>Stanford Human Trafficking Data Lab</i>	June 2021 – Present <i>Stanford, CA</i>
<ul style="list-style-type: none">• Developed a containerized, distributed, cloud-native data pipeline to process hundreds of thousands of square kilometers of satellite imagery daily (Docker, Kubernetes, Redis, RabbitMQ, Google Cloud Platform).• Developed a computer vision algorithm to identify charcoal production sites in satellite imagery (PyTorch, GDAL).• System will be deployed for use by federal prosecutors in Brazil to identify locations where forced labor is being used.• Papers forthcoming.	
Research Assistant <i>Stanford Center for Ocean Solutions</i>	June 2020 – June 2021 <i>Stanford, CA</i>
<ul style="list-style-type: none">• Created a deep-learning-based computer vision algorithm to identify small fishing vessels in satellite imagery (PyTorch, GDAL, OpenCV).• Analyzed entire near-shore region of the Peruvian EEZ and identified previously unknown locations where illegal, unreported, or unregulated fishing was occurring (Google Cloud Platform, Statsmodels, R).• Code available here.• Paper forthcoming.	
Tutor <i>Stanford University Mathematical Organization</i>	September 2018 – June 2019 <i>Stanford, CA</i>
<ul style="list-style-type: none">• Tutored Stanford students taking classes in linear algebra, vector calculus, and differential equations.	

PROJECTS

Light-Pipe <i>Python, Celery, Flask, GDAL, PyTorch, Tensorflow</i>	
<ul style="list-style-type: none">• A Python framework designed to facilitate the deployment of geospatial data pipelines at scale.• Provides platform-agnostic API for efficient ETL with cloud-optimized geotiffs.• Allows for model training with heterogeneous geospatial data sources and formats.• Handles task scheduling and queuing to facilitate distributed processing of geospatial data.• Facilitates data ingest from major data providers.	
“Weak Supervision with Incremental Source Accuracy Estimation”	
<ul style="list-style-type: none">• Developed a method to estimate the dependency structure and accuracy of weak supervision sources incrementally using precision matrices and robust principal components analysis.• Allows for model training with weakly-supervised training data in on-line settings.• Paper available here.	
“Effect of Behavioral Factors on Average Heart Rate During Cardiovascular Exercise”	
<ul style="list-style-type: none">• Designed a fractional factorial experiment to determine whether certain behavioral factors affect average heart rate during cardiovascular exercise.• Found that four of the two factors tested were associated with a significant effect on average heart rate during exercise.• Paper, data, and code available here.	

SKILLS

Languages: Python, C++, SQL, R, BASH
Tools: Git, Docker, Kubernetes, Apache Airflow, Apache Spark, Argo, PostgreSQL, PostGIS, Rabbit MQ, Redis, Google Cloud Platform, RESTful APIs
Libraries: Celery, Flask, Pytorch, Tensorflow, Scikit-Learn, GDAL, OpenCV, Statsmodels, Rasterio, Numpy, Pandas

AWARDS

National Merit Scholar	April 2017
-------------------------------	------------