

# Yu Cheng

✉ [yu@yu-cheng.co](mailto:yu@yu-cheng.co)  
🐙 [github.com/fischcheng](https://github.com/fischcheng)  
🌐 [yu-cheng-boston](https://www.linkedin.com/in/yu-cheng-boston)  
☎ (786) 503-3366

## PROFILE

Resourceful Data Scientist with a strong background in data wrangling, machine learning, and software engineering. Proven expertise in developing and deploying ML solutions, innovating through research and effectively collaborating with cross-functional teams.

## EDUCATION

### University of Miami

PhD in Meteorology and Physical Oceanography, 2018

### National Taiwan University

BSc in Atmospheric Sciences, 2010

## SOFT SKILLS

- Oral and written communication
- Creative problem-solving
- Research project management
- Team and culture building

## TECHNICAL SKILLS

- Statistics and advanced analytics, time series analysis
- Machine Learning: Scikit-Learn, PyTorch, Tensorflow
- Languages: Python, SQL, Matlab
- Workflow: Airflow, Docker, Kubernetes, dagster
- GCP: BigQuery, Looker Studio, Pub/Sub; AWS: EC2, SageMaker
- Others: QGIS, GDAL, rasterio, DBT, datadog, matplotlib, plotly/dash, tableau

## CERTIFICATES

- [Deep Learning Specialization](#)
- [AWS Certified Machine Learning – Specialty](#)

04.19.2024

## PROJECTS

### Dashboard of Dark Ship Detection and Monitoring

developed a CNN-based detection system for illegal fishing vessels using [Synthetic Aperature Radar images \(SAR\)](#), showcasing skills in designing end-to-end data workflows and extracting insights from multimodal data.

### Flood Detection Data Challenge

participated in the DrivenData Flood Water Detection Challenge, where I trained a segmentation model using PyTorch. This model, based on SAR images, achieved a validation IoU score of 0.77.

## EXPERIENCE

### Massachusetts Institute of Technology

Sustainability Data Scientist

2023 - Present

Boston, MA

- Devised and developed a comprehensive data platform, encompassing pipeline orchestration, data warehousing, and cataloging for campus-wide sustainability data. This initiative ensured data quality and transparency to meet MIT's Scope 3 reporting needs.

### Web3 Builders

Founding Senior Data Scientist

2022 - 2023

Boston, MA

- Developed labeled datasets by integrating 10+ data sources, enabling ML models to predict scam probabilities of Ethereum wallets and smart contracts, powering the transaction protection browser extension, TrustCheck.

### SFL Scientific (acquired by Deloitte)

Senior Data Scientist

2021 - 2022

Quincy, MA

- Implemented ML solutions tailored to clients' objectives, leading to improved business performance by conducting rigorous exploratory analysis, defining metrics, and supporting the full data science life-cycle.

### Tomorrow.io

Senior Atmospheric Data Scientist

2018 - 2021

Boston, MA

- Created novel weather products by applying computer vision and ML techniques, resulting in a near real-time global precipitation pipeline using geostationary satellite images and cloud infrastructures with [Google Cloud Platform](#), [xarray](#), [dask](#) and [satpy](#). [\[blog post\]](#). The pipeline shows a significantly smaller latency than that of a similar public product, and enabled the company to expand internationally.
- Developed a pytorch dataloader and trained a U-NET based model combining Geostationary satellites images and Lincoln Lab [SEVIR Dataset](#) to detect precipitation intensity and coverage.