

Yoni Nachmany
New York City • yoninachmany@gmail.com • 646-717-3333

Machine learning and software engineer proficient in Python and deep learning, driven by frontier AI.
Thrives in small, ambitious, and collaborative teams **translating research innovations into real-world impact.**

Experience

Zeus AI

Remote

Head of Data Engineering, First Hire

May 2024 – May 2025

Zeus AI spun out of NASA to design **novel AI weather models with Earth observations**. I developed software, infrastructure, and workflows for **generating pretraining data**, evaluating new modalities, and serving models.

- Productionized **PyTorch foundation model** for global weather forecasting **50x faster than NWP**, applying the **Massively Multimodal Masked Modeling transformer architecture** to remote sensing with **robust, real-time, serverless ingestion and inference** using Modal GPUs and AWS SNS/Lambda.
- Crawled and **tokenized TB-scale spatiotemporal data on supercomputers** using Slurm and MPI.
- Deployed and efficiently served models of hourly solar potential for **energy traders** using FastAPI.

The Linux Foundation

Remote (Contract)

Senior Software Engineer, Data Pipeline Task Force

March 2024 – May 2024

Overture Maps Foundation is a collaboration with **member companies like Amazon, Meta, and Microsoft** for open data. I helped the **GA launch of a global geospatial data standard**, the Global Entity Reference System.

- Prototyped **Parquet Data Changelog with Spark and Athena**, benchmarking partitioning schemes.

Gro Intelligence

New York City

Software Engineer, Data Platform Project Lead

April 2023 – March 2024

Gro Intelligence built an **agricultural data platform with predictive models** mapping the impact of climate change on commodities. I led **development of large-scale ETL workflows** for training machine learning models.

- Orchestrated **Airflow DAGs for a century of climate projections** using ECS and an S3 tensor data lake.
- Contributed to and oversaw development of internal **Python library** to standardize Xarray and Zarr.

The New York Times

New York City (Contract)

R&D Engineer

December 2021 – September 2022

The NYTimes R&D Team develops **new technical capabilities and story formats** with emerging technologies. I contributed to 3D mapping research and **led the CV/NLP work to accurately digitize the news archive**.

- Scaled **multimodal transcription and evaluation of 10M articles** from labeled scanned newspapers.
- Implemented **3D photogrammetry generation** with stereo satellite libraries to support newsroom coverage.

Mapbox

Washington, D.C.

Geospatial Software Engineer

September 2019 – November 2021

Mapbox's Imagery team **improves the global satellite basemap** by sourcing, updating, and enhancing datasets. I extended the capabilities of the **PB-scale image data processing pipeline** and derived value-added products.

- Processed 135M km² of satellite data to **launch 3D map product and saved \$1M** with open aerial data.
- Improved the UX for all 3D satellite map users with a **novel multi-resolution data fusion algorithm**.
- Implemented and evaluated **deep learning-based super-resolution and detection** for enterprise users.

Education

University of Pennsylvania

Philadelphia

Master of Science in Engineering (MSE), Data Science

May 2019

Bachelor of Science in Engineering (BSE), Networked & Social Systems Engineering

May 2018

Hunter College High School

New York City

Teaching

TA for NETS 213: Crowdsourcing & Human Computation with Prof. Chris Callison-Burch	Spring 2019
TA for ESE 305: Foundations of Data Science with Prof. Victor Preciado	Fall 2017

Publications

Global Atmospheric Data Assimilation with Multi-Modal Masked Autoencoders	ICDM 2025
Learning Word Ratings for Empathy and Distress from Document-Level User Responses	LREC 2020
Design Considerations for High Impact, Automated Echocardiogram Analysis	ICML 2020
Mitigating Geographic Bias of Image Classifiers with Multilingual Image Data	D4GX 2019
Detecting Roads from Satellite Imagery in the Developing World	CVPR 2019
Generating a Training Dataset for Land Cover Classification to Advance Global Development	NeurIPS 2018
Decoded: Crowdsourcing Summaries of Internet Privacy Policies	HCOMP 2017

Fellowships/Internships

Data Science for Social Good Fellow at Imperial College and The Alan Turing Institute	Summer 2019
Machine Learning Intern at Radiant Earth Foundation	Summer 2018
Inaugural Earth on AWS Fellow at Sinergise (acquired by Planet)	Spring 2018
Open Source Software Fellow at Azavea	Summer 2017
Software Engineer Intern at Meetup	Summer 2016
Machine Learning Research Intern and hackNY Fellow at Clarifai	Summer 2015

Talks

Generating Elevation Data from Stereo Pairs of Satellite Imagery	GeoPython 2023
Depth Perceptions - Perspectives on 3D & Digital Twins	SatSummit 2022
Mapping Historical “Street View” Images of New York City: Visualizing Geotagged Archival Photos	FOSS4G 2022
Under-Mapped Spaces: New Methods and Tools for Critical Storytelling	Stanford 2022
Deep Transfer Learning for Land Cover Classification on Open Multispectral Satellite Imagery	FOSS4G 2019