Maayan Matsliah

Boston, MA / New York, NY | (408) 480-7792 | matsliah.m@northeastern.edu | linkedin.com/in/maayan-matsliah | github.com/maayanmatsliah-tech | maayanmatsliah-tech.github.io

Education

Northeastern University Boston, MA

B.S. in Computer Science, AI Concentration | GPA: 3.88/4.0 | Dean's List *Expected May 2029*

- **Relevant Coursework:** Mathematical Foundations of AI, Program Design & Implementation II, Discrete Structures, Linear Algebra, Statistics
- **Certificates:** Machine Learning Foundations (Cornell), AWS Cloud Practitioner Essentials (AWS)

Technical Skills

Languages: Python, Java, JavaScript, HTML, CSS, SQL, Shell

Frameworks/Tools: TensorFlow, Keras, PyTorch, torchvision, scikit-learn, Flask, NumPy, Pandas, Matplotlib, Seaborn, DuckDB, REST APIs

ML/AI Concepts: RAG, Agentic AI Workflows, Convolutional Neural Networks, Supervised & Unsupervised Learning, LangChain, Data Encoding, Trees, KNN, Feature Engineering, Model Evaluation, Computer Vision, ResNet, Word2Vec, Hyperparameter Tuning

Experience

Cornell Tech

Break Through Tech AI/ML Fellow *Mar 2026 – Present*

- Trained and benchmarked **Decision Tree and neural network classifiers** on Airbnb price prediction (**81% vs. 82% accuracy**) and chose the Decision Tree for production, cutting training time **˜100x** with no meaningful accuracy loss.
- Built a **RAG-powered chatbot** using **LangChain, Word2Vec, and OpenAI's API** for a camping company, grounding responses in their internal files so the model can answer business-specific questions beyond its base knowledge.
- Matched to a **6-person** AI Studio team with MathWorks, building a fault-diagnosis model for smart appliances, applying **pruning and quantization** to compress it for embedded deployment.

GitHub

Open Source Contributor *Jan 2026 – Present*

- Authored pull requests that were **merged across a dozen production codebases** with **200–700+ GitHub stars**, including conda/constructor, python-humanize/humanize, and the Munich Quantum Toolkit.
- Refactored a scientific-notation utility in python-humanize, simplifying the logic and improving benchmarked performance by **5%** (**75.3µs → 71.5µs**).
- Identified and fixed a silent error-handling bug in a production data platform by adding proper **API response unwrapping**, ensuring failed requests surface accurate errors instead of false success messages — reviewed and approved by **5 maintainers** before merge.

Princeton Neuroscience Institute (FlyWire)

Citizen Scientist & Internship Coordinator *May 2023 – Present*

- Designed and ran a weekly progress-report system across **15 interns**, analyzing self-reported status to flag at-risk progress and surface common blockers; co-led resulting 1:1 check-ins with a Princeton researcher, contributing to **12 of 15 interns** being cleared to present to **30 scientists** at the Princeton Neuroscience Institute symposium.
- Designed weekly sync-up meeting structure and organized talking points based on common questions and roadblocks.
- Annotated neuron type and location for the **BANC connectomics dataset** using the **Codex AI platform**.

Technical Projects

Citation Graph Analysis | *Python, SQL, DuckDB, OpenAlex API, Claude Code* | GitHub
Independent research using a citation graph to test how field diversity and the rise of remote work reshaped scientific collaboration.

- Architected a high-throughput graph processing pipeline in Python and DuckDB to ingest **413M+ JSON records** and process a **2.9B-edge citation graph** under strict memory constraints (**12GB RAM**).
- Designed partitioned batching algorithms to detect reciprocal edges across **˜2.9B unnested citations**, reducing memory footprint by **>90%** while achieving **100% referential integrity validation**.
- Engineered automated data-cleaning and taxonomy backfill pipelines resolving left-boundary metadata censorship across **34M historical works** and enabling rapid decile aggregation across **115M+ papers**.

Marine Debris Classifier | *Python, PyTorch, torchvision, ResNet18* | GitHub
A computer vision classifier built to support automated ocean cleanup by distinguishing marine life from debris.

- Engineered an end-to-end data pipeline in PyTorch across **3 public and custom datasets (˜26K underwater images)**, resolving an initial **10:1 class imbalance down to 3:1** and applying dynamic augmentations (random resized cropping, horizontal flips, ImageNet normalization) to prevent overfitting.
- Fine-tuned a pretrained **ResNet-18** classifier using SGD with momentum (0.9) and StepLR decay, overcoming an initial **46% baseline** to achieve **99.81% validation accuracy** across **10 epochs** on an **NVIDIA T4 GPU**.
- Evaluated class-level performance via **Scikit-Learn confusion matrices** and classification reports, prioritizing high marine-life recall to minimize false debris classifications and prevent wildlife capture in autonomous ocean cleanup systems.

Savings Made Simple | *Python, Flask, HTML, Render, SQLAlchemy* | GitHub | Live Demo
A personal budgeting tool for tracking spending against a savings goal while studying abroad and guide real-time budget adjustments.

- Built and deployed a **Flask/SQLAlchemy** web app on **Render** with relational data models, session-managed CRUD, and dynamic weekly burn-rate calculations.
- Engineered in-memory **Matplotlib** visualizations (**Base64/BytesIO**, no disk I/O) and **pytest** suites covering financial projection edge cases.