

Aviv Gelfand

📍 Jerusalem, Israel ✉ avivgelfand@gmail.com ☎ 054-300-7257 in aviv-gelfand 🌐 AvivGelfand

Summary

- 1st year MSc candidate in Data Science (HUJI); Thesis: Dimensionality reduction of Deep Learning problem space through statistical methods and graph theory. BSc in Statistics and Data Science (HUJI; GPA 89)
- Dedicated to long-term growth in a cutting-edge company with a positive atmosphere; available for 80%-100% time role as a Data Scientist / ML or AI Researcher / Engineer / SW Dev.
- Demonstrated success in AI and software development through academic/private projects and roles, with a track record of implementing machine learning models to solve complex problems requiring strong collaboration and rapid self-learning skills, consistently delivering efficient and accurate models.

Experience

NLP Data Scientist / Research Assistant

Oct 2023 – May 2024

The Hebrew University of Jerusalem

- Successfully executed a Supervised-Fine-Tuning training of large language models to classify defendant arguments in trials for a broader statistical legal research project; reduced labeling time and errors, replacing manual tagging with LLM-based tagging.
- Drove strategic alignment by delivering situational analyses and action plans to US-based stakeholders, effectively prioritizing objectives and focusing efforts to meet deadlines.
- Dealt with significant imbalance using embedded cluster stratified sampling methods. Corrected errors in the labeled data. Reached over 90% accuracy and 80% F1 scores in most target labels. Substantially reduced model complexity and training time using efficient sampling.
- Acquired all relevant knowledge in PyTorch and LLMs through fully independent learning, unguided.
- Presented findings at a prominent statistics and data science conference in May 2024.

Statistics & Machine Learning Instructor

Oct 2022 – Sept 2023

Upscale Analytics

- Built dozens of ML hands-on projects for a Machine Learning and Statistical Inference course, with constant feedback from a Senior Data Scientist (and head lecturer), Michael Winer.
- Instructed data analysts through their first steps in statistical inference and classic machine learning: linear and logistic regressions, ensemble learners such as random forests & XGBoost, and clustering methods.
- Advanced from teaching assistant to instructor, then to a partner role. 3 successful semesters of weekly TA sessions with a significant presence of students.

Research & Development Assistant

Apr 2022 – Mar 2023

The Hebrew University of Jerusalem

- Self-taught and programmed smart contracts with a C++-based programming language; developed simulations interface using react.js.
- Designed and analyzed online surveys; conducted literature and academic peer reviews. Web-scraped data monthly.

Data Analyst & Research Team Leader

Dec 2018 – Dec 2020

Israel Defense Forces

- Conducted research, assessments, and analytics for high-ranked decision-makers. Triplicated, updated, and maintained strategic databases while collaborating with programmers and economists.
- Led cross-organizational projects resulting in several two in-depth research papers with an unprecedented scale of collaboration between two research branches.

Education

The Hebrew University of Jerusalem

Oct 2023 – July 2026

MSc in Statistics

- Thesis on Dimensionality Reduction in Deep Learning (LLMs focused), guided by Prof. Ariel Jaffe.

The Hebrew University of Jerusalem

Oct 2021 – Aug 2024

BSc in Statistics & Data Science

- GPA: 89 ; Also completed 33% of the required courses for master's degree while working in several roles.

Technologies

Scripting Languages and tools: Python, Java, Solidity (C++ based), R, SQL, JS

Programming Libraries: Pandas, PyTorch, Transformers, SKlearn, LLM's SDK (OpenAI, groq, etc), Selenium, Beautiful soup

GenAI: Prompt engineering (COT, etc), Crusur, Claude, ChatGPT, Perplexity, 'Langtalks' fan

EDA: Basic and complex Visualizations, Advanced Statistical Inference

ETL: Cleaning, Encoding / Embedding, Imputing

ML: Efficient Sampling, Feature Engineering, Parameters Optimization, Evaluation Metrics

NLP: Fine-Tuning open source LLMs, LORA, PEFT, text cleaning techniques with Regex and NLTK

CI/CD: GIT, Linux, SLURM Cluster system, GitHub workflows, shell scripts

Big Data Frameworks: Big Query, Data-Bricks, Apache Spark