Aviv Gelfand

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

Summary

- 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.
- 1st year MSc candidate in Data Science (HUJI); Thesis: Dimensionality reduction of Deep Learning problem space through statistical methods. BSc in Statistics and Data Science (HUJI; GPA 89)
- 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

- Fine-tuned (SFT) Large Language models for a US-Israeli statistical legal research team, classified arguments in trial records.
- Applied LORA, PEFT, text cleaning, and statistical methods.
- Dealt with the significant imbalance and noisy labels and features, using embedded cluster stratified sampling methods and an iterative active learning approach.
- Corrected errors in the labeled data.
- Reached over .90 accuracy and over .80 F1 scores in most target labels, over 20% improvement from the initial benchmark.
- Utilized the university's Cluster resources (with Linux) to fine-tune open-source foundational models with PyTorch Lightning.
- Acquired all relevant knowledge in PyTorch and LLMs with nothing but a laptop and the Internet.
- Presented my work at the annual Israel Statistics and Data Science Association conference in May 2024.

Statistics & Machine Learning Teaching Assistant   *Oct 2022 – Sept 2023*
Upscale Analytics

- Built and taught dozens of machine-learning projects in the company's 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 (random forests & XGBoost), clustering with K-means, and DBSCAN.
- 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 economic research, assessments, and analytics for high-ranked decision-makers.
- Led a joint project with a programming team for an economics dashboard app for a Major General (Aluf).
- Created, updated, and maintained databases of macro-economical data via Microsoft Excel.

Education

The Hebrew University of Jerusalem *Oct 2023 – July 2026*
MSc in Statistics
- Thesis on Dimensionality Reduction of 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 ; Finished 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