# Aviv Gelfand

⚲ Israel   ✉ avivgelfand@gmail.com   in aviv-gelfand   ⌨ AvivGelfand

## Summary

3rd year BSc (GPA 89) / 1st year MSc in Statistics and Data Science with thesis at HUJI.

Seeking an 80%-100% position as a Data Scientist / ML/AI Engineer / Researcher / SW Dev.

Solid experience in SW / AI student positions and projects, with independent learning and teamwork.

## Experience

**The Hebrew University of Jerusalem**, NLP Data Scientist / Research Assistant
Israel
Oct 2023 – present
11 months

- 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.

**Upscale Analytics**, Statistics & Machine Learning Teaching Assistant
Israel
Oct 2022 – present
1 year 11 months

- Built and taught dozens of machine-learning projects in the companys 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.
- Promoted to be a lecturer in the course after 3 successful semesters of weekly TA sessions with a significant presence of students.
- Advanced from teaching assistant to instructor, then to a partner role.

**The Hebrew University of Jerusalem**, Research & Development Assistant
Israel
Apr 2022 – Mar 2023
11 months

- 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 reviews and academic peer reviews.
- Web-scraped data monthly.

**Israel Defense Forces**, Data Analyst & Research Team Leader
Tel Aviv, Israel
Dec 2018 – Dec 2020
2 years 1 month

- 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

**Bachelor of Science - BS**    **The Hebrew University of Jerusalem**, Statistics & Data Science      Dec 2021 – Dec 2024
- GPA: 89
- Coursework: Human Language from a Computational Perspective (NLP), Linear Algebra for Statisticians, Statistical Inference and Its Applications, Modern Statistical Data Analysis, Regression and Statistical Models, Intro to Machine Learning, Big Data Mining, Data Structures and Algorithms

**Master of Science - MS**    **The Hebrew University of Jerusalem**, Statistics      Sept 2023 – July 2026
- Coursework: Signal Processing, Geometry of Data

## Projects

**Fine Tuning Large Language Models**      2023-09 to 2024-04
- This repository presents my (Aviv Gelfand) work on applying supervised fine-tuning (SFT) methods to build classifiers for labels as part of a broader project in statistical legal research collaboration led by Dr. Adi Leibovitch and Sharon Levy from The Hebrew University of Jerusalem Faculty of Law, Prof. J.J.Prescott and Grady Bridges from the University of Michigan Law School.

**De Grandmizer Shtick! GenAI Hackathon**      May 2023
- De Grandmizer Shtick is a Python application that generates conversations using GPT-3 and converts them into audio files using the Google Cloud Text-to-Speech API.
- The project is designed to help create the illusion of an occupied home to intimidate burglaries and was created during HUJI Generative AI Hackathon 2023.

**Alfred - a Personal Moodle Bot**      Apr 2023
- Alfred is a Python-based bot Scraping & Messaging bot that scrapes The Hebrew University of Jerusalem's Moodle and sends a message through Telegram about any new update or upload in upcoming deadlines for assignments.

**Driver Behavior Detection Using Smartphone Signals**      2024-05 to 2024-06
- A Digital Signal Processing journey to an application for detecting and reporting aggressive bus driver behaviors.

## Technologies

**Programming Languages & Tools:** Python, Pandas, NumPy, Matplotlib, Plotly, Scikit-Learn, TensorFlow, Keras, PyTorch, Git

**Machine Learning & Data Science:** Machine Learning, Deep Learning, Natural Language Processing (NLP), Large Language Models (LLM), Transformers, Statistical Inference, Predictive Modeling, Data Analytics, Data Cleaning, Data Wrangling, EDA, Big Data Analytics

**Other Skills:** Problem Solving, Analytical Skills, Communication, Teamwork, Self Learning, Independent Thinking