

GABRIEL ALON

(650) 799-9295 | gabrielalon257@gmail.com | California

www.linkedin.com/in/gabrielalon | github.com/galonpy | Portfolio: galonpy.github.io

EXPERIENCE

BluelightAI

Senior Data Scientist

February 2024 - Current

- Developed a software integration and corporate partnership from scratch with the CEO of Marqo, an ecommerce recommendations company. Built a python toolkit that improves neural network fine-tuning and model selection for their customers by summarizing error patterns. Each error translates to users getting the wrong product search results.
- Created our data summarization algorithm. It intuitively summarizes the contents of vast unstructured customer data with keywords that are unique to distinct groups in the data. Combined topic modeling algorithms with topological data analysis.
- Published demos that make topological data analysis easy and useful for curating what training data should go into an LLM.

University of Michigan

Machine Learning Researcher

May 2022 – February 2024

- Created a personalized YouTube video browser that won a scholarship prize. Fine-tuned a BERT model in Pytorch to filter out toxic content and created a production OpenAI pipeline. Deployed into a live [website](#) for watching videos, using Streamlit.
- [Authored](#) a 90+% effective solution to a famous unsolved cybersecurity issue for LLMs described in the NYTimes. Combined GPT-2's perplexity property with a gradient boosting model to detect when a user is trying to circumvent safety restrictions.

McD Tech Labs (Purchased by IBM) | Mountain View, California

Data Scientist

June 2019 - July 2021

- Directly improved \$14,000,000 worth of customer interactions (2% of sales a day across 100 stores) by fixing Python bugs in a logistic regression model that was causing the Siri-like voice AI to ignore customer speech. Fixed the evaluation metrics.
- Created performance tracking visualizations with Python and SQL for our engineering and business teams that summarized millions of food ordering customer interactions a day. Used pandas, scikit-learn, matplotlib, tableau and superset.
- Developed time series models in Python and SQL with AWS RedShift to identify which versions of the software were performing better over time. Used the causal inference technique difference-in-differences for the time series analysis.
- Wrote production level code with unit testing, CI/CD, static analysis, Jira tickets and code reviews.
- Presented a weekly performance report used for prioritizing engineering resources. This report highlighted product trends, root cause analyses of specific software failures, and proposed product requirements.

Wefi | Tel Aviv, Israel and Santa Cruz, California

Data Analyst (Part-Time)

July 2015 - January 2016

- Wrote SQL, Tableau, and Excel reports on broad mobile app usage patterns from a dataset of 1 million Android phones.
- Tableau visual published on Venturebeat.com in an article titled “Attention Shoppers: Shopping with a store app...”
- Produced reports by custom request, i.e: informing investors prior to Square's IPO and relating ads to app behavior.

SKILLS

Languages: Python (6 years), SQL (5 years), Object-oriented programming, Unix, Bash, Shell

Tools: Jupyter Notebook, VScode, GitHub, AWS, Tableau, Superset, Excel, Jira

EDUCATION

University of Michigan, Ann Arbor

Master of Applied Data Science

3.7 GPA

May 2022 - August 2023

University of California, San Diego

Bachelor of Science in Management Science

Graduated 2019