

AARON GUROVICH

aargurov@ttu.edu • +1 949-505-0956 • aarongurovich.com

Education

Texas Tech University

B.S. in Computer Science, Minor in Mathematics

Lubbock, TX

Expected: May 2026

- **GPA:** 3.5
- **Relevant Coursework:** Software Engineering, Applied Data Science, Concepts of Programming Languages, Design and Analysis of Algorithms, Data Structures, Object-Oriented Programming, Computer Architecture.
- **Activities:** President of Hillel, Member of Chess Team (USCF: 1907).

Experience

Johnson & Johnson Electrophysiology

Clinical Data Engineer Intern (Extended)

Irvine, CA

June 2025 – November 2025

- **Architected** an **R Shiny** site selection tool for clinical trials using a custom **genetic algorithm**, automating optimal site selection to **accelerate trials**, **ensure diverse participant representation**, and **support FDA compliance**.
- Boosted a **Generative AI** pipeline's accuracy in classifying **protocol deviations** from **75% to 98%** by optimizing data ingestion in **Python**, applying advanced **prompt engineering**, and refining the **AWS** infrastructure.
- **Conceptualized** and developed an **AI-powered assistant** in **R** that converts **natural language** into optimized **SQL queries**, reducing costly AI ingestion and enabling **faster, more reliable insights** in **R Shiny dashboards**.
- **Transformed** clinical device data by refactoring legacy code into an **OOP Python** framework across multiple **Lambda functions**. Built **unit tests**, used **Bitbucket** with **Jira** for version control, and **Jenkins** for **CI/CD** tracking.
- **Collaborated** with clinicians, statisticians, and engineers to translate business needs into scalable data solutions, and **co-authored 3 abstracts** showcasing my projects for the **Johnson & Johnson Data Science Showcase**.

Texas Tech University

Research Assistant

Lubbock, TX

Oct 2024 – May 2025

- Built a **full-stack platform (MalScope)** with **Python**, **Flask**, **D3.js**, and **Cytoscape.js** that helps analysts **detect anomalies in network traffic** through real-time, interactive **visualizations**.
- Developed a high-performance **Python data pipeline** to process raw network logs and PCAP files, generating **20+ engineered features** and scaling reliably to **billions of records**.
- Applied **machine learning (Louvain clustering in NetworkX)** to automatically **group suspicious traffic patterns**, making hidden anomalies easier to identify.
- Created a user-friendly **interface** with **heatmaps** and **network graphs**, adding features like cluster highlighting and tooltips to make complex data **easy to explore**.
- Delivered an **end-to-end solution**, from **data ingestion** to **feature engineering**, **ML clustering**, and **visual analytics**, that speeds up investigations and improves decision-making.

ProofPerks

Software Engineer Intern

Remote

Aug 2024 – Dec 2024

- Fine-tuned **transformer models** from **Hugging Face** in **Python** for race/ethnicity verification, boosting authentication accuracy and deploying them as **scalable microservices** on **Google Cloud Platform (GCP)**.
- Built a cloud-native **data pipeline** on **GCP** to automate ingestion, preprocessing, and augmentation of large-scale biometric datasets, supporting both **model training** and **real-time inference**.
- Optimized system performance by implementing a **multi-layer caching strategy** with **Redis**, cutting API latency and reducing database load for live biometric verification requests.

Projects

UStartKit <https://ustartkit.com/>

- **Built a full-stack starter kit generator** with **React + TypeScript** and **Supabase Edge Functions (Deno)**, integrating **OpenAI GPT-3.5** and the **Amazon Product API** to generate curated essential, premium, and luxury kits.
- **Engineered a product pipeline** that sanitized titles, parsed prices, filtered low-quality items, enforced rating thresholds, de-duplicated entries, and ranked results to deliver consistent, high-quality recommendations.
- **Implemented persona-aware retrieval** where age, gender, and skill level shaped queries; merged **LLM-curated brand suggestions** with multi-page searches to boost recall, diversity, and personalization.
- **Optimized scalability and UX** with sequential API calls to handle rate limits, tiered product selection for clear choices, and concise AI-generated reasons for inclusion to keep the experience engaging and reliable.

Skills

Programming: Python, R, Java, SQL, Pandas, Numpy, Scikit-learn, PySpark, Scapy, dplyr, ggplot2, tidyR
Cloud & Data: AWS (S3, Lambda, SQS, Redshift, Glue), GCP, Azure, Docker, PostgreSQL, Cloudflare, Supabase
Frameworks: Flask, React, Shiny, D3.js, REST APIs, Cytoscape.js, Node.js, Next.js
Dev Tools: Git, Linux, Bash, CI/CD, GenAI, RStudio, VS Code, Tableau