Josh Rauvola

Dallas, TX 75207 • jrauvola@gmail.com • jrauvola.com • LinkedIn

Professional Summary

AI Engineer with 4+ years of experience specializing in LLMs, multimodal AI, and advanced machine learning solutions. Hold Master's degree in Applied Data Science from University of Chicago (GPA: 3.98) and dual B.S. degrees in Computer Science and Economics. Demonstrated expertise in delivering $10M+ business value through innovative AI implementations across diverse industries including logistics, healthcare, and financial services, with focus on scalable solutions and regulatory compliance.

Education

University of Chicago, Chicago, IL *Sept 2023 – Dec 2024*

- Master's in Applied Data Science with a focus in AI/Machine Learning
- GPA/Awards: 3.98/4.0, Data Science Institute Scholarship Recipient

California Polytechnic State University, San Luis Obispo, CA *Sept 2018 - June 2022*

- Dual Bachelor of Science: Computer Science (Focus in AI/ML) & Economics (Focus in Quantitative Analysis)
- Graduated in four years while working full time and receiving two degrees, National Green & Gold Scholar

Professional Experience

U.S. Bank

AI Engineer *Feb 2025 – Present*

- Deployed Torque Clustering, an unsupervised ML algorithm inspired by gravitational physics, achieving 15-20% higher accuracy than traditional methods (K-means/DBSCAN) for customer segmentation and anomaly detection
- Architected Data Dev, an executive data analysis tool with federated database connections using Spice AI, achieving 90%+ accuracy on semantic parsing benchmarks and enabling real-time business intelligence
- Implemented foundational time series models for ATM cash-forecasting initiative, accounting for seasonal patterns and reducing manual interventions by 30%, optimizing cash distribution across thousands of terminals.
- Built explainable product recommendation engine using transparent clustering algorithms to ensure regulatory compliance while driving cross-sell opportunities

Harbor Global

AI Engineer *May 2024 – Feb 2025*

International logistics and supply chain management company

- Architected comprehensive AI solution utilizing LLMs to generate assessments, query data, and create reports using multi-shot prompting and context-aware responses, projected to save $10.8M annually
- Implemented advanced data ingestion pipelines for real-time data integration, reducing manual research hours by 70%
- Developed robust database infrastructure with vectorized databases for complex queries across 15M+ records

Amaze Health, Denver, CO

Data Engineer *Jan 2023 – Sept 2023*

Comprehensive virtual healthcare service provider

- Supported the development of a new patient learning center by setting up data pipelines and implementing data migration practices that enabled efficient, real-time transfer of terabytes of data
- Engineered a systematic enrollment and billing system to streamline customer onboarding, reducing lost billing income and increasing overall process efficiency by 86%

Critical Insight, Seattle, WA

Software Developer *Jun 2021 – Oct 2022*

Managed Detection and Response cyber-security company

- Built and deployed new pages and features for an internal customer health portal to improve analytical functionality and information display
- Designed predictive models for the portal using TensorFlow and scikit-learn to enhance respondent experience
- Architected and developed a Cassandra-based system to clean and deploy 10+ years of AWS S3 bucket data

Prime Trust, Las Vegas, NV

Data Scientist *Oct 2021 – Jun 2022*

Financial infrastructure fintech and digital asset company

- Implemented NLP algorithms within Airflow data pipelines, improving reporting efficiency by 60%
- Collaborated with finance and data science teams to design and implement quantitative time series forecasting in customer pricing models using TensorFlow and Spark
- Advised technical and business leaders on best practices for effective and accurate data flow across systems

Research Experience

Northwestern University
Graduate Research Lead *Jun 2025 – Present*

- Directed a multidisciplinary team of graduate researchers to develop and validate a comprehensive language-agnostic framework for measuring developer productivity in the open-source landscape
- Completed comprehensive research study with manuscript under review at ISE 2025

University of Chicago
Graduate Research and Teaching Assistant *Mar 2025 – Present*

- Supported advanced research initiatives in AI/ML while mentoring graduate students on thesis/capstone projects
- Contributed to ongoing research projects focused on AI and energy-efficient computing solutions

Digital Emissions
Research Fellow *May 2024 – Jan 2025*
Non-profit dedicated to reducing digital footprints and promoting sustainable technology

- Developed collaborative platform for Green Software Foundation, garnering 300+ contributions within 90 days
- Published benchmarks showing 35% reduction in energy consumption for NLP tasks using carbon-aware scheduling

Technical Skills

Languages & Frameworks: Python, R, C, TensorFlow, PyTorch, Keras, Spark, SQL, Cassandra
AI/ML Expertise: LLMs, Computer Vision, NLP, Time Series Analysis, Clustering, Neural Networks, RAG, LoRA
Tools & Platforms: Google Cloud Platform, AWS, Docker, Kubernetes, Git, Jupyter, Apache Airflow

Awards, Honors & Media

Winner — UChicago AI+Science Hackathon (Molecular GNN Challenge) *May 2024*
1st overall among UChicago & external PhD teams; built a message-passing GNN on ChEMBL-scale data to model molecular interactions (Article).

Podcast Guest — Backstage: Awesome Green Software *Nov 2024*
Deep-dive on the new platform directory's system architecture: Notion-backed CMS with Algolia search and contributor workflow design; highlighted available carbon-aware/measurement tooling (Episode).

Projects

LIA: Large Interview Assistant
GitHub Repository *Dec 2025*

- Engineered a multi-agent interview coach (RAG, CoT, baseline) that delivers real-time, rubric-scored feedback and personalized Q&A across behavioral and technical domains
- Built a custom AR glasses prototype that runs LIA—head-mounted mic/camera streaming to the backend with in-frame prompts and confidence gauges for hands-free practice
- Designed and productionized voice & facial analytics (fluency, pause rate, pitch variance, gaze/engagement) plus a privacy-preserving resume parser for user context
- Shipped full-stack platform (React front end, Flask services, LangChain orchestration) on GCP Cloud Run; integrated Gemini 1.5 Pro for question generation and expert exemplars
- Authored a STAR-aligned scoring rubric and evaluation agent; ran ablations comparing RAG vs. CoT to improve answer structure and relevance

Publications

- ***Fair Developer Score: Build-Adjusted Measurement of Effort and Impact***. published in ASE 2025
  Wang X., Zhu J., Feng J., Zhang Z., Ali A., **Rauvola J.**, Delgado D., Antar A.
  — Proposes build-normalized metrics to compare developer output across heterogeneous codebase utilizing commit-centric clustering to form build-level units, define effort/importance metrics and validate against process outcomes
  — Developed validation methods/experiments and implemented torque clustering section as Graduate Research Lead.
- ***ThoughtTrim: Anchor-Driven RL Modification***. Apart Research, 2025.
  Briand A., **Rauvola J.**
  — Reinforcement learning approach that steers models toward high-influence "anchor" steps for faster, cheaper reasoning; co-designed the algorithm and evaluations.

Personal Interests

• University of Chicago XC & TF Athlete • 4:17 mile time • Green AI Research • Open Source Contributor