Jonathan A. Rocha

Data Scientist · AI/ML Engineer · Full-Stack Web Developer

(737) 397-0800 · jarocha@smu.edu · jonathanaaronrocha.com · LinkedIn · GitHub

Professional Summary

Data Scientist and AI/ML Engineer with over 20 years of full-stack web development experience across financial services (Wells Fargo), automotive retail, marketing technology, and AI/ML consulting. Strong interdisciplinary foundation spanning machine learning, NLP, topological data analysis, and cloud infrastructure, paired with advanced training in English and historical research. Currently completing an M.S. in Data Science at SMU with a research focus on ensemble transformer models, sentiment-based market regime detection, and topology-aware deep learning. Founder and CEO of DataSalt.ai, a boutique AI/ML consultancy, and creator of the Formal educational platform series (formalml.com, formalstatistics.com, formalcalculus.com) for rigorous mathematics and ML content.

Education

Master of Science in Data Science

Southern Methodist University (SMU) — Dallas, TX | 2025 – Aug 2026 GPA: 3.63

Coursework: Artificial Intelligence (DS 7331), Database Management Systems (DS 7330), Applied Statistics I & II (DS 6730/6731), Machine Learning II

Capstone: Sentiment-based market regime detection using ensemble transformer models (BERT-family + GARCH(1,1) + Statistical Jump Model); advised by Dr. Lin

Master of Arts in English

Texas A&M University–Central Texas (TAMUCT) — Killeen, TX | 2023 – 2024

Bachelor of Arts in History

Texas A&M University (TAMU) — College Station, TX | 1999 – 2004

Professional Experience

Founder & Chief Executive Officer

DataSalt.ai (datasalt.ai) — Austin, TX (Hybrid) | Feb 2025 – Present

- Founded and led a boutique AI/ML consultancy serving small and mid-sized businesses across Texas, specializing in NLP pipelines, predictive modeling, and data-driven decision systems
- Head of Data Science and principal consultant: retrieve and analyze sensitive client data across retail, agriculture, legal, healthcare, and other verticals; transform inputs into actionable insights through intuitive data storytelling
- Published portfolio of 10 case studies and 7 technical blog posts on datasalt.ai (Next.js 14 / Tailwind CSS); built SaltyDog, an AI chatbot with a custom avatar, and a programmatic hero image generator
- Built finrag.io, a multimodal financial RAG system (Gemini Embeddings 2, Qdrant, Cloudflare R2, FastAPI on Fly.io, Claude Sonnet for synthesis, Next.js on Vercel)
- Created the Formal educational platform series — formalml.com (ML-specific mathematics), formalstatistics.com (probability and inference), formalcalculus.com (calculus and analysis) — built on a shared Astro 5 / React 18 / MDX / Tailwind CSS 4 / D3.js v7 / KaTeX stack with 130+ published topics across 29 curriculum tracks

Senior Web Developer & Full-Stack Engineer

Fullsteam / Fullsteam Marketing — Austin, TX (Hybrid → Remote) | Jan 2015 – Dec 2025

- 11-year tenure across three role transitions: Web Developer (Jan 2015) → Full-Stack Engineer at Fullsteam Marketing (Jan 2016, concurrent) → promoted to Senior Web Developer (Jan 2019)

- Architected and maintained React-based web applications aligned with the company's digital strategy; executed full front-end and back-end development using HTML, CSS, JavaScript, and Python
- Managed database systems and cloud infrastructure (AWS); contributed to digital growth strategy through data-informed development decisions
- Optimized site performance for SEO and user experience, significantly boosting visibility and search engine rankings of Fullsteam's digital properties
- Collaborated with cross-functional teams to troubleshoot issues, integrate new features, and deliver data-driven, user-centered products

Independent Web Development Consultant

Self-Employed — Austin, TX (Hybrid) | 2004 – Present

- Deliver end-to-end custom web solutions specializing in React, responsive design, and UX/UI for clients ranging from startups to established enterprises across two decades of practice
- Manage every stage of the project lifecycle: requirements gathering, technical architecture, front-end and back-end development, API and back-end integration, performance optimization, accessibility, and ongoing maintenance
- Translate complex business needs into clear technical solutions; provide strategic guidance on best practices to non-technical stakeholders
- Emphasize clean, maintainable code, intuitive user experiences, and mobile-first responsive layouts that perform seamlessly across devices

Senior Web Developer

Amaru Motors LP dba Charlie Clark Nissan — Harlingen, TX | Mar 2009 – Jan 2015

- Developed, optimized, and maintained user-facing websites and web applications using HTML, CSS, JavaScript, and Python, aligned with the company's digital strategy
- Led mobile-first web design initiatives and contributed to the initial design and development of the company's mobile app
- Optimized site performance for SEO and usability, improving search engine rankings and customer engagement across digital properties
- Collaborated with cross-functional teams to troubleshoot issues, implement new features, and meet strict standards for quality, performance, and functionality

Web Developer

Wells Fargo — San Antonio, TX | Mar 2004 – Mar 2009

- Developed, optimized, and maintained secure, user-friendly web applications supporting Wells Fargo's online banking experience for account management, transactions, and customer self-service
- Worked across the stack with a back-end focus using HTML, CSS, JavaScript, and Python; built reliable, high-performing features under strict regulatory and security requirements
- Collaborated with product, design, and security teams to ensure compliance with banking standards while optimizing backend services and site performance for response time, stability, and usability
- Aligned technical solutions with broader business and regulatory objectives to enhance the online banking customer journey and reinforce Wells Fargo's long-term digital strategy

Research Interests

Topology-aware deep learning; time-series data mining; natural language processing applied to financial markets; sentiment-based market regime detection; topological data analysis (persistent homology, Mapper algorithm); ensemble transformer models; scalable ML systems; distributed computing architectures; computational text analysis.

Academic Projects & Research

Sentiment-Based Market Regime Detection

SMU Capstone (market-sentiment.io) | 2025 – Present

- Building a two-layer pipeline combining GARCH(1,1) volatility modeling with a Statistical Jump Model for cross-asset sentiment regime detection from news and social media text
- Ensemble approach integrates BERT-family transformers with classical time-series methods; dashboard deployed at market-sentiment.io with hybrid RAG + live-context chatbot
- Advised by Dr. Lin; capstone paper substantially revised with tracked changes across Methods, Results, Discussion, Conclusion, and dataset appendix

Persistent Homology for Financial Regime Detection

ML 2 Final Project | 2025

- Applied topological data analysis methods (persistent homology, Vietoris–Rips complexes, persistence diagrams) to financial time-series classification
- Produced full video script, Jupyter notebook, and PowerPoint presentation deck

Machine Learning Classification

DS 7331 — Artificial Intelligence | 2025

- Implemented classification algorithms, achieving a 93.72% F1-score on the Dry Bean Dataset

Clustering Analysis

DS 6372 — Applied Statistics | 2025

- Developed and evaluated K-means, hierarchical clustering (DIANA/AGNES), and DBSCAN algorithms on the Bank Marketing Dataset

NoSQL Database Design

DS 7330 — Database Management Systems | 2025

- Designed and implemented MongoDB database solutions for social media applications; received excellent assessment scores

Independent Projects

Formal Educational Platform Series — formalml.com · formalstatistics.com · formalcalculus.com

2025 – Present

- Built a unified network of three sister educational platforms for rigorous mathematics and ML content, deployed on Vercel and built on a shared stack: Astro 5, React 18, MDX, Tailwind CSS 4, D3.js v7, KaTeX (pnpm). GitHub: jonx0037/formalML, jonx0037/formalStatistics, jonx0037/formalCalculus
- **formalml.com — Mathematical foundations of modern machine learning** (65+ published topics across 13 tracks): Topology & TDA, Linear Algebra, Probability & Statistics, Optimization, Differential Geometry, Information Theory, Graph Theory, Category Theory, Supervised Learning, Unsupervised & Generative, Nonparametric & Distribution-Free, Bayesian & Probabilistic ML, Learning Theory & Methodology
- **formalstatistics.com — Pure statistics curriculum** (32 published topics across 8 tracks): Foundations of Probability, Core Distributions & Families, Convergence & Limit Theorems, Statistical Estimation, Hypothesis Testing & Confidence, Regression & Linear Models, Bayesian Statistics, High-Dimensional & Nonparametric
- **formalcalculus.com — Calculus and analysis curriculum** (32 topics across 8 tracks): Limits & Continuity, Single-Variable Calculus, Multivariable Differential Calculus, Multivariable Integral Calculus, Sequences & Series, Ordinary Differential Equations, Measure & Integration, Functional Analysis Essentials

- Combined: 130+ published topics across 29 tracks. Established editorial voice guidelines applied uniformly across all three sites: informed-peer tone, collaborative "we" pronoun, assumed knowledge of linear algebra/calculus/probability, explicit notation introduction
- Two-deliverable workflow per topic: Jupyter notebook for mathematical content + Claude Code handoff brief for site implementation

finrag.io — Multimodal Financial RAG System

2025 – Present

- Built a multimodal financial document intelligence platform using Google Gemini Embeddings 2, Qdrant vector database, Cloudflare R2 for storage, FastAPI backend on Fly.io, and Next.js frontend on Vercel
- Integrated Claude Sonnet for LLM synthesis and Gemini Flash for TTS audio generation of earnings call transcripts
- Published companion case study and blog post on datasalt.ai

CounselOS — Multi-Agent AI Legal Matter Intake

2025

- Built a multi-agent AI legal matter intake system (FastAPI + Next.js) featuring a five-agent pipeline with a custom state machine orchestrator
- Deployed to Railway (backend) and Vercel (frontend); GitHub: jonx0037/counselos
- Demonstrated multi-agent orchestration patterns for portfolio and technical interview preparation

datasalt.ai — AI/ML Consultancy Platform

2024 – Present

- Built and maintains datasalt.ai on Next.js 14/Tailwind CSS/Vercel with 10 case studies covering South Texas verticals (boat sales, beach resort, shrimping, citrus/agriculture, healthcare, construction, law firm, used-car dealership)
- Developed SaltyDog, an AI chatbot with a custom avatar (Amelie the French Bulldog), a programmatic hero image generator, and 7 technical blog posts

Works in Progress

Applied NLP for Finance: Building Market Intelligence Systems with Language Models — Book manuscript (12–13 chapters) with companion GitHub repository (applied-nlp-finance); dual MIT/CC BY-NC 4.0 license; covers knowledge graphs, LLM workflows, and applied financial NLP.

Technical Skills

Programming Languages: Python, R, JavaScript (ES6+), SQL **Frameworks & Libraries:** React, Next.js, Astro, Node.js, FastAPI, Flask **ML / AI Tools:** Scikit-learn, Hugging Face Transformers, BERT, GARCH, TDA (GUDHI, Ripser, giotto-tda), Qdrant, Ollama, ensemble methods **Data & Visualization:** D3.js v7, matplotlib, Plotly, KaTeX, MDX, pandas, NumPy **Databases:** MongoDB, NoSQL systems, SQL **Cloud & Infrastructure:** Amazon Web Services (AWS), Vercel, Fly.io, Cloudflare R2, Railway, Docker **Specializations:** Full-stack web development, NLP, topological data analysis, machine learning engineering, statistical analysis, time-series analysis, data visualization, UX/UI design

Additional Information

- 20+ years of professional web development experience spanning financial services, automotive retail, marketing technology, and AI/ML consulting, with deep expertise in React/Next.js
- Native financial-services domain experience from five years at Wells Fargo's online banking platform, complemented by current applied financial NLP research (finrag.io, *Applied NLP for Finance*)

- Strong interdisciplinary foundation combining humanities scholarship (M.A. English, B.A. History) with applied technical expertise
- Pursuing a PhD in data science/computer science with research interests in topology-aware deep learning and time-series data mining
- Proficient in translating complex technical problems into accessible written and verbal communication