

Shivam Mokha

[LinkedIn](#) | [Github](#) | [ResearchGate](#)

160 E Station Drive Square
Pittsburg, PA 15219 USA
+ 1 (510) 862 9799
shivammokha94@gmail.com

EDUCATION

Arizona State University | Arizona, USA

Master of Science, Business Analytics (3.80/4.00)

August 2018 - May 2019

Nagpur University | India

Bachelor of Engineering, Electronics (74.5%)

August 2012 - June 2016

EXPERIENCE

Aurora Innovation | Pittsburg, Pennsylvania, USA

Staff Data Scientist, Safety

March 2026 - Present

Led end-to-end safety analytics strategy for autonomous systems, integrating large-scale proprietary data (vehicle logs, sensor, testing) with public crash datasets (FARS, CRSS, CISS, Texas CRIS) to build predictive risk models, leading indicators, and system-level behavioral simulations; produced crash-rate analysis (per 100M VMT) used in California DMV permit renewal and federal/state regulatory filings.

Designed and shipped an end-to-end, LLM-powered safety evaluation platform (FastAPI, React/TypeScript, Docker/Kubernetes) - an incremental Claude/Bedrock scoring engine with skip logic and parallel workers plus a four-tool RAG assistant - serving ~40ms responses and replacing ad-hoc notebooks with a production product used across the safety org.

Architected scalable data infrastructure and risk frameworks - automated pipelines, baseline risk-coverage models, cross-industry benchmarking, and executive dashboards to quantify, monitor, and communicate operational safety performance - and re-platformed operational analytics (Mode → Holistics; Workday/Samsara/Jira/Redshift) with net-new SLA and certification metrics.

Influenced executive and cross-functional decision-making by presenting rigorous technical analyses to senior leadership, regulators, and industry forums; set technical direction for safety modeling and forecasting, established org-wide ownership standards for shared safety tooling as DRI, and mentored junior data scientists.

Lucid Motors Inc. | Fremont, California, USA

Senior Data Scientist, Software Engineering

August 2021 - March 2026

Accelerated Lucid Air's European launch by partnering with cross-functional US/EU teams to define autonomy, charging, and release metrics, and ran validation drives across 8 countries (ES, FR, IT, CH, DE, DK, SE, NO) that surfaced [N] software improvements across charging, navigation, and infotainment.

Architected and maintained Python backend data-extraction pipelines integrating GitLab, JamaCloud, and Artifactory (~1,300 LOC each), and re-engineered legacy code to cut time/space complexity - delivering faster, more consistent data feeds that powered timely, data-driven dashboards.

Built and productionized ML and statistical models on historical vehicle data to predict component failure, paired with real-time anomaly detection that flagged deviations from normal operating patterns and enabled early intervention before field issues.

Partnered with Infrastructure and ADAS teams to design a real-time sensor-data logger database, filtering and summarizing high-volume vehicle telemetry into live statistical views that advanced Lucid Air's autonomous capabilities.

Developed backend systems and executive dashboards used by PMs, program managers, directors, and engineers to steer feature development-to-deployment across every software release for Lucid Air and Gravity, surfacing critical gaps in requirements traceability, test coverage, and customer-reported feature issues.

Set technical direction and standards for the Quality Engineering org - mentoring engineers in full-stack development, authoring reusable technical documentation, and running technical interviews to scale hiring.

Technical Skills & Tools

Python, SQL, Redshift, Pandas, Numpy, SciPy, TensorFlow, PyTorch, Scikit-learn, Apache Spark, Hadoop, Airflow, Tableau, Git, Docker, Kubernetes, BigQuery, RestAPI, Flask, Django, JupyterHub, Cassandra, SQLAlchemy, AWS, GCP, Azure, Anaconda, DBT, Holistics, Mode

Machine Learning & AI

Supervised & Unsupervised Learning, Deep Learning, Neural Networks, NLP, Time Series Analysis, Feature Engineering, Model Deployment (MLOps), A/B Testing, Experimentation, Reinforcement Learning, Large Language Models (LLM), RAG (Retrieval-Augmented Generation), LLM Fine-Tuning (LoRA, RLHF), Claude Code, Co-Pilot, Gemini

Data Engineering & Analytics

ETL/ELT Pipelines, Data Warehousing, Data Cleaning & Preprocessing, Data Mining & Feature Selection, Big Data Processing, Real-time Data Streaming

Certificates

MIT – Calculus 1A: Differentiation
Introduction to Big Data (Coursera/UCSD)
Machine Learning (Coursera/Stanford University)
Core & Advanced JAVA (Microsoft)
C++ (Microsoft)

Waymo LLC (Google Inc.) | Mountain View, California, USA

Engineering Associate, Planner/Engineering

October 2019 – July 2021

Partnered with the Planner Eval team to [design new behavioral metrics and analyzers](#), mining [large-scale simulation](#) and log data to statistically quantify how the self-driving stack responded to algorithm changes via rigorous test-vs-baseline comparison.

Built analysis-ready database tables from raw sensor data and designed, executed, and analyzed [A/B experiments on UI/UX changes](#), [driving measurable improvements](#) in rider experience across [N] releases.

Managed complex data pipelines and metric dashboards end-to-end, partnering with cross-functional teams and TPMs to shape workflows and [consistently ship deliverables](#) within quarterly timelines.

Collaborated across teams to define new datasets that [improved self-driving performance and reduced collision risk](#), diagnosing vehicle behavior throughout the SDLC to surface trends and problematic areas before they reached the road.

Find Your Influence, Inc. | Tempe, Arizona, USA

Data Science Intern

January 2019 – May 2019

Worked on a capstone project for a cloud-based company that enables businesses to manage their brand ambassador and social influencer campaigns.

Created Name Entity Recognition & Word Sense Disambiguation algorithms using Python 3.6 to optimize keyword search. Performed [Natural Language Processing \(NLP\) on multiple sets of text data](#) to be used as algorithm input parameters and created multiple domain-based dictionaries. Perform [text analytics to develop search capabilities](#) (Search Engine Optimization) based on the keyword frequency and sentiment of an influencer's social media postings.

Accenture PLC | Bangalore, India

Data Analyst

August 2016 – July 2018

Designed the [process and developed metrics to produce data visualizations reports](#) and dashboards using Tableau and QlikView to help drive key business decisions. Analyzed existing product and purchase data across 1.5 million consumers; Compared against new vendor products and recommended the most value-add products for the business.

Managed projects for deploying CS-Oracle applications across client environments. Identified key code and data issues across environments and developed new solutions to cater to business needs. Collaborated with business and technology stakeholders to help define new products or enhancements to existing products.

RESEARCH PUBLICATIONS

Sinha, A. & Mokha, S. (2017). [Classification & fraud detection in the finance industry](#). International Journal of Computer Application, 176 (3), pp. 45-52

Bandhaokar, S., Jangir S., Mokha S. & Pimple S. (2016). [Design of Flexible Microstrip Patch Antenna of 2.4 GHz Operating Frequency Using HFSS](#). International Research Journal of Engineering and Technology, 03(05), pp. 638-644

Choure M., Mokha S., Naik, R., Nikam K. (2016). [Automated Counting of HMTVs](#). International Research Journal of Engineering and Technology, 03(05), pp. 239-241