

AANSH SHAH

Security, ML, Infrastructure & Distributed Systems Engineer

aansh.shah@gmail.com • aanshshah.com • New York & Boston, MA

EDUCATION

Brown University — M.S. Computer Science (2020), B.S. Applied Mathematics & Computer Science (2019)

PROFESSIONAL EXPERIENCE

Software Development Engineer - Security & ML Infrastructure at Amazon (Alexa & AGI) *July 2020 - July 2025*

- Architected **OpenTelemetry observability platform** using Golang microservices and Kinesis streams → **\$3.2M quarterly revenue increase**; automated customer feedback processing **8hr→3min (97% reduction)** using Apache Airflow and ML sentiment analysis
- Built **Rust-based data compaction service** with custom LZ4 compression algorithms transforming petabyte-scale data infrastructure → **63% cost reduction (\$1.2M/year)**; reduced S3 throttling **90%**, automated recovery workflows **5hr→5min**
- Designed **enterprise IAM platform** with OAuth 2.0/SAML workflows processing **8M+ daily auth requests at 99.999% availability**; reduced dataset provisioning **5 days→3 hours** for 40+ teams using CDK (TypeScript) automation
- Led **security audits for 15 PII services** creating reusable compliance frameworks with automated vulnerability scanning → **zero high-severity findings**, **45% faster reviews**, **28% fewer incidents**
- Streamlined **ML model deployment** with Git-based artifact tracking and containerized CI/CD pipelines **3wk→2wk**; built **federated learning platform** using SageMaker, PyTorch, and TensorFlow on Kubernetes reducing voice model onboarding **30%**
- Implemented **real-time anomaly detection system** using Apache Flink and TensorFlow, monitoring 500+ microservices with sub-second alerting reducing mean time to resolution **60%**

Founder & CEO at Nexus Legal AI

July 2025 - Present

- Building an **AI-powered intelligence platform** transforming how legal professionals conduct research and case strategy. Developed proprietary **pattern recognition technology** reducing research time by **50x** while improving prediction accuracy. Led product and engineering from concept to pilot, integrating **large language models** with structured case law databases to deliver actionable insights

AI Instructor at Inspirit AI

June 2020 - April 2021, June 2025 - Present

- Teach AI fundamentals to 50+ high school students annually, building real-world projects in healthcare/climate/social good using Python, TensorFlow, and Scikit-learn; lead 1:1 research mentorship resulting in 8 student publications
- Developed curriculum covering neural networks, computer vision, and NLP with hands-on labs; created interactive Jupyter notebooks and deployed student models to cloud platforms

Founder & CEO at AI Recruitment Platform

2019-2020

- Built **NLP-based candidate screening platform** using spaCy and BERT models with Twilio Voice API integration → **1000+ candidates/month processed**; created telephony SDK reducing partner integration **weeks→days**, improving candidate quality **40%**
- Developed **automated interview scoring system** using speech-to-text and sentiment analysis; deployed on AWS Lambda with DynamoDB for real-time candidate ranking

Teaching Assistant at Brown University

2016-2019

- CSCI1410: AI** (Head TA) - Managed 18 TAs for 140 students; built Python auto-grading infrastructure using Docker containers; created search algorithm assignments (BFS/DFS/A*)
- DATA2040: Deep Learning** - Co-designed new course curriculum; managed 7 TAs; developed CNN/RNN/Transformer assignments using Keras/TensorFlow with GPU clusters
- APMA340: Differential Equations** - Created computational models using NumPy/SymPy; developed accessible scripting tutorials for numerical methods and visualization with Matplotlib

Software Engineering Intern

2018-2020

MSBAI (Jun-Jul 2020)

- Built ML regression models for restaurant sales predictions using XGBoost; deployed Docker containers on AWS using Fargate, ECS, and Lambda with automated CI/CD

Upserve (Jan-Apr 2020)

- Built ETL pipeline evaluating ML model predictions using DynamoDB, SQS, and Apache Spark; analyzed time series performance data with statistical forecasting

Wayfair (2019)

- Finance: Created React/PHP/SQL management system; 3D: Developed Unity mobile app with depth sensors for AR/VR catalog digitization using C# and computer vision

Oblix VR (2018)

- Built multi-user social VR applications using Unity/C# for Oculus/GearVR; integrated Watson Speech APIs for voice navigation and spatial audio processing

TECHNICAL SKILLS

Programming Languages

Python, TypeScript, Java, Rust, JavaScript, C#, Go, SQL, Bash

Cloud & Infrastructure

AWS (EC2, S3, Lambda, DynamoDB, Kinesis, Flink, Fargate, ECS), Docker, Kubernetes, Terraform

ML/AI & Data

TensorFlow, PyTorch, Federated Learning, NLP, Computer Vision, Apache Spark, Kafka, scikit-learn

Security & Monitoring

OAuth 2.0, SAML, RBAC, OpenTelemetry, Prometheus, Grafana, vulnerability scanning

LEADERSHIP & RESEARCH

Data Scientist - Brown Political Review (2017-2019)

Led 17 data journalists; built "First Drafts" interactive platform for data-driven articles using D3.js, Python, and PostgreSQL for political analysis

Founder - GameCraft Charity (2011-Present)

Providing video games and consoles to underprivileged youth (14+ years); organized 25+ community events, distributed 500+ gaming systems

Coach - RI Urban Debate League (2015-2019)

Coached underserved students in policy debate; managed tournament operations for 200+ participants; students achieved 85% college acceptance rate

PUBLICATIONS & SPEAKING

First Author Publications:

"Comparing Global with Disease-specific ML Readmission Models" - AMIA 2020
"Graph Embedding Priors for Multi-task Deep RL" - NeurIPS KR2ML 2020
"Stackelberg Punishment and Bully-Proofing Autonomous Vehicles" - ICSR 2019

Conference Speaker:

"Scaling Real-time Stream Processing for ML Pipelines" - Apache Flink Forward 2023

AWARDS & PROJECTS

AWARDS

1st Place Intel ISEF (APA) **NYS Debate Finalist**

KEY PROJECTS

Chrome Extension - Toxicity Classifier

Led 4-person team building LSTM toxicity classifier processing millions of YouTube comments using TensorFlow; deployed Flask API with 95% accuracy and sub-200ms latency

Medical AI Platform

ICU readmission prediction using ensemble methods on MIMIC-III dataset; achieved 0.84 AUC using XGBoost and feature engineering with clinical time series data

Real-time Analytics Dashboard

Twitter sentiment analysis using D3.js and PubNub for live visualization; processed 10K+ tweets/minute with streaming NLP pipeline and interactive geographic mapping

DQN Atari Agent

Deep reinforcement learning game agent using TensorFlow and CNNs; implemented experience replay and target networks achieving superhuman performance • **20+ GitHub repos**