

Abhiram Vadlapatla

908-392-4256 • v.abhiram97@gmail.com • linkedin.com/in/abhiram-vadlapatla/ • github.com/abhiramvad

EDUCATION

M.S. Computer Science, Arizona State University, Tempe, AZ

Dec 2025

Relevant Coursework: Artificial Intelligence, Statistical Machine Learning, Data Intensive Systems for ML

B.Tech in Computer Science, Manipal Institute of Technology, India

Aug 2018

Relevant Coursework: Data structures and Algorithms, Object Oriented Programming, Software Engineering

TECHNICAL SKILLS

Programming Languages: Python, Java, Typescript, Javascript, Bash, C++

Backend Frameworks/Libraries: Flask, SQL Alchemy, Celery, Django, Spring Boot, Hibernate, Nodejs, OAuth2.0, RabbitMQ

UI Frameworks/Libraries: React, Angular, NX, HTML, CSS, SCSS, Bootstrap, Webpack, Yarn

Databases: SQL Server, Prometheus, Elasticsearch, MySQL, Postgres, Big query, SQLite,

Cloud/SaaS: Terraform, Cloudfoundry, AWS ECS, Docker, Kubernetes, Lambda, AWS Kinesis, Fluent bit, Newrelic, Grafana

Protocols: REST, GraphQL, gRPC, WebSockets, Server-Side events (SSE), SMTP

PROFESSIONAL EXPERIENCE

Goldman Sachs, Bangalore, India:

Jul 2022 - Dec 2023

Senior Software Engineer

- Implemented IAM knowledge bot using **React, Socket IO, Flask, Langchain**, and **GPT-3.5** for 30% faster query resolution.
- Accelerated go-live time from 1 week to 48 hours by facilitating instant thematic customizations and streamlining change requests with a **React Drag and Drop** live preview tool for IAM consumers.
- Achieved 100% data transparency and real-time analytics within a 1-minute window by implementing a data ingestion pipeline processing 2GB of data hourly with **Python, Apache Airflow, AWS Kinesis, AWS Lambda, BigQuery, and Grafana**
- Shipped a high-performance **JS analytics library** using **web beacon API** for **client-side observability**, boosting dev efficiency 1.5x via **builder pattern** and React hooks for capturing user journey events.
- Enhanced the security profile of the **IAM platform** by integrating **ThreatMetrix** into the Spring Boot server, achieving a 30% improvement in fraudulent login detection and abstracting the low-level API to support future integrations.
- Keynote speaker at Goldman Sachs**, sharing insights with 200 developers on Identity and Access Management.

Philips Healthcare, Bangalore, India:

Jan 2018 - Jun 2022

Software Technologist

Apr 2022 - Jun 2022

- Halved CI build time to 16.54 seconds by utilizing **Yarn's** offline mirror feature, hosting 312 MB of locally compressed dependencies for a 1.3 GB Node modules repository.

Senior Software Engineer

Apr 2020 - Mar 2022

- Increased user engagement by 50% building a real-time **chatbot** using **React, DialogFlow, Soul Machines SDK**, and **Express**.
- Spearheaded code quality measures: 100% test coverage(**pytest**), 5% duplication limit (**jscpd**), mutation testing via **git hooks**
- Implemented database integration with **Flask, SQLAlchemy**, managing database migrations using Alembic and Flask Migrate, resulting in a 50% reduction in deployment times and database schema management errors.
- Integrated **New Relic** with Flask app to support APM and latency metrics, achieving a 45% improvement in performance monitoring and issue resolution times.

Software Engineer

Jul 2018 - Mar 2020

- Improved developer and third-party adoption of the API services available by 20% by designing an API monetization ecosystem and developing a developer portal using **Kong API gateway** for third-party API consumption.

PROJECTS

EpicDiffusion Playground - Text to video generation app

Apr 2024

- Created a creation tool using GPT-4 prompt injection for text-to-video diffusion (I2VGen, Stable Diffusion), reducing latency by 40% with gRPC and debouncing.
- Enhanced model pipeline spatial and temporal coherence with Prompt Weighting, Free U, and Chain of Thought, boosting CLIP Score from 0.21 to 0.26 for out-of-distribution data.

Video Spotlight - Semantic Search in Videos using BERT and Timesformer

Apr 2024

- Engineered a React video player featuring custom markers and AI powered in-video search with 0.5ms latency, supporting lexical, fuzzy, and semantic queries.
- Optimized MobileNet and Timesformer video AI models, achieving 4x size reduction (488MB to 126MB) through int8 quantization for efficient cloud and on-device performance.