

# Yosha Mundhra

San Francisco Bay Area | mundhra.yosha09@gmail.com | [LinkedIn.com/in/yoshamundhra](https://www.linkedin.com/in/yoshamundhra) | [GitHub.com/YoshaM09](https://github.com/YoshaM09) | [Yoshamundhra.com](https://yoshamundhra.com)

## SKILLS

**Languages** *Proficient* Python(5yrs) · SQL(5yrs) · Java(3yrs) *Intermediate* JavaScript(2yrs) *Beginner* Typescript(1yr) · React(1yr)

**Software** Azure · Langchain · TensorFlow · JupyterNotebook · Pandas · PySpark · Numpy · Apache Spark · VS Code · SSMS

## EXPERIENCE

### Data Scientist

Kaiser Permanente via Spotline Inc

Jan 2024 - Present

San Jose, CA

- Designed an ETL pipeline in Azure Databricks to build an identity data warehouse, for fast & accurate IAM analytics
- Leveraged PySpark to accelerate data preprocessing aligned with data governance framework improving data reliability
- Analyzed identity data using SQL & Python to detect and resolve abnormalities, enhancing system integrity by 70%
- Improved LLM chatbot responses and reduced hallucinations by 20% via prompt engineering & RAG with LangChain
- Automated and tested CyberArk SaaS upgrades using Python and Selenium, reducing deployment time by over 85%

### Software Engineering Resident

Headstarter

Apr 2025 - Present

San Jose, CA

- Building 12+ machine learning, ai-engineering and full-stack projects in fast-paced software team environments
- Engineered an AI voice assistant with React, Groq API, LLaMA 3, and Deepgram STT/TTS, for human-like interactions
- Saved users time by designing a smart chat feature to answer natural language queries based on content of a video
- Designed MCP server for 100% automated trip planning integrating Gemini API, Exa Search, and Cursor as MCP client

### Data Science Intern

CDK Global

Jun 2022 - Aug 2022

San Jose, CA

- Identified top 10 profitable products and customers, improving decision-making for new service pitches by 50%
- Analyzed and evaluated profit/loss margins post-migration using SQL & Python, leading to improved financial strategies
- Created Power BI dashboards using insights from CSV and Excel data, boosting stakeholder data accessibility by 60%

### Software Engineer

Tata Consultancy Services

Dec 2018 - Aug 2021

India

- Engineered 55+ features using Python, Java, & JavaScript to enhance [Humana's mail-order pharmacy platform](#) by 80%
- Streamlined prescription processing from provider submission to doorstep delivery, improving patient convenience
- Delivered all user stories within 2 weeks Azure sprint cycles, maintaining 100% agile compliance with project releases
- Resolved 30+ defects via root cause analysis, debugging, and robust error handling reducing production bugs by 70%
- Mentored 15 new team members through code reviews & documentation, elevating overall code quality team efficiency

### Software Engineer

ITC Infotech

Jun 2018 - Nov 2018

India

- Integrated REST APIs with Spring for [ITH TravelSMART](#), a B2B platform offering 20+ travel services like flights, & hotels
- Connected to 10+ APIs including Amadeus GDS, streamlining and boosting corporate travel booking efficiency by 60%

## AI/ML PROJECTS

**MCP Server for Travel Planning** | Open-Source, LLM, MCP, NLP (~40 hours) - [GitHub Link](#)

Jun 2025 - Jul 2025

- Developed an AI-powered MCP server to automate trip planning by integrating Gemini for natural language queries, Exa Search for travel data, and Cursor as the client, significantly improving user experience and planning efficiency.

**Multimodal Video Analysis** | Open-Source, LLM (~40 hours) - [GitHub Link](#)

May 2025 - Jun 2025

- Built a React and TypeScript frontend to enable natural language chat with YouTube videos using Gemini, generating timestamped answers to redirect to relevant sections, saving users time to find specific content in long videos.

**Toxic Comment Classification** | Open-source, NLP, Keras, TensorFlow (~30 hours) - [GitHub Link](#)

May 2023 - Jun 2023

- Designed pipeline for data pre-processing using Textthero to clean text data, form tokens and apply padding
- Built text classification model with cell state LSTM, dense layers, batch normalization layers and dropout layers to avoid overfitting. Trained 20 epochs with 35 batch size to attain a test accuracy of 0.90.

## EDUCATION

**College of Science, San Jose State University (SJSU)**

Dec 2023

Master of Science in Data Science (Computer Science and Mathematics) | 3.5 GPA

**College of Engineering, SRM University (SRMIST)**

May 2018

Bachelor of Science in Information Technology | 3.8 GPA

## ACTIVITIES

**AUTHOR:** Mundhra, Yosha, "An Incentivization Mechanism towards Network-Slice as a Service". DOI: [doi.org/10.31979/etd.y8sf-pd42](https://doi.org/10.31979/etd.y8sf-pd42)