Contact

9530043790 (Mobile)
nothefakedevesh@gmail.com

www.linkedin.com/in/devxm
(LinkedIn)

Top Skills

Slack APIs

rrweb

Redux Thunk

Languages

Hindi (Native or Bilingual)

English (Full Professional)

Certifications

React, Node.js, Express, MongoDB -
The MERN FullStack Guide

Next.js Course

Devesh Meena

Software Engineer | Applied AI @Orca

India

Summary

At Orca AI, my team and I have crafted a cutting-edge RAG system, harnessing the power of OpenAI and Anthropic LLMs, which we've meticulously optimized for peak performance. Our robust evaluations and deployment innovations have not only enhanced user interaction but have also streamlined real-time Q&A processes.

Complementing this practical expertise is my academic journey at NIT Jalandhar, where a Bachelor's in Computer Science is on the horizon. My commitment to full-stack development is evident from my impactful tenure at HopStair®, where I significantly boosted user engagement and operational efficiency through technical enhancements and optimized GitHub workflows.

Experience

Orca AI

AI Engineer, Founding Team

December 2023 - Present (1 year 4 months)

San Francisco, California, United States

Founding Engineer at Orca AI—wore multiple hats across Software Engineering, AI, and Full-Stack Development. Worked with incredible people, developed a strong product mindset, and combined it with technical expertise to build user-centric solutions.

1. AI & Multi-Agent Systems
- Built autonomous multi-agent systems capable of complex reasoning, self-correction, and optimal task execution.
- Expertise in prompt engineering methodologies like Zero/Few-Shot, Chain of Thought, and more for open/closed-source LLMs & LRMs.
- Designed and implemented MCP-based agent communication, enabling seamless interaction between AI agents.

- Developed efficient RAG (normal & graph-based) pipelines for fast, accurate information retrieval from vector DBs like Pinecone, Weaviate, MongoDB, and Neo4j.
- Applied advanced chunking, embedding, and metadata strategies for optimizing vector search performance.
- Worked with libraries/frameworks like Langchain, LangGraph, LlamaIndex, HuggingFace, and experimented with Ollama, ONNX, SiLLLM for local fine-tuning.
- Prefer raw LLM APIs over abstracted frameworks for greater control, flexibility, and customizability in AI workflows.
- Implemented guardrails, data cleaning, and custom prompt engineering to mitigate prompt injection and abuse.
- Used Langfuse for LLM observability/evaluation, along with Datadog, Helicone, and other monitoring tools.

2. Full-Stack & DevOps
- Built interactive, secure frontends using React, TypeScript, with a focus on seamless UX and chat interfaces.
- Deployed multi-agent AI systems at scale using AWS, optimizing for performance and reliability.
- Developed Slack integrations using Slack APIs for workflow automation and enterprise use cases.

buildspace
S5 Cohort
June 2024 - July 2024 (2 months)
San Francisco Bay Area

HopStair®
Full-stack Developer
August 2023 - January 2024 (6 months)
London, England, United Kingdom

- Developed and maintained React Native features and an admin portal using Firebase & React, enhancing user engagement and team collaboration.
- Optimized GitHub workflows for Expo and PlayStore deployments, reducing deployment times.
- Contributed technical insights in weekly standup meetings, driving project progress.

Fundwave

Full Stack Developer
May 2023 - July 2023 (3 months)
Gurugram, Haryana, India

- Engineered scalable, secure authentication microservices with TypeScript, Java, Node.js, and Keycloak IAM, reducing response times by 24%.
- Collaborated with cross-functional teams to design efficient, maintainable code architectures for long-term scalability.

Shardings
Frontend Developer
October 2022 - January 2023 (4 months)
New Delhi, Delhi, India

Microsoft
Engage 2022 Mentee
May 2022 - June 2022 (2 months)

_____

Education

NIT Jalandhar
Bachelor of Technology - BTech, Computer Science · (November 2020 - May 2024)

Subodh Public School
High School Diploma · (2018 - 2020)