

Sriharsh Sai

 Email: sriharsh1997@gmail.com  Phone: +1-647-528-6479

Summary

Software Engineer with experience in full-stack development across AI, healthcare, and enterprise sectors.

- **Specialized AI Expertise:** Architecting Retrieval-Augmented Generation (RAG) systems and deploying autonomous agents using OpenAI and Claude SDKs.
- **Cloud & Backend:** Skilled in developing server-side applications and microservices using **Node.js**, **PostgreSQL**, **MySQL**, **AWS Lambda**, and **Docker**.
- **Modern Frontend:** Expert in building performant user interfaces with **React**, **Next.js**, and **Tailwind CSS**.
- **Versatile Foundations:** Proficient in **C++**, **C#**, and **TypeScript** for diverse system requirements.

Experience

Software Engineer

 **SpinSci Technologies LLC** | *January 2026 – Present*

- **Enterprise Application Development:** Developing and integrating high-quality, mission-critical health-tech applications using JavaScript, React, and Node.js to deliver scalable business solutions.
- **Microservices & Cloud Computing:** Implementing scalable microservices and serverless functions via AWS Lambda and Docker to optimize system performance and resource efficiency.
- **Technical Architecture:** Translating complex user requirements and operational specifications into functional system designs and robust product-based business solutions.
- **Database Optimization (PostgreSQL):** Managing the integration of software components with relational databases, ensuring high data accuracy and resolving complex backend performance issues.
- **Lifecycle Management & Security:** Maintaining high-availability applications through **Git-based version control**, implementing secure code updates and troubleshooting production environments.
- **Technology Mentorship:** Providing high-level technical guidance and subject matter expertise in emerging technologies to ensure scalable and efficient software delivery.

Full Stack Developer

 **XceptAI** | *November 2024 – January 2026*

- **Generative AI Engineering (OpenAI):** Engineered an intelligent, conversational SMS platform and Telegram bots using the **OpenAI SDK**, automating complex customer interactions with natural, context-aware responses.

- **Financial Analysis (Anthropic Claude):** Leveraged **Anthropic's Claude API** and its large context window capabilities to ingest, analyze, and summarize massive financial datasets into actionable forecasts.
- **RAG System Architecture:** Designed and deployed a **Retrieval-Augmented Generation (RAG)** pipeline using **PostgreSQL** and **pgvector**, enabling AI models to perform high-speed semantic search and reduce hallucinations.
- **Vector Database Implementation:** Built robust embedding pipelines to vectorise domain-specific knowledge bases, facilitating precise information retrieval for automated agents.
- **AI Agent Automation:** Developed autonomous AI agents capable of parsing natural language intent to execute real-time tools, including **Google Calendar** booking and **CRM** data entry.
- **Prompt Engineering & Context Logic:** Implemented advanced backend logic in **TypeScript** to manage conversation state, inject dynamic variables (e.g., weather data), and optimize system prompts for highly personalized user experiences.

AI Technical Stack: OpenAI API, Anthropic Claude API, RAG, Vector Embeddings, PostgreSQL (pgvector), Natural Language Processing (NLP), Telegram API, TypeScript.

Software Developer

 **Digy4** | Jul 2021 – Oct 2022

- Developed a **CRM application** using **Node.js, Express, MongoDB, and EJS**.
- Implemented **TestNG test cases, Selenium WebDriver scripts**, and **retry logic** for test automation.
- Set up a **User Account Module** using **Next.js, TypeScript, and Chakra-UI**.
- Established **development environments** with **WordPress, XAMPP, and FileZilla**.

Frontend Web Developer

 **Evolve Branding** | Apr 2020 – Jun 2021

- Developed the **front-end** using **ReactJS, Bootstrap, and Tailwind CSS**.
- Integrated **Eventbrite API** for event listings.
- Assisted in **backend development** with **MongoDB and Express**.

Education

 **Dalhousie University**

Bachelor's in Computer Engineering | Sep 2015 – May 2019

Skills

AI & Machine Learning: Generative AI (OpenAI/Claude APIs), RAG Architecture, Vector Databases (pgvector), NLP, AI Agent Automation.

Languages & Frameworks: JavaScript (ES6+), TypeScript, C++, C#, React, Next.js, Node.js, Express.js, ASP.NET Core.

Cloud & Infrastructure: AWS Lambda, Docker, Microservices, RESTful APIs, Git.

Databases: PostgreSQL, MySQL, MongoDB.