PRIYANSH SAHU

bepriyansh.vercel.app

priyanshsahu.2002@gmail.com
linkedin.com/in/bepriyansh
+91 7668747527

SUMMARY

AI Engineer with experience designing and deploying RAG-based systems, multi-agent workflows, and production-grade AI applications. Skilled in FastAPI, LangGraph, MCP, microservices, REST APIs, and distributed systems, with full-stack experience in React and Next.js. Strong foundation in data structures, algorithms, system design, and scalable backend development.

WORK EXPERIENCE

AI Engineer 2, Bayer
Dec 2025 - Present

- Designed RAG pipelines, improving search relevance accuracy by 40%, resulting in more accurate and context-aware AI responses.
- Developed Generative UI systems using LangGraph to transform agentic outputs into dynamic, real-time frontend components.
- Implemented a concurrency-safe API caching layer, eliminating redundant queries and preventing 'thundering herd' traffic spikes.
- Developed production-grade FastAPI backends and MCP servers, leveraging LangGraph to orchestrate multi-step agent workflows.
- Used LiveKit for real-time, low-latency communication within agentic applications using WebRTC-based audio-video streaming.

Senior Software Engineer 2, HCLTech
Jul 2025 - Nov 2025

- Designed and implemented a multi-agent orchestration system using LangGraph to coordinate specialized AI agents.
- Built and deployed MCP servers for internal applications, standardizing agent communication and tool access across systems.
- Integrated Microsoft SSO using OAuth 2.0 OBO (On-Behalf-Of) flows to enable secure, delegated enterprise authentication.
- Developed FastAPI-based backend services to support agent orchestration, system integrations, and AI-driven workflows.
- Worked on improving reliability and maintainability of backend services through cleaner abstractions and better service structure.

Academic Trainee, HCLTech
May 2025 - Jul 2025

- Gained hands-on experience with microservice architecture and distributed systems, working with independently deployable services.
- Built an AI-native social media platform using a modular microservice architecture (API gateway, auth, post, notification services) with event-driven components for asynchronous workflows.
- Implemented messaging and background processing using queues/events to improve scalability and decouple core services.
- Developed a GitHub PR reviewer agent that analyzed pull requests for code quality, best practices, and potential issues.
- Strengthened foundations in backend development, system design, and AI-assisted tooling through end-to-end project ownership.

EDUCATION

Indian Institute of Technology, (BHU), Varanasi
Dec 2021 - May 2025

Bachelor of Technology in Ceramics Engineering

- Applied supervised ML techniques to estimate band gaps of mixed compounds by learning from known data of individual materials.
- Studied how adding different rare-earth dopants changes ionic conductivity in zirconia using machine learning models.

PROJECTS

On-Device Mobile AI
April 2026

- Built a secure, privacy-first, on-device AI chat application that safely runs powerful language models completely offline.
- Developed intuitive user controls allowing for complete customization of AI response behavior and hardware resource usage.
- Engineered a highly robust background system to reliably download and seamlessly manage massive gigabyte-scale AI models.
- Designed a real-time event system to provide precise download tracking and instant asynchronous AI inference alerts.
- Orchestrated application architecture to ensure perfectly smooth, predictable user transitions across all asynchronous AI workflows.
- Technologies: Flutter, Dart, SQLite, Llama inference Engine, Background Fetch, State Management (Bloc), REST APIs.

Legal AI Agent
Feb 2026

- Built an automated AI legal assistant that safely extracts contract data and mathematically eliminates all AI hallucinations.
- Engineered a multi-agent system using Knowledge Graphs to securely extract and accurately validate complex legal documents.
- Deployed advanced parsing tools to reliably preserve exact document structures, embedded tables, and critical contract metadata.
- Designed an auditable UI powered by Knowledge Graphs, dynamically linking insights directly to exact document clauses.
- Reduced tedious document review cycles from hours to minutes by fully automating data extraction and verification workflows.
- Technologies: Python, LangGraph, Pydantic, Knowledge Graphs, Neo4j, Next.js, Express, Custom Document Parsers, Docker.

AI Analyst
Nov 2025

- Built an interactive AI data analyst that translates natural language directly into secure, executable SQL queries.
- Engineered a strict rule-based pipeline to validate generated SQL syntax against database schemas before safe execution.
- Integrated dynamic frontend components to automatically render database query results into intuitive charts and data tables.
- Optimized response latency by implementing semantic caching to instantly return results for previously asked business questions.
- Eliminated technical bottlenecks, enabling stakeholders to independently extract critical business intelligence in real time.
- Technologies: Python, React, Node.js, Express, PostgreSQL, Redis, LangChain, Docker, RESTful APIs, Git, CI/CD Pipelines.

Skills & Achievements

- Solved 800+ data structures and algorithms problems on LeetCode, GeeksforGeeks, and Coding Ninjas.
- Led public relations and event operations as PR Executive at Technex'23, coordinating speakers, technical talks, exhibitions.
- Languages & Core: Python, TypeScript, JavaScript, SQL, C++
- AI & Backend: LangGraph, RAG, MCP, FastAPI, Node.js, Microservices, REST
- Frontend & Tools: React, Next.js, Docker, Git, PostgreSQL, MongoDB, Linux