

KASINATHAN B

AI Software Engineer | LLM, RAG & AI Agents | Speech AI | Full Stack (React / Angular / Node.js / .NET)

Kerala, India | +91 62358 37610 | way2kasinathb35@gmail.com

linkedin.com/in/kasinathanb | github.com/kasinathan28

PROFESSIONAL SUMMARY

AI Software Engineer with 2+ years of experience building production-grade AI applications and scalable full-stack systems. Shipped a real-time clinical documentation product: a browser extension that transcribes doctor-patient consultations, generates structured SOAP notes with an LLM, and writes them directly into the EMR. Specialized in integrating Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), embeddings-based semantic search, AI agents, and speech-to-text pipelines into React, Node.js, and .NET products. Delivers end to end: frontend interfaces, REST APIs and backend services, database architecture, and cloud deployment on AWS. Applications built and shipped serve approximately 4,000 to 20,000 monthly active users across healthcare, sales automation, and Web3 domains.

TECHNICAL SKILLS

AI and LLM: OpenAI API, Google Gemini, Prompt Engineering, Retrieval-Augmented Generation (RAG), Embeddings, Semantic Search, Vector Databases, AI Agents, Multi-Agent Workflows, Function Calling, Structured JSON Output, Context Engineering, Conversational AI, AI Workflow Automation

Speech AI: Whisper API, Speech-to-Text, Real-Time Audio Transcription, Streaming Audio Capture, Voice AI Integration

Languages: JavaScript, TypeScript, Python, C#, C, C++

Frontend: React.js, Next.js, Flutter, Browser Extension Development (Chrome Extension APIs), HTML5, CSS3, Tailwind CSS, Bootstrap, Responsive UI Development

Backend: Node.js, Express.js, .NET, REST API Design, JWT Authentication, OAuth 2.0, Third-Party API Integration

Testing and Automation: Browser Automation Testing, End-to-End (E2E) Test Automation, .NET Automation Frameworks, AI-Assisted Test Generation

Databases: MongoDB, MySQL, Firebase, Redis, Vector Databases

Cloud and DevOps: AWS EC2, AWS S3, Docker, Vercel, Firebase Hosting, Git, GitHub, Production Deployment

Integrations: WhatsApp Business API, Meta Graph API (Facebook, Instagram), Google Ads API, Google APIs, Twilio, Firebase Authentication, EmailJS, Moralis, Ethereum

Healthcare Domain: Electronic Medical Records (EMR), Clinical Documentation, SOAP Note Generation, Ambient Clinical Documentation

Practices: System Design, Scalable API Architecture, Performance Optimization, LLM Cost Optimization, Agile Delivery, Code Review

PROFESSIONAL EXPERIENCE

Software Engineer

AOT Technologies | June 2024 - Present

- Built and shipped a browser extension for ambient clinical documentation that captures doctor-patient consultations in real time, transcribes the conversation via speech-to-text, generates structured SOAP notes with an LLM, and writes them directly into the EMR, removing manual note-taking from the clinician workflow.
- Engineered the transcription-to-note pipeline end to end: real-time audio capture and streaming, speech-to-text integration, and prompt design that maps unstructured clinical dialogue into the Subjective, Objective, Assessment, and Plan sections clinicians expect.
- Building a .NET browser automation testing platform with a zero-token architecture that executes repetitive test steps deterministically without LLM calls, reserving model inference for novel scenarios and cutting per-run inference cost.
- Develop enterprise-grade healthcare web applications, including Electronic Medical Record (EMR) modules used in day-to-day clinical workflows, using React and REST APIs.
- Build responsive, reusable component libraries that standardize the UI across product modules and eliminate duplicated frontend code.

- Integrate secure payment gateways into patient-facing flows, handling transaction state, failure recovery, and compliance-sensitive data.
- Profile and optimize frontend performance across render cycles, bundle size, and API call patterns; partner with backend engineers on REST API contracts to reduce integration defects during feature delivery.
- Contribute to feature planning, peer code review, and production release cycles within an Agile team.

Freelance AI and Full Stack Engineer

Self-Employed | Remote | September 2024 - Present

- Design and ship custom AI-powered products for startups and small businesses, owning the full lifecycle from requirements and system design through frontend, backend, database architecture, and cloud deployment.
- Build RAG pipelines using embeddings and vector search to ground LLM responses in client-specific knowledge bases, improving answer relevance and reducing unsupported model output.
- Implement multi-agent AI workflows with function calling and structured JSON output to automate document generation, lead handling, and repetitive internal business processes.
- Develop speech-to-text and audio transcription pipelines using Whisper-class APIs to power real-time voice-driven product features.
- Deploy and operate production services on AWS EC2 and S3 with Redis caching and MongoDB data modeling, delivering applications used by approximately 4,000 to 20,000 monthly active users.
- Optimize prompt design, context construction, and token usage to improve response quality and control per-request LLM cost.

Software Engineering Intern

Techfriar | June 2024 - December 2024

- Built and shipped MERN stack (MongoDB, Express.js, React, Node.js) application features from requirement through production release.
- Developed reusable React components and integrated backend REST APIs across multiple product modules.
- Diagnosed and resolved production defects, improving application stability and frontend rendering performance.

KEY PROJECTS

AI Sales Pilot | Google Gemini, Node.js, React, MongoDB, Redis, AWS, Meta APIs

- Built an AI sales automation platform that centralizes inbound leads from WhatsApp, Instagram, Facebook, and Google Ads into a single CRM pipeline.
- Engineered LLM-based lead qualification and conversation summarization with Google Gemini, automatically scoring leads and condensing message history for sales representatives.
- Implemented semantic search across lead and conversation data using embeddings and vector similarity, enabling natural-language retrieval over the CRM.
- Designed a multi-agent workflow architecture that routes enrichment, qualification, and follow-up drafting to specialized AI agents.
- Deployed to AWS with Redis caching and MongoDB persistence, integrating Meta Graph APIs for real-time multi-channel message ingestion.

AI-Enabled Web3 Form Registry | Node.js, MongoDB, Ethereum (Sepolia), Moralis

- Developed a blockchain-backed document verification platform that anchors cryptographic form hashes to the Ethereum Sepolia network for tamper-evident records.
- Integrated Moralis APIs for on-chain read and write operations, backed by a MongoDB off-chain store linked to each on-chain proof.
- Implemented an end-to-end verification workflow allowing third parties to validate document authenticity against the immutable on-chain hash.

Premium Interior Design Platform | Next.js, Responsive UI, SEO

- Designed and delivered a high-performance marketing website with modern responsive UI, on-page SEO optimization, and a self-serve blog management system for the client team.

Story-Based AI Dating Platform (Concept) | Flutter, Cross-Platform Mobile

- Architected a cross-platform Flutter application centered on narrative "storybook" profiles with AI-assisted compatibility matching in place of attribute-based filtering.

EDUCATION

Diploma in Computer Engineering

Carmel Polytechnic College | 2021 - 2024 | CGPA: 8.32 / 10

THSLC (Higher Secondary)

Govt. THS Haripad | 2020 - 2021 | Score: 98%