

Akash Kumar

Ranchi, Jharkhand • +91-9431553235 • aakashkkr04@gmail.com

Portfolio: akashkkr04.github.io • GitHub: github.com/theakash04

PROFESSIONAL SUMMARY

Full-stack developer focused on building scalable, real-time applications and AI-powered tools. Experienced in end-to-end product development with a strong emphasis on performance and usability.

PROJECTS

Recall

- **Overview:** Built a fast, user-centric smart bookmarking system enabling users to save and search web content using vague memory cues, solving the common problem of lost or hard-to-recall bookmarks.
- **Key Features:**
 - Developed a seamless one-click URL saving feature with automatic background scraping and **rescraping** support.
 - Implemented advanced hybrid search (keyword + semantic) using custom-built vector search with **pgvector** and PostgreSQL, optimizing for both **speed** and **accuracy**.
 - Designed a deduplicated relational DB schema (GlobalBookmarks, UserBookmarks, BookmarkContent) to **avoid redundant scraping, embedding, and storage, reducing resource and AI usage costs**.
 - Reused embeddings through content hashing and chunk matching to ensure storage efficiency and **faster search performance**.
 - Secured user data with Supabase Auth and backend-only access logic; ensured **privacy** even with public anon keys.
 - Delivered a minimalist, intuitive UI/UX with Google OAuth, customizable search settings, and a fast, frictionless user experience.
- **Tech Stack:** Next.js, Supabase Auth, PostgreSQL, pgvector, Express.js, Bull-Mq, Redis (upstash), Google AI, Docker

Stuttering Detection

- **Overview:** Built an AI-powered application that detects stuttering in real-time speech, aimed at helping users identify and improve their speech fluency.
- **Key Features:**
 - Fine-tuned OpenAI's Whisper (Tiny) model using PyTorch and a stuttering-labeled dataset to classify speech as stuttered or fluent with ~92% training accuracy
 - Engineered a FastAPI backend to process and analyze audio data (WAV format) from web and mobile clients in real-time
 - Designed a responsive Next.js frontend for recording and submitting speech with minimal user friction
 - Integrated Gemini API to generate personalized, AI-driven feedback for users, focusing on non-pathological stuttering causes like confidence or anxiety
 - Delivered a seamless end-to-end experience with accurate detection, fast response time, and high usability across platforms
- **Tech Stack:** FastAPI, Next.js, Whisper (Tiny), PyTorch, Gemini API

Termify

- **Overview:** Developed a real-time chatbot that simplifies complex legal documents like terms and privacy policies using LLMs and custom retrieval.
- **Key Features:**
 - Designed a RAG architecture with Cortex Search, Snowflake, and Mistral (mistral-large2) to generate context-aware, human-readable summaries of legal text
 - Built a personalized search pipeline allowing users to upload documents, stored in isolated schemas, enabling secure and custom querying
 - Evaluated response accuracy, context quality, and coherence using Trulens across four system versions, driving measurable performance gains
 - Delivered an intuitive Streamlit-based UI and containerized deployment with Docker for rapid scalability and testing
- **Tech Stack:** Mistral LLM, Cortex Search, Snowflake, Trulens, Streamlit, Docker

Technical Skills

- **Languages:** Typescript, Python, Rust
- **Frameworks & Libraries:** React.js, Next.js, Tailwind CSS
- **Backend:** FastAPI, Nodejs
- **Databases:** PostgreSQL, mongoDB (NoSQL), Redis
- **DevOps:** Docker
- **Others:** Git, Jest

Additional Skills: Remote teamwork, problem-solving, Time management, Adaptability and quick learning

Education

- **Bachelor of Computer Applications (BCA)**
Birla Institute of Technology, (Lalpur Extension) Ranchi, Jharkhand
Expected Graduation: 2026
- **Senior Secondary (Class 11-12)**
St. John's Inter college, Ranchi - JAC Board
Year of Completion: 2023