Nikhil Mahajan

Bangalore, India nikm5502@gmail.com 7310992130 bio.nik-55.dedyn.io
nikhilmahajan123 nik-55

Skills

- Languages: Python, TypeScript, C, C++
- Software Development: React, Next.js, Django, FastAPI, PostgreSQL, TimescaleDB, MongoDB, Redis, Celery, Selenium
- GenAI: LLM Fine-tuning, LangGraph, Gemini APIs, Video Generation (VAEs, Latent Diffusion Transformers)
- DevOps: AWS (EC2, ECS, ECR, S3, Lambda, SQS, WAF, Load Balancer, Cloudwatch), Terraform, Docker, Git, Linux, Nginx

Experience

Software Engineer (Full Time) *Bangalore*
Astrikos.AI *July 2025 – Present*

- Engineered domain-specific conversational AI chatbots by fine-tuning open-source LLMs and developing backend infrastructures for diverse enterprise clients.
- Deployed models across multi-GPU environments, implementing model partitioning techniques to overcome single-node memory constraints.
- Developed centralized Speech-to-Text (STT) and Text-to-Speech (TTS) backend services, standardizing audio processing and integrating it seamlessly across multiple company products.
- Architected an automated data ingestion pipeline to periodically extract high-throughput sensor data from industrial historians.
- Optimized TimescaleDB configurations and chunking intervals, significantly accelerating the performance of complex time-series analytics queries.
- Managed cloud and network infrastructure, including AWS provisioning, Nginx reverse proxy configuration, and domain management.
- Packaged custom ISO images to facilitate secure, standardized, and offline on-premise software deployments.

Software Engineer (Intern) *Remote*
Astrikos.AI *April 2024 – June 2025*

- Designed and developed a flow engine to enable dynamic execution of Python code through a node-based system.
- Built a dynamic console for managing global datacenter operations, improving visibility and control.
- Worked with real-time telemetry data to support data-driven infrastructure management.
- Collaborated across multiple teams to coordinate deployments and ensure smooth delivery of features.
- Utilized ReactJS for the frontend and Django, Celery, PostgreSQL, and TimescaleDB on the backend.

Software Engineer (Intern) *Remote*
Recepto.AI *Nov 2024 – March 2025*

- Designed and implemented a workflow to regularly retrieve the latest and most relevant leads from the internet.
- Developed custom web scraping tools to extract leads from multiple online sources, including a specialized LinkedIn scraper for profiles, job listings, and posts.
- Integrated all tools into the lead generation system.
- Contributed to both backend (FastAPI, MongoDB) and frontend (React) development during the product's early stages.
- Actively participated in product discussions, proposed new features, and collaborated closely with the founders, gaining hands-on experience in a fast-paced startup environment.

Software Engineer (Intern) *Remote*
Repello.AI *March 2024 – May 2024*

- Designed and developed the user interface for an automated red teaming platform, based on a thorough understanding of product needs.

- Built the platform with features including real-time log streaming and PDF report generation.
- Conducted red teaming on emerging AI products to identify and report security vulnerabilities.
- Implemented frontend components using React and backend services with Flask.
- Contributed to the improvement of the automated red teaming engine.

Open Source Contributor (Google Summer of Code) *Remote*
Wikimedia Foundation *May 2023 – Aug 2023*

- Developed end-to-end (E2E) tests for various user journeys to improve application reliability.
- Contributed to a large-scale codebase by fixing minor bugs in Wikipedia.
- Gained practical experience working within a global, distributed open-source development team.
- Utilized WebdriverIO for browser automation to streamline testing workflows.

Full Stack Developer (Freelance) *IIT Roorkee*
Self-Employed *Nov 2022 – Dec 2022*

- Developed a web-based platform for simulating stock trading.
- Integrated third-party APIs to fetch and display real-time stock market trends and data.
- Built a responsive frontend using React.js and Bootstrap, and implemented backend using Django.

Projects

Redis From Scratch

- Built a lightweight, in-memory key-value database from scratch in C, replicating core Redis functionality and underlying data structures.
- Engineered a concurrent TCP server using standard C socket programming to handle multiple client connections simultaneously.
- Parsed the REdis Serialization Protocol (RESP) to decode raw network byte streams into executable database commands.
- Developed support for essential commands (PING, ECHO, SET, GET) and implemented time-to-live (TTL) logic for automatic key expiration.
- Implemented master-replica synchronization, enabling replica nodes to connect, perform handshakes, and reliably synchronize state with the master server.

Variational Autoencoder - QuickDraw Dataset

- Trained a custom Variational Autoencoder (VAE) from scratch on the Google QuickDraw dataset to autonomously generate and interpolate hand-drawn sketches.
- Designed an advanced encoder-decoder architecture incorporating ResNet-style Residual Blocks (with SiLU activations) and Scaled Dot-Product Attention to capture long-range spatial dependencies.
- Implemented the reparameterization trick and a composite loss function, specifically utilizing KL Annealing to successfully prevent posterior collapse during training.
- Stabilized mixed-precision training by implementing gradient clipping to resolve gradient explosion issues.
- Engineered custom inference methods, utilizing Aggregate Posterior sampling rather than standard normal priors to significantly improve output quality and prevent blank generations.

Bhagavad Gita Conversational AI

- Curated a custom, domain-specific dataset on the Bhagavad Gita from scratch by scraping websites, running OCR on scanned books, and transcribing YouTube audio.
- Leveraged the Gemini API to process raw, unstructured text and programmatically generate synthetic question-answer pairs, yielding a high-quality dataset for LLM instruction tuning.
- Fine-tuned the Google Gemma-3-1B-IT model using Hugging Face libraries to create an AI capable of answering questions based on the text.
- Managed the training pipeline on rented GPUs from runpod.io and published the final model weights to the Hugging Face.

Shopmate

- Engineered an AI-powered personal shopping assistant using a multi-step LangGraph agent to automate product discovery, autonomously comparing items and recommending the best options to reduce search time.

- Created an intent understanding module that analyzes user requests, conducts preliminary searches on product categories, and dynamically asks context-aware clarifying questions before initiating a deep search.
- Built scraping pipeline to extract and structure product details, pricing, and user reviews directly from Amazon, Flipkart, and Google Search.
- Designed a recommendation system that scores products based on user preferences and shows the top three matches for easy decisions.

Election Portal

- Developed and maintained the official platform for conducting campus-wide elections at IIT Roorkee, enabling over 10K voters to elect Positions of Responsibility (PORs).
- Implemented robust security measures to ensure secure and tamper-proof voting processes.
- Wrote clean, extensible, and production-grade code across frontend and backend.
- Built using Next.js, Django, PostgreSQL, Strapi, and Docker.
- Managed scalable infrastructure with AWS and Terraform, ensuring high availability and fault tolerance.

IKS Console

- Developed for the Indian Knowledge Systems (IKS) Division, Ministry of Education, Government of India
- Built a dynamic certificate generation platform that allows users to upload templates via a web console and generate personalized certificates from excel data.
- Automated the emailing of generated certificates directly to participants.
- Utilized Django, React.js, PostgreSQL, Docker, Pillow, and Nginx.
- Managed infrastructure including AWS SQS, Lambda, and S3 for scalable and efficient processing.

Education

Indian Institute of Technology Roorkee *Nov 2021 – April 2025*
B.Tech in Chemical Engineering

Extracurricular Activities

Project Manager *IIT Roorkee*
MDG Space (Technical Club) *April 2022 – April 2025*

- Led multiple technical projects end-to-end from ideation to deployment.
- Organized and conducted workshops, mentorship programs, and tech events attended by over 100 students. Delivered technical lectures on various tech stacks including frontend and backend development, and cloud. Mentored juniors for hackathons, internships, and technical problem-solving.
- Fostered a collaborative and growth-oriented team culture.
- Contributed to strategic discussions on integrating emerging technologies into club initiatives.

Participant *Remote*
Google Gen AI Hackathon *Sep 2024 – Oct 2024*

- Developed a GenAI-powered application to support cancer diagnosis and treatment by analyzing MRI scans, identifying and segmenting tumors, estimating cancer probability, and generating detailed diagnostic reports with personalized treatment plans. Included a GenAI chatbot for real-time access to up-to-date research and treatment insights.
- Collaborated in a 3-member team; placed in the top 5 in the healthcare track (India).
- Invited to Google's Gurgaon office; networked with VCs and participants, and received swag and recognition.

Finalist (Top 100 / 2000+) *Bangalore*
OpenEnv India 2026 — Hugging Face, Meta PyTorch, Scaler *April 2026 – May 2026*

- Built MedChain, an RL environment where an LLM agent coordinates hospital supply chains across multiple wards and enterprise systems, with a 10-component deterministic reward formula and no LLM-judged rewards.
- Implemented GRPO training directly against the live simulation using a custom multi-turn rollout, fine-tuning Qwen3.5-2B with context resets between rounds to keep sequences tractable.
- Solo submission; advanced to the finals round and placed in the top 100 among 2000+ participants.