Nirajan Karki

Senior Backend & AI Engineer · Python

Kathmandu, Nepal · +977-9844395719 · kneeraazon@gmail.com
kneeraazon.com · github.com/karkinirajan · linkedin.com/in/kneeraazon

INTRODUCTION

Senior Backend and AI Engineer with **7+** years building production APIs, retrieval systems, and full-stack products for global clients. Top Rated on Upwork with a **100%** Job Success Score and **80+** delivered projects for **75+** clients. Specializes in Python, FastAPI, Django, RAG, and cloud infrastructure, with results including **41%** lower LLM inference cost, **72%** lower API latency, and **35%** lower AWS spend.

7+	80+	75+	100%
YEARS IN PRODUCTION	PROJECTS DELIVERED	GLOBAL CLIENTS	UPWORK JOB SUCCESS

SELECTED PROJECTS

RevOps Lead Scoring Platform

Built a B2B lead-scoring platform combining a live transparent formula with a shadow XGBoost model, Celery workers, a score simulator, and signed HubSpot and Salesforce webhooks.

Python · FastAPI · Celery · XGBoost · PostgreSQL · Next.js 16

github.com/karkinirajan/revops-lead-scoring-platform

Pilot — AI Operations Console

Built the Next.js console for an AI assistant triaging customer email, Slack, and calendar events as ALERT, DRAFT, or NOTHING, with a live feed, trace drawer, and Claude chat.

Next.js 16 · TypeScript · PostgreSQL · Claude · Zep Cloud

pilotready.vercel.app

EkJhalak — Bilingual News Briefings

Built and operate a bilingual news reader ingesting public RSS feeds, deduplicating coverage, and generating labelled Nepali and English summaries and translations.

Next.js 16 · TypeScript · RSS · Groq API · Supabase

ekjhalak.news

Delta Hedging Platform

Built an options risk platform on IG's demo API that prices positions with Black–Scholes, tracks portfolio delta, recommends hedges, and places only confirmed orders.

Python · Flask · IG API · Black–Scholes

github.com/karkinirajan/delta-hedging-platform

SmartClaim AI — Medical Billing Compliance

Built a wound-care compliance app evaluating **46** Medicare LCD L35041 criteria across a 6-step checklist with structured OpenAI output.

Next.js 16 · TypeScript · OpenAI · Neon Postgres

smartclaimai.vercel.app

Nestpane — Local-First Productivity Dashboard

Shipped a local-first Chrome new-tab workspace for Markdown notes, tasks, bookmarks, focus timer, and habits, with encrypted Drive sync.

Chrome Manifest V3 · JavaScript · Web Crypto

nestpane.kneeraazon.com

Spinoza's Ethics — Interactive Visualisation

Interactive visual guide to Spinoza's Ethics with 6 SVG diagrams and a Groq-powered Geometric Advisor.

Next.js 16 · TypeScript · SVG · Groq API

ethicsbyspinoza.vercel.app

Web Platforms — TexecN, CBT Coach & Sheesh Shoes

Built an AI Kanban task board, a FastAPI + Next.js CBT recovery coach, and a 3D athletic footwear store.

www.toexecutenow.com · therapist.kneeraazon.com · sheeshshoes.com · kneeraazon.com

Additional work: Clarity Enterprise, Everest Smart Traders, deterministic quantitative routing, automated trading arena, FastKwan TCM RAG API, AI recruitment CLI, VoiceTech radar, and budget tracker. Full catalogue: github.com/karkinirajan

WORK EXPERIENCE

Senior Backend Python and AI Engineer
Nov 2021 – Present

Freelance via Upwork (Top Rated, 100% JSS)

- Architected production RAG and LLM systems for chat, semantic search, and agentic workflows using OpenAI, Claude, Gemini, LangChain, and LlamaIndex.
- Cut LLM inference cost **41%** through embedding caches and semantic deduplication; reduced API latency **72%** with async FastAPI, Redis, batching, and query tuning.
- Scaled a Django platform to **12,000** requests per minute at **99.99%** uptime through a zero-downtime microservices migration.
- Built a containerized ingestion pipeline for **2M+** events per day at **37ms** average latency; reduced AWS spend **35%** through infrastructure and RDS optimization.
- Automated CI/CD with GitHub Actions, reducing release cycles from 2 days to 2 hours; delivered **80+** client projects end to end.

Full-Stack Developer
Aug 2020 – Sep 2022

Pathway Technologies and Services

- Built Django/DRF platforms and APIs for React and Next.js clients; improved MySQL and MongoDB models and query performance under production load.

Web Developer
Jan 2019 – May 2021

Merohealth

- Shipped Django, DRF, React, and Next.js features for a healthcare platform, including components used by a **10,000+** user dashboard.
- Improved relational and document data models to support reliable, responsive healthcare workflows.

Web Developer
Feb 2018 – Feb 2019

Food and Beverage Nepal

- Solo-built a Django product and catalogue platform, delivering responsive UI, clean APIs, and improvements to performance, accessibility, and SEO.

SKILLS

AI & Retrieval: RAG, hybrid retrieval, reranking · LLM agents and tool orchestration · OpenAI, Claude, Gemini · LangChain, LlamaIndex- · pgvector, Qdrant, ColBERT · RAGAS, DeepEval, Langfuse

Backend: Python, FastAPI, Django, DRF, Flask · AsyncIO, Celery, WebSockets · Pydantic, REST API design · Caching, idempotency, retries

Frontend & Data: React, Next.js, TypeScript · PostgreSQL, MongoDB, Redis · Kafka, Elasticsearch, Supabase

Cloud & Delivery: AWS, GCP, Azure · Docker, Kubernetes, Linux · GitHub Actions, Prometheus, Grafana

EDUCATION

Bachelor of Engineering, Electronics and Communications
Tribhuvan University (IOE), Pokhara · 2016 – 2021