



NIKHIL BHAGAT

programmer@nikhilbhagat.com.np | ☎ +977-9705651002 |

[Github](#) | [Linkedin](#) | [Portfolio](#)

Address: Hathkhola, Biratnagar, Nepal

Objective

Full Stack Engineer with 8+ years of experience building scalable APIs, real time systems, and AI powered products. Specialized in Python based backends, cloud native architectures, and integrating Generative AI using LLMs like OpenAI and LangChain. Passionate about system design, performance optimization, and building intelligent, automation first backend solutions.

Technical Skills

- ✓ Languages: Python, JavaScript (ES6+), SQL
- ✓ Backend Frameworks: Django, Django REST Framework, FastAPI, Node.js (Express.js)
- ✓ Frontend: React.js, Next.js, Redux, TailwindCSS
- ✓ AI / GenAI: OpenAI API, LangChain, LlamaIndex, Vector DBs (FAISS, Pinecone), Transformers
- ✓ Databases: PostgreSQL (with PostGIS), MongoDB, Redis
- ✓ Cloud & DevOps: Docker, Kubernetes, AWS (Lambda, EC2, S3), GitHub Actions, Jenkins, Terraform, Ansible
- ✓ Architecture: RESTful APIs, Microservices, Event-Driven Design, Kafka, Webhooks, Background Workers (Celery + Redis)
- ✓ Realtime & Notifications: Firebase, WebSockets, Push Notifications
- ✓ Testing & Observability: Pytest, Postman, Django Debug Toolbar, Logging & Metrics

Projects

1. Rani (Code Debugger AI Tool)

Jan 2024 – Mar 2024

- GitHub: [Link](#)
- Built an AI chatbot using OpenAI API and LangChain to analyze and explain large codebases, reducing debugging time by 30%.
- Integrated prompt templates, error tracing, and code to natural language explanations.

- Solved token limit bugs in large file inputs and optimized context compression logic.
- Stack: Python, LangChain, OpenAI, Django REST Framework

2. Tung Tung (Location Based Game Full Backend mechanics API's)

Aug 2024 – June 2025 (Client Project: Japan)

- Github: [Link](#)
- Built a real time backend for a GPS based multiplayer game where users can claim zones, engage in strategic attacks, and take ownership of rival territories.
- Implemented geofencing, real time push notifications, and live synchronization to deliver dynamic gameplay and territory battles across an interactive map.
- Integrated secure user authentication, scalable data handling, and an official leaderboard system tracking the top 100 players for competitive ranking.
- Built with Django, Django REST Framework, and GeoDjango, using PostgreSQL + PostGIS for spatial data, JWT authentication, Celery + Redis for background tasks, Firebase for push notifications, and PostGIS spatial queries to power real time zone battles, leaderboards, and geolocation based gameplay.
- Stack: Django, DRF, GeoDjango, PostgreSQL + PostGIS, Redis, Firebase

3. Neural Network Visualizer Tool

Jul 2018 – Dec 2018

- Live Demo: [Link](#)
- Created an interactive React.js visualization tool to demonstrate how neural networks propagate data.
- Solved rendering lag issues by implementing virtualized D3.js nodes for large networks.
- Adopted by 200+ learners for educational purposes.
- Stack: React.js, D3.js, TensorFlow

4. Wahi AI (Nepal Financial Reporting Standards (NFRS)) [Full Backend Api's]

Mar 2025 – Oct 2025 (Worked for my Client : Nepal)

- Github: [Link](#)
- Developed an intelligent document driven conversational AI system to assist users with Nepal Financial Reporting Standards (NFRS), delivering accurate, context aware guidance in both English and Nepali.

- Implemented secure JWT authentication, multilingual support, semantic document processing, and vector based retrieval using FAISS for precise, real-time financial query responses.
- Built on Django REST Framework, FAISS for semantic search, JWT for authentication, and advanced NLP models to power AI driven conversational experiences.
- Stack: Django REST Framework, FAISS, LangChain, JWT, PostgreSQL

5. Microservice Generator Tool (Turbogen)

Mar 2024 – Jan 2025

- npm Package: [Install it](#) (`npm i turbogen`)
- Automated Node.js microservice scaffolding, cutting setup time by 50%.
- Fixed recursive dependency installation and improved CLI UX using Commander.js + Chalk.
- Stack: Node.js, CLI Tools

6. DevLinker : Social Dev Network (React + Node Full Stack App)

Feb 2020 – June 2022 (Worked for my Italy Client)

- Created a full stack platform for developers to share portfolios and collaborate.
- Built React frontend with JWT authentication, Redux state management, and Axios APIs.
- Designed REST APIs in Express.js, handled image uploads via Cloudinary, and integrated CI/CD with GitHub Actions.
- Fixed CORS errors, API throttling issues, and improved build performance by 40%.
- Stack: React.js, Node.js (Express), MongoDB, JWT, Cloudinary, GitHub Actions

7. Battisputali Backend (Community App Just like a TikTok)

➤ **Sep 2022 – Oct 2023 (Client Project: Hong Kong)**

- GitHub: [Link](#)
- Developed a scalable backend for a 10k+ user community app supporting real-time posts, likes, and comments.
- Solved Cloudinary upload timeouts and optimized video caching via stream chunking.
- Built WebSocket based notification and MongoDB aggregation pipeline for analytics.
- Stack: Node.js, Express.js, MongoDB, Cloudinary, WebSocket

8. TaskPilot (Real Time Task Management App)

➤ **Feb 2025 – Aug 2025 (Client Project: Philippines)**

- Built a Trello style Kanban board with real time socket updates for task assignment and progress tracking.
- Integrated drag and drop (React Beautiful DnD), role based auth, and activity logs.
- Resolved WebSocket disconnect issues by implementing Socket.IO heartbeat checks.
- Stack: React.js, Node.js, Express.js, Socket.IO, PostgreSQL

9. Hauba – Transformer based LLM (Custom AI Model)

Jul 2024 – Present (Private Github Repo)

- Designed and built Hauba, a custom Transformer based Large Language Model (LLM) fine tuned for code understanding, generation, and debugging assistance.
- Pretrained on a curated dataset of 50K+ code files across Python, JavaScript, and TypeScript to improve syntax understanding and logical reasoning.
- Implemented multi head self attention, positional encoding, layer normalization, and causal masking from scratch using PyTorch.
- Integrated function summarization, error explanation, and unit test generation capabilities for developers.
- Applied LoRA fine tuning for efficiency and quantization (int8) for deployment on limited GPU resources.
- Served real time inference using FastAPI microservices, integrated into LangChain pipelines for retrieval augmented coding assistance.
- Debugged gradient explosion and overfitting issues using gradient clipping and adaptive learning rate scheduling.
- Stack: PyTorch, Transformers, FastAPI, LangChain, CUDA, Tokenizers, LoRA

(Additional projects: [Portfolio](#) | [Github](#))

Educational Qualification

Bachelor of Computer Application (BCA)

Mahendra Morang Aadarsha Multiple Campus | 2025

Secondary School Certificate

Merryland College | 2013

Experience

CineART • Hong Kong • 06/2022 – Present [Project Based Contract]

Senior Backend Developer

Lead backend development for a LinkedIn style platform for Nepal, building scalable APIs using Django REST Framework.

- Designed and deployed microservices with Docker and Heroku; planning AWS migration.
- Implemented engagement based algorithm to rank user posts and boost content visibility.
- Set up CI/CD with GitHub Actions and integrated third party services.
- Implemented secure auth systems (JWT), optimized performance with Redis, and improved server efficiency.
- Collaborated with cross functional teams and mentored junior developers for clean, test-driven code.
- Collaborated with other team members to troubleshoot and resolve issues efficiently.

Certifications & Open Source

- ✓ npm Contributor: Published Turbogen (1k+ weekly downloads) and 4+ utility packages.
- ✓ Open Source: 75+ repositories with 50+ stars; active contributor to AI/ML tools.

Core Competencies

- ✓ API Design & Scalability | CI/CD Pipelines | Agile Development
- ✓ Problem-Solving | Cross-functional Collaboration | Technical Documentation

Interests

Generative AI Research, Cloud Native Architecture, Open Source Development