Muhammad Zeeshan

Agentic AI Engineer | Senior Python Backend Engineer

Lahore, Pakistan | zeeshare.com | Open to relocation and travel: Riyadh / Dubai
linkedin.com/in/zeeshan-ai-python-backend-engineer | github.com/Zeeshan138063

SUMMARY

Senior Python engineer with 8+ years of experience delivering production backends, integrations, data platforms and applied AI for startups and global enterprises, including clients in the USA and UK.

I build scalable, secure applications on both serverless and traditional architectures: Python, Django, FastAPI and Go, AWS (Lambda, API Gateway, Cognito), and microservices. I work across relational (PostgreSQL), document (MongoDB), key-value (Redis, DynamoDB) and time-series databases.

On the AI side I'm hands-on with Claude, OpenAI, LangChain, RAG, vector search, structured outputs, and tool and API integration.

I'm working toward Forward Deployed Engineer roles, where the job is to turn an ambiguous requirement into a working prototype and then into a reliable production service.

SKILLS

AI agents and conversational AI: Claude / Claude Code, OpenAI, LangChain, RAG, Embeddings, Hybrid search, Semantic search, Structured outputs, Tool and API integration, Voice APIs, Speech-to-text, Groq

Languages and frameworks: Python, Django / DRF, FastAPI, Flask, Pydantic, Go, JavaScript, TypeScript, Angular, PHP, Django Ninja, Pandas, React / Next.js

Integration and reliability: Async Python, REST, GraphQL, Webhooks, Queues, Retries, Idempotency, Error handling, Monitoring and observability, gRPC, WebSockets

Databases: PostgreSQL, PostGIS, MySQL, MongoDB, DynamoDB, Redis, pgvector, ChromaDB, Pinecone, InfluxDB, Prometheus

Architecture and messaging: Microservices, Monoliths, Domain-Driven Design, Event-driven systems, Cadence, API gateways, Celery, RabbitMQ, Kafka, SQS

Cloud and delivery: AWS Lambda, API Gateway, Cognito, S3, RDS, EC2, CloudFront, GCP, Docker, Kubernetes, Terraform, GitHub Actions, GitLab CI/CD, Bitbucket Pipelines, Grafana, Step Functions, ElastiCache, CloudWatch

Testing and code quality: Pytest, Test-driven development, Pre-commit hooks, Black, Flake8, isort, Bandit, Pylint

Automation and data acquisition: Scrapy, Playwright, Selenium, SeleniumBase, BeautifulSoup, Crawl4AI, Firecrawl, Zyte, Oxylabs, High-concurrency worker systems

Forward-deployed delivery: Technical discovery, Workflow mapping, Solution architecture, Rapid prototypes, Stakeholder collaboration, Documentation, Production support, Scrum, Jira

EXPERIENCE

Founder & Agentic AI Engineer (part-time), Independent

Jun 2026 - Present | Remote

- Prototype and productionize applied AI workflows using Claude, OpenAI, Python, FastAPI, RAG, structured outputs and API-based tools for search, research, operations and knowledge retrieval.
- Built a voice-enabled RAG search workflow: a voice API, speech-to-text conversion, then a retrieval pipeline that returns results from indexed knowledge sources.
- Convert successful prototypes into deployable FastAPI services with asynchronous processing, queues, persistence, retries, monitoring and clear API boundaries.

Software Engineer, Diametos GmbH

Sep 2025 - Present | Germany (remote)

- Backend engineering for sleep-diagnostics healthtech products at a German medical-technology company, in a regulated, quality-critical product environment.
- Own features from technical design through implementation, testing and deployment.
- Work daily with a Germany-based engineering team, keeping EU-hours overlap.

Senior Software Engineer, Nextbridge Ltd.
Jan 2023 - Sep 2025 | Lahore, Pakistan

- Developed DeepSerp, an AI SEO/GEO optimization platform that helps websites rank on AI search engines such as ChatGPT, Perplexity and Gemini. It crawls complex websites, converts content into structured data, and powers LLM-assisted analysis and recommendations.
- Built DeepSerp's universal web scraping layer, which extracts content from any public website stack (React, Vue, WordPress) and feeds the analysis pipeline.
- Designed and launched a centralized AI search engine in 2024 using Retrieval-Augmented Generation, LangChain and OpenAI embeddings. It syncs products from four backend systems, serves three frontends through one hybrid search (vector plus text) with millisecond-range responses, and solved the client's main product-discovery problem.
- Led development of a distributed scraping system that reduced end-to-end processing time from 4.5 hours to under an hour (about 75%) and lowered infrastructure cost by about 40% through adaptive concurrency, queue-based load balancing and resource optimization.
- Engineered high-throughput ingestion pipelines processing 1.4M+ products daily with 200+ concurrent workers, using enterprise proxy networks (Zyte, Oxylabs) with region-aware routing for geo-accurate requests across the US and Canada.

Python Developer, Ten Lifestyle Group
Dec 2021 - Dec 2022 | London, UK (remote)

- Built a RESTful API and a microservice from scratch using Django, Django REST Framework, Django Ninja and FastAPI, and debugged and resolved issues in the existing application.
- Developed Python services and low-latency REST, GraphQL and gRPC APIs for customer-facing and internal applications serving multiple global regions, with Celery for async tasks and Redis and AWS RDS for data and caching.
- Designed services following Domain-Driven Design principles, with clearly defined bounded contexts and aggregate boundaries, inside a microservices architecture.

Software Engineer, Ethyca
May 2021 - Dec 2021 | New York, NY (remote)

- Worked on Ethyca's privacy platform (Fides), which automates privacy requests across databases, data warehouses and third-party vendors.
- Took part in code reviews, debugging and refactoring to improve maintainability and long-term stability.
- Wrote unit and integration tests so new features met functional and non-functional requirements before release.

Backend Software Engineer, Emumba
Sep 2020 - Dec 2021 | Islamabad, Pakistan

- Contributed to the backend of Zero Trust Network Access (ZTNA) systems for Extreme Networks, ensuring only authenticated users and devices can reach internal applications in an identity-aware network environment.
- Built secure enterprise backend services and REST APIs in Python (Django/DRF) and Go across monolithic and microservice architectures.
- Implemented event-driven and long-running workflows using Cadence, and integrated Ambassador API Gateway for service routing and internal and external API management.

Python Developer (contract), Citibank
Mar 2020 - Apr 2021 | Canada (remote)

- Designed Python automation and a custom ETL pipeline processing millions to billions of records for enterprise banking reporting.
- Built shell scripts and Python tools to automate daily operational workflows, improving stability and reducing execution errors in production.
- Worked with stakeholders to understand business and compliance requirements and turn them into backend solutions that fit financial system constraints.

Full Stack Engineer, Crewlogix Technologies
Jan 2018 - Sep 2020 | Lahore, Pakistan

- Built web applications and APIs with Python and Django, securing them with OAuth and JSON Web Tokens.

- Designed and optimized databases in PostgreSQL, MySQL and DynamoDB, with caching in Redis and Amazon ElastiCache.
- Handled geographic data with PostGIS, including mapping and spatial queries.

PROJECTS

DeepSerp / DeepAudit (Nextbridge Ltd.)

A platform that crawls any public website, converts its content into structured data, audits it for AI search visibility, and recommends fixes. I built the universal scraping layer, the core AI modules and the backend that scores and tracks content in real time.

Technologies: Python, Django, PostgreSQL, Redis, MongoDB, ChromaDB, OpenAI API, LangChain, RAG

AI search engine for product catalogs (Nextbridge Ltd.)

A centralized AI search engine using Retrieval-Augmented Generation, shipped in 2024. It keeps products in sync from four backend systems, serves three frontends through one search API, handles complex product catalogs with many variations in size, model, memory and storage, and solved the client's main search problem.

Technologies: Python, FastAPI, LangChain, OpenAI embeddings, MongoDB vector search, Redis

Distributed scraping system (Nextbridge Ltd.)

I led the development of a distributed scraping system that cut end-to-end processing time from 4.5 hours to under an hour, about 75%, and lowered infrastructure cost by about 40%.

Technologies: Python, Django, Scrapy, Playwright, Zyte, Oxylabs, Redis, MongoDB, AWS

Voice-enabled RAG search (Independent work)

A search workflow you can talk to. I integrated a voice API, converted speech to text, and passed the transcribed query into a retrieval-augmented pipeline that returns relevant results from indexed knowledge sources.

Technologies: Python, Voice API, Speech-to-text, RAG

Location-based search with PostGIS (Crewlogix Technologies)

PostGIS-backed proximity search: finding all matching records within a set distance of a position, for location-aware application features.

Technologies: PostgreSQL, PostGIS, Python, Django

EDUCATION

B.S. Computer Science

FAST National University of Computer and Emerging Sciences, Pakistan | 2013 - 2017

CERTIFICATIONS

- Anthropic: Claude 101; Claude Code 101
- Google Cloud (2025): 9 skill badges and 22 completion badges, including Use Machine Learning APIs on Google Cloud; Engineer Data for Predictive Modeling with BigQuery ML; Vector Search and Embeddings; Machine Learning Operations (MLOps) for Generative AI; Generative AI Explorer - Agent Platform
- Full list with verification links: zeeshare.com/credentials
- In progress: CCAR-F: Claude Certified Architect, Foundations

COMMUNITY

- Moderator and LeetCode Instructor, iCodeGuru. Mentoring engineers, running code reviews and supporting interview preparation.