

ABHAY MUNJ

Python Developer | Django Backend | Web Scraping | Generative AI & LLM Systems

Chandigarh, India | +91 8288817478 | abhaymunj.20@gmail.com

[linkedin.com/in/abhaymunj](https://www.linkedin.com/in/abhaymunj) | github.com/Abhay96Munj | [Upwork](https://www.upwork.com)

PROFESSIONAL SUMMARY

Python Developer with 4+ years building production Django/FastAPI backends, large-scale web scraping pipelines, and LLM-powered applications. Experience with hybrid-retrieval RAG pipelines, LLM tool-calling agents, LLM observability, N8N workflow automation, and OpenAI API integration, plus data extraction at scale using Selenium and BeautifulSoup. Shipped end-to-end systems from a multilingual patent-processing CLI for an international legal team to a production AI travel platform.

TECHNICAL SKILLS

Languages: Python, SQL

Backend: Django, Django REST Framework (DRF), Django ORM, FastAPI, RESTful API Design, Celery, Redis

Web Scraping & Automation: Selenium, BeautifulSoup (bs4), requests, Regex, Bright Data (proxy rotation), Rate Limiting, Data Pipelines, N8N

AI / GenAI: OpenAI API, Claude API, Gemini API, RAG (Retrieval-Augmented Generation), Hybrid Retrieval (Dense + BM25), Embeddings, Vector Search (FAISS), Reranking (CrossEncoder), RAG Evaluation (hit-rate, MRR), Tool-Calling Agents, Function Calling, LLM Pipeline Design, Prompt Engineering

AI Observability: LangFuse, LangSmith

Databases: PostgreSQL, MySQL, Redis, NoSQL

Cloud & DevOps: AWS S3, Docker, CI/CD Pipelines, GitHub Actions

Libraries & Tools: PyMuPDF, pdfplumber, OCR, Websockets, Firebase, Tkinter, PyInstaller (Windows/macOS packaging), Git

AI Dev Tools: Cursor IDE, Codex, Claude Code, Google Stitch

3rd Party APIs: Google Maps API, Stripe (Checkout, Webhooks), Apollo, Firecrawl, Booking/Flight/Hotel APIs

PROFESSIONAL EXPERIENCE

IDS Infotech — Python Developer Oct 2023 – Present

Chandigarh, India

- Developed scalable RESTful APIs using Django, Django REST Framework, and FastAPI, consumed by frontend and mobile clients across multiple production projects.
- Built and integrated LLM-powered workflows into production applications using the OpenAI API.
- Implemented asynchronous background task processing with Celery, decoupling long-running workflows from the request cycle and improving throughput.
- Optimized Django ORM queries using select_related/prefetch_related and caching strategies, reducing redundant database hits and improving API response times.
- Mentored junior developers on Python best practices and led code reviews.

INT-Factor — Python Developer Dec 2021 – Oct 2023

Chandigarh, India

- Designed, developed, and shipped Python applications including API integrations between software components and modular, reusable backend systems.
- Implemented data scraping, web crawling, and workflow automation pipelines across 3 client engagements.
- Improved code reusability through modular programming; owned full design-develop-test-deliver cycles independently.

PROJECTS

AI Travel Planner — orahee.com

Stack: Python, Django, OpenAI API, LangFuse, Google Maps API, Celery, Redis, PostgreSQL, Websockets, Firebase, AWS S3, Docker

- Built an AI-driven travel platform that generates personalized day-by-day itineraries from a user's dates and destination using the OpenAI API, with orchestration logic built in-house rather than on a framework; deployed to production with ~200 active users.
- Cut generation latency and OpenAI costs by moving Google Places and maps lookups out of the in-LLM call path into the backend pipeline, enriching model output with API data after generation rather than during it.
- Integrated weather, maps, and flight/hotel booking APIs with rescheduling logic that re-adjusts itineraries in real time; instrumented with LangFuse to trace and debug model behaviour in production.
- Built the real-time backend for multi-user co-planning (Websockets + Firebase) powering in-app messaging and voting, plus budget tracking and dynamic per-trip AWS S3 provisioning.

CORSEARCH — International Patent Processing CLI Tool

Stack: Python, PyMuPDF, pdfplumber, OCR, Regex, JSON, GitHub Actions

- Built a CLI tool for an international legal team automating parsing and structured data extraction from government patent filings across 20+ countries.
- Engineered parsers resilient to inconsistent, non-standardized filing layouts across jurisdictions, sustaining a 90-95% extraction success rate on documents supplied by the upstream team.
- Parsed PDFs across 6+ languages (English, Mandarin, Russian, French, Spanish, and others) into normalized JSON consumed by downstream legal applications.
- Extracted embedded patent images and built automated validation scripts that cut manual QA effort by 30-40%; shipped a GitHub Actions CI/CD pipeline for cross-team deployment.

Large-Scale Web Scraping & Data Extraction Pipelines

Stack: Python, Selenium, BeautifulSoup, Bright Data, Regex, PostgreSQL

- Built production scrapers for e-commerce review mining (Amazon, Medela AG) and property data aggregation (Zillow, Yelp) across multiple high-volume sources.
- Handled dynamic JS-rendered pages, pagination, session management, and rate limiting; implemented retry logic and Bright Data proxy rotation, sustaining a ~95% success rate on e-commerce review extraction.
- Normalized scraped output into PostgreSQL schemas with automated validation to catch layout drift and silent extraction failures.

SALESLY — AI Sales Automation & Lead Intelligence Platform

Stack: Python, OpenAI API, LangFuse, N8N, Firecrawl, Apollo API, PostgreSQL, Docker

- Designed a staged LLM pipeline orchestrated in N8N — intent classification, lead sourcing and enrichment via Apollo, then OpenAI-generated outreach copy — returning 20+ qualified leads per query.
- Built the backend powering a ChatGPT-style query interface for company and industry data — parsing incoming queries and routing them to an N8N workflow enriched via Firecrawl and Apollo APIs.
- Instrumented end-to-end LLM observability with LangFuse — tracking prompt versions, latency, token usage, and output quality in production; used trace data to identify agent failure modes and introduce prompt guardrails that improved output reliability.

RAG-101 — Hybrid Retrieval RAG API | Personal Project (github.com/Abhay96Munj/RAG-101)

Stack: Python, FastAPI, FAISS, BM25, sentence-transformers, CrossEncoder Reranking, Gemini API, LangFuse, Docker

- Built a RAG API with hybrid retrieval — FAISS dense embeddings and BM25 sparse search fused via Reciprocal Rank Fusion, then reranked with a CrossEncoder to the top-3 chunks passed to the LLM.
- Implemented a reranker score threshold that refuses to answer rather than hallucinate when no chunk clears the bar; every answer returns page-level source citations.
- Built an evaluation harness over a golden set measuring hit-rate and MRR per pipeline stage plus LLM-judge faithfulness, with config snapshots enabling regression detection across runs.
- Added a multi-turn tool-calling agent with nested LangFuse tracing across retrieval, rerank, and generation spans; containerized via Docker Compose.

ADVILEX — Backend & CMS for Regulatory Compliance Platform (advilex.com)

Stack: Python, Django, Django REST Framework, PostgreSQL, AWS S3, Docker

- Built the Django backend powering the public platform of a regulatory compliance consultancy serving Fortune 500 clients across 50+ countries (Heineken, Mars, TotalEnergies, Fonterra).
- Designed an admin panel-driven CMS letting the client publish services, case studies, and announcements independently with no developer involvement, and built the REST API layer and data models behind a region-segmented regulatory news feed (North America/LATAM, EU, Asia Pacific) consumed by the frontend team.
- Implemented dynamic per-content AWS S3 provisioning with presigned-URL asset delivery for client-facing media.

INTELLIX — Subscription Commerce & Licensing Backend (bluknox.com)

Stack: Python, Django, Django REST Framework, PostgreSQL, Stripe (Checkout Sessions, Webhooks), PyInstaller, AWS S3, Docker, Tkinter

- Built the Django backend for the storefront of a data-encryption product (AES-256, HIPAA/PHI-oriented), covering product tiers, monthly/annual plans, add-ons, coupon codes, and free trials.
- Integrated Stripe Checkout Sessions and webhook handlers to manage the full subscription lifecycle, reconciling plan and add-on state server-side as the source of truth for entitlements.
- Designed a licensing and entitlement API that resolves each customer's active plan and purchased add-ons (storage tiers, devices, sharing) and serves that state to the desktop client, which gates feature access.
- Packaged cross-platform installers (Windows 10+, macOS 13+) with PyInstaller requiring no admin privileges; managed AWS S3 assets and Docker deployment with a staging environment for pre-release QA.

EDUCATION & AWARDS

B.Com — Chandigarh, India

Graduated 2020

Employee of the Quarter, Q2 2025–2026 (IDS Infotech) • Best Team Award, Q2 2025–2026 (IDS Infotech)