

Muhammad Umar Nawaz Janjua

Islamabad, Pakistan | +92-310-9122679 | umar.111nawaz@gmail.com
linkedin.com/in/muhammad-umar-nawaz-janjua-99a702233 | github.com/umar052001 | janjua.dev

PROFESSIONAL SUMMARY

Versatile Software Engineer with 3+ years of experience across backend systems, AI/ML integration, full-stack development, and cloud infrastructure. Proven ability to architect scalable platforms processing millions of records, design LLM-powered applications with RAG and agentic workflows, and deliver end-to-end solutions. Demonstrated success reducing operational costs by 90%, winning international hackathons, and building production systems on GCP and AWS.

PROFESSIONAL EXPERIENCE

Software Engineer

Apex AI Arabia | Saudi Arabia (Remote)

Jul 2026 -- Present

- Architecting high-throughput backend microservices and distributed systems using Go, Rust, and Python with strict latency SLAs and thread-safety targets.
- Engineering enterprise-grade RAG pipelines and multi-agent routing frameworks with vector search indexing and low-latency LLM inference patterns.
- Driving performance profiling and query optimization across PostgreSQL and Redis, developing custom data-extraction modules to remove processing bottlenecks in high-volume pipelines.
- Authoring clean, scalable gRPC and RESTful interfaces with robust observability, error handling, and automated CI/CD deployments for cross-team consumption.

Software Engineer

Modern Intelligent Solutions | Riyadh, Saudi Arabia

Jun 2024 -- Jan 2026

- Architected an Enterprise Agentic RAG System enabling complex reasoning over 5,000+ Arabic and English research documents using LangGraph and Vertex AI.
- Engineered a production B2B Outreach Ecosystem using LLMs for personalized outreach and data extraction, reducing operational costs by 90%.
- Deployed multi-modal AI Voice Agents with Gemini and LiveKit for real-time customer interaction and automated scheduling.
- Architected high-concurrency backend services in Golang (Chi Router) for AI-driven credit scoring with strict data validation and transactional integrity.
- Built full-stack components including React dashboards, RESTful and gRPC APIs, PostgreSQL optimization, and real-time data streaming pipelines.
- Managed the full ML lifecycle including data preprocessing, model deployment, CI/CD, and infrastructure scaling on GCP.

Software Engineer

Voilabs | Paris, France (Remote)

Nov 2023 -- May 2024

- Fine-tuned DONUT (Document Understanding Transformer) models for high-accuracy, OCR-free data extraction from complex French financial documents.

- Built a full-stack AI E-Learning Platform using Next.js, FastAPI, and PostgreSQL, integrating LLMs with Tavily for real-time content generation.
- Developed a Resume Automation Engine leveraging NLP and XML manipulation to automate formatting and extraction for 1,000+ documents.
- Implemented React dashboards for monitoring financial document processing workflows with real-time status tracking.

Open Source Developer / Hackathon Winner

Self-Employed | Remote

Jan 2022 -- Present

- Authored Blitz Parse, a high-performance PDF extraction library in Rust with Python bindings, outperforming PyMuPDF for large-scale document processing and RAG ingestion.
- Won the Stability AI Hackathon (LabLab.ai) building a VS Code extension for low-latency code completion using Stable Code LLMs, competing against 500+ global participants.
- Created Invoice Watcher, a cross-platform Electron.js desktop application with React UI for real-time directory monitoring and automated Google Sheets data entry.

PROJECTS

Imagini -- B2B Leads Platform | Nov 2025 -- Jan 2026

FastAPI, React, PostgreSQL, Redis, Docker, GCP

- Architected a full-stack data platform hosting millions of B2B leads with streaming uploads and real-time administrative control via React dashboard.
- Implemented Async FastAPI with streaming parsers for massive CSV/Excel uploads without memory spikes, achieving sub-second search.
- Built a fraud-resistant credit consumption system with tiered access (Verified vs Unverified) and advanced search/export.
- Deployed on GCP with Docker, autoscaling, Cloud Functions, and Celery/Redis task queues.

Email Verifier Engine | Sep 2025 -- Nov 2025

Python, AsyncIO, FastAPI, React, GCP, Redis, Celery

- Built a proprietary email verification system replacing commercial tools, reducing costs by 90% while processing hundreds of thousands of emails daily.
- Implemented multi-layered verification (syntax, DNS/MX, deep SMTP simulation) with IP rotation, rate limiting, and a Guesser module.
- Developed a React analytics dashboard with real-time stats, batch upload management, and collaboration features.
- Deployed on GCP with autoscaling, Cloud Functions, and Celery/Redis distributed workers.

Blitz Parse -- High-Performance PDF Extraction in Rust | 2025 -- Present

Rust, Python Bindings, PyPI, PyMuPDF Benchmarks

- Authored a high-throughput PDF extraction library written in Rust with Python bindings, achieving 3x throughput over PyMuPDF in benchmarks.
- Rust core with zero-copy parsing and memory-efficient page handling for multi-thousand-page documents.
- Published on PyPI with automated cross-platform CI/CD across Linux, macOS, and Windows via GitHub Actions.

AI Research Agent CLI | 2025 -- Present

Python, AsyncIO, LLM, CLI, Web Crawling

- Autonomous multi-stage research engine combining web crawling, concurrent async pipelines, real-time synthesis, and local LLM reasoning in an interactive CLI.
- Orchestrates multi-step research workflows: query expansion, parallel crawling, content extraction, relevance filtering, and synthesized reports.
- Supports local LLM inference via Ollama for fully offline research with configurable models and context management.

Invoice Watcher | 2024 -- Present

Electron.js, React, File Watcher, Google Sheets API

- Automated system monitoring incoming directories for real-time document change detection and structured extraction from financial invoices.
- Cross-platform Electron.js application with React UI for real-time directory monitoring and automated Google Sheets data entry.
- Implements filesystem watchers with debounce logic, structured extraction, and automatic spreadsheet synchronization.

EDUCATION

Bahria University | Islamabad, Pakistan

Bachelor of Science in Software Engineering | Sep 2020 -- Sep 2024

Coursework: Data Structures, Algorithms, Database Systems, Software Engineering, Machine Learning, Artificial Intelligence, Computer Networks, Operating Systems

TECHNICAL SKILLS

Core: Python, Go, Rust, TypeScript, SQL, Docker, Linux, Git

CERTIFICATIONS

- Machine Learning Specialization -- Stanford Online (Coursera)
- Meta Frontend Developer Specialization -- Meta (Coursera)
- Build JavaScript Applications with Node.js -- Microsoft
- Generative AI Fundamentals -- Google
- Introduction to Large Language Models -- Google
- Introduction to Generative AI -- Google
- Introduction to Responsible AI -- Google
- The Future of Payment Technologies -- University of Michigan
- Integrating AWS with the SDK -- Amazon Web Services
- Amazon S3 Basics -- Coursera
- Calculus for Machine Learning and Data Science -- DeepLearning.AI
- Linear Algebra for Machine Learning and Data Science -- DeepLearning.AI
- Winner -- AudioCraft 24-Hours Hackathon -- LabLab.ai
- Runner-up -- StableCode 24-Hours Hackathon -- LabLab.ai
- Eleven Labs AI Hackathon Certificate -- LabLab.ai