

WAQAS SHARIF

AI Engineer | LLM Reliability | Agentic Systems in Production

Karachi, Pakistan (UTC+5) | Open to remote, freelance and contract work | ma.sharif.waqas@gmail.com
linkedin.com/in/ma-waqas-sharif | github.com/Waqas01CP | waqas01cp.github.io

PROFESSIONAL SUMMARY

Software Engineering graduate (BESE, KIET) specialising in LLM reliability and behavioural testing. Builds AI-native: designs the architecture and the algorithms, makes the engineering trade-offs, delegates implementation to coding agents such as Claude Code for speed, and verifies their work through automated tests, mutation testing and independent audits.

3.7 s first response | 592 automated tests | 19 behavioural probes | 5.4 h median job discovery | 3 named failure modes

TECHNICAL SKILLS

Agentic AI and LLM: LangGraph, Google ADK, Gemini API, OpenAI API, Anthropic API, Guardrails, Structured Outputs
AI-Native Development: Claude Code, Claude Design, Agent Skills, Model Context Protocol (MCP), Spec-Driven Development, Directing and Reviewing Coding Agents
Languages: Python (primary), JavaScript, Dart, C/C++, C#, Java
Backend and Infrastructure: FastAPI, PostgreSQL, Supabase, MySQL, Docker, Railway, Render, Vercel, Data Pipelines (ETL)
Tools and Frontend: Git / GitHub, GitHub Actions, LangSmith, Flutter
Software Engineering Practices: ADRs, Mutation Testing, Fitness Functions, Independent Code Audits

PROJECTS

Rahzaan: AI Career Guidance Platform Feb 2026 to Present

Primary Architect and Director

Python, LangGraph, FastAPI, Flutter, Supabase. Deployed agentic counselling platform for Pakistani secondary school students.

- Evaluation harness.** Built the platform's evaluation harness: **19 behavioural probes** using paired before-and-after measurement, **60 per-turn traces** across five personas, and 592 automated tests catching regressions before users hit them.
- Root cause.** Refuted my own diagnosis of why the system over-asserted: measuring across **five data-completeness levels** showed the thinnest persona was the most honest, and **internal signal divergence** was the real trigger. Redirected the fix, then added a deterministic post-generation gate as backstop.
- Architecture.** Architected a **LangGraph** pipeline of 24 nodes with fan-out across 9 concurrent nodes and 52 typed state fields, cutting first-response latency from 20 to 25 seconds down to **3.7 seconds**.

Job Aggregator: Automated Job Discovery Pipeline Sep 2026 to Present

Architect and Director

Python, GitHub Actions, Airtable. In development, and running in production since 17 Sep 2026. I design the system and its algorithms and delegate implementation to Claude Code for speed, under specifications I approve.

- Outcome.** Cut job discovery from a by-hand baseline I estimated at three to six days to a measured **median of 5.4 hours** after publication, over 114 postings. Across **30 production runs**, zero requests failed.
- Method.** Built it through a four-role AI-native development method of my own design: I set the blueprint, decide every brief and audit and make the trade-offs; an architecture chat refines the design with me, Claude Code implements and a cold read-only reviewer audits. **55 decision records**, 754 tests and 410 recorded mutations that prove the tests notice.

RESEARCH

Applied LLM Workflow Research Sep to Dec 2025

Python, Streamlit. Behavioural study identifying reproducible failure modes in production LLM workflows.

- Failure modes.** Named three failure modes, Optimal Answer Trap, Final Synthesis Bias and Conversational Gravity, each reproduced through documented prompt sequences, with sampling experiments and **1M-token** context stress tests.

WORK EXPERIENCE

HubIT Karachi, Pakistan

AI Automation Engineer (Intern)

June 2026 to July 2026

- Built a **Python** pipeline extracting recurring utility documents into **schema-validated billing records** for a residential property management client, replacing a manual data entry step; enforced strict validation with repair and retry on malformed model output, blocking bad records before they reached billing.

EDUCATION

Karachi Institute of Economics and Technology (KIET)

Karachi, Pakistan

BE Software Engineering, CGPA 3.47 / 4.0

Sep 2022 to Aug 2026

CERTIFICATIONS

5-Day AI Agents Intensive, Google and Kaggle (2025) | Prompt Engineering Specialization, Vanderbilt (2025) | Google Prompting Essentials (2025) | Python 3 Programming Specialization, Michigan (2025)