

SOHAIB ALI KHAN

Backend Software Engineer · AI Systems · iOS Infrastructure

+92 307 4313968 · hello@sohaibalikhan.com · linkedin.com/in/sohaibali38 · sohaibalikhan.com · github.com/sohaib1268

SUMMARY

Backend engineer with experience building distributed systems, AI-powered platforms, and production iOS infrastructure. Currently leading iOS infrastructure and AI agent development at Sofy.ai, where I own a dedicated iOS microservice, ship autonomous agents that execute on real devices, and work across Java/Spring Boot and Python daily. Before that, I built real-time game backends at Frag Games using AKKA to support thousands of concurrent players. I care about writing systems that stay up, recover gracefully, and actually work under pressure.

EXPERIENCE

Software Engineer, iOS Infrastructure & AI Lead

03/2025 – Present

[Sofy.ai](https://sofy.ai)

Lahore

AI-powered mobile infrastructure platform used by QA teams to automate workflows on real iOS and Android devices.

- Built and own a dedicated Java/Spring Boot microservice for iOS that handles device lifecycle, app installation, pairing, trust flows, and execution orchestration. Also own the broader iOS infrastructure layer including JAMF MDM integration and Appium/XCUI Test tooling.
- Architected and shipped an AI Agent for iOS using Python and our existing Spring Boot backends. The agent is backed by a knowledge base and can run workflows on real devices autonomously.
- Developed the platform's core element-matching algorithm in Java for identifying UI elements reliably even when XML hierarchies shift between runs. Added OpenCV-based visual matching as a hybrid fallback for complex UI flows.
- Introduced deep linking support, driver recovery logic, and intelligent retry mechanisms, which together brought execution failures down by roughly 35–40%.
- Optimized the Java-based communication layer between execution services and device controllers, cutting end-to-end execution time by about 30% under load.
- Built scalable Spring Boot microservices for XML parsing, workflow orchestration, execution lifecycle management, and result aggregation. Deployed and maintained on Azure.
- Drove observability, monitoring, and recovery workflows across the iOS device farm. Kept downtime consistently low through proactive health checks and automated recovery.

Junior Software Engineer

04/2024 – 03/2025

[Frag Games](https://fraggames.com)

Lahore

Game studio building real-time multiplayer titles.

- Built concurrent, fault-tolerant backend services in Java using AKKA that supported over 10,000 concurrent players across live titles including Antara and Elowyn: The Quest of Time.
- Shipped core gameplay features like matchmaking, leaderboards, real-time chat, shield mechanics, troop systems, and reward multipliers.
- Designed REST and gRPC APIs using Protocol Buffers for efficient, low-latency communication between services.
- Integrated RabbitMQ for async event processing and built a MERN stack CMS and ops dashboard for live monitoring of game state and player activity.
- Deployed services on AWS, contributed to CI/CD pipelines, and improved overall production stability.

Research Assistant

08/2023 – 04/2024

[Computational Biology Research Lab, FAST Lahore](https://fastlab.org)

Lahore

- Developed and deployed a React Native + MERN stack application for the Pakistan Anti-TB Association (USAID-funded), covering TB screening, case tracking, and field data reporting.
- Designed and built an AR-based indoor navigation system in Unity for a major hospital in the UK. Delivered as a Final Year Project and actively used by students for campus navigation.

PROJECTS

AI Agent for iOS | *Sofy.ai*

- Built an autonomous agent with a structured knowledge base, integrated over existing Python services and Spring Boot microservices. The agent interprets workflow goals, selects UI actions, and handles device-specific edge cases without manual scripting.

AR Scanner | *Personal Project* [[GitHub](#)]

- AR application that lets users scan real-world objects, converts them into 3D models, and places them back into a live AR environment. Built to explore the intersection of computer vision and spatial computing.

Indoor Navigation System | *King's College London*

- AR-based navigation system built in Unity for a UK hospital. Implemented gaze interaction, Firebase integration, and generative-AI-driven advertisement overlays for indoor spaces.

Anti-TB Platform | *USAID-funded*

- Full-stack React Native + MERN application supporting tuberculosis screening, case tracking, and stakeholder reporting. Deployed via Netlify/Vercel for the Pakistan Anti-TB Association.

Skin Cancer Melanoma Detection | *Academic Project*

- Image classification pipeline using EfficientNet on the HAM10000 dataset. Focused on preprocessing, augmentation, and model evaluation to improve melanoma detection accuracy.

SKILLS & TECHNOLOGIES

Languages:	Java, Python, Golang, JavaScript / TypeScript, C/C++
Backend:	Spring Boot, AKKA, Node.js, REST, gRPC / Protocol Buffers, RabbitMQ, WebSockets
AI / Agents:	LLM integration, AI agent development, knowledge base design, OpenCV, EfficientNet
Mobile / iOS:	Appium, XCUITest, libimobiledevice, JAMF MDM, APNs, iOS device orchestration
Cloud & Data:	Azure, AWS, Docker, CI/CD (Azure DevOps, GitHub Actions), MongoDB, PostgreSQL, Firebase

EDUCATION

Bachelor of Software Engineering

[National University of Computer and Emerging Sciences \(FAST\)](#)

Dean's List Fall 2023 · GPA: 3.74 / 4.0

09/2020 – 05/2024

Lahore

CERTIFICATIONS

[Go Bootcamp: gRPC and Protocol Buffers \(HTTP/S, HTTP2\)](#)