Mohammad Yehya Hayati

Portfolio: mohammadyehya.github.io
Github: github.com/MohammadYehya
Email: mohammad7446@gmail.com
Mobile: +923352495576

EDUCATION

- **National University of Computer & Emerging Sciences** — Karachi, Pakistan
  *Bachelor of Science - Computer Science; **GPA: 3.90/4.00*** — *Aug 2021 - Jun 2025*

SKILLS SUMMARY

- **Programming Languages**: Python, JavaScript / TypeScript, C/C++, Rust, SQL, Bash, PowerFX
- **AI/ML Frameworks & Tools**: Agentic RAG, Multi-Agent Orchestration (LangGraph, LangChain, CrewAI), MCP (Model Context Protocol), Fine-tuning (LoRA/QLoRA), Transformers, Computer Vision, TensorFlow, PyTorch, Scikit-Learn
- **Production & MLOps**: LLM Evaluation (RAGAS, DeepEval), Experiment Tracking (MLflow, W&B), Docker, Kubernetes
- **Web & Backend Stack**: FastAPI, Next.js, Node.js, Express.js, Socket.IO, TailwindCSS
- **Cloud & Infrastructure**: Azure, AWS, GCP, Vector DBs (FAISS, Chroma, Pinecone), Vercel

(Full List on Portfolio)

EXPERIENCE

- **Systems Limited** — Onsite
  Associate Consultant — *Jul 2025 - Present*
  - **Agentic AI RFP Automation**: Architecting an autonomous multi-agent system to automate the Request for Proposal (RFP) lifecycle. Developed hierarchical agent workflows using LangGraph that perform document ingestion, multi-step planning, and autonomous tool-calling to generate technical proposals.
    * **Impact**: Reduced manual proposal generation cycles by an estimated 80% and improved technical compliance matching accuracy.
  - **Mashreq Bank Cash Digitization (Backend Engineering)**: Engineering high-performance backend modules using Java and Spring Boot for Mashreq's "Islamic-first" digital retail platform.
    * Developing secure RESTful APIs for the Cash Digitization module, enabling paperless cash-heavy workflows and straight-through processing for retail customers.
    * Implementing Spring Security and Shariah-compliant business logic to ensure international-grade security standards and regulatory compliance (SBP).
  - **Enterprise Operational Excellence**: Implementing automated testing frameworks and CI/CD pipelines for internal business units, reducing manual deployment overhead by 40%.

- **Imperium Dynamics** — Onsite
  Software Engineer Intern — *Apr 2025 - Jun 2025*
  - **Pro-Code Extension Development**: Developed complex C# server-side plugins for the Microsoft Dataverse environment to enforce mission-critical business logic and handle high-volume data transformations.
  - **System Integration**: Integrated Dynamics 365 with external third-party APIs and legacy services, ensuring seamless data flow across the Microsoft Power Platform.
  - **Workflow Automation**: Engineered Power Automate flows for automated email notification systems and multi-stage approval cycles.

- **Toyota Indus Motor Company** — Remote
  Full Stack ML Engineer - Collaboration — *Oct 2024 - May 2025*
  - **Production Anomaly Detection**: Engineered a real-time vehicle defect detection system using PyTorch and YOLOv8. Integrated a high-concurrency FastAPI backend with a Next.js dashboard to monitor 4 to 9 camera feeds.

- ∗ **Impact**: The anomaly detection system minimized vehicle defect-related downtimes, reducing resource wastage and improving overall efficiency. Worker sequence verification ensured standardized manufacturing processes, leading to higher-quality vehicles and fewer post-production corrections, ultimately enhancing Toyota's productivity and cost-effectiveness.

- **EyeconAI**   Onsite
  Intern ML Engineer   *Jun 2024 - Sep 2024*
  - ○ **Production-Ready WebApp**: Designed and developed a full-stack web application using Next.js, Express.js, and FastAPI, integrating user authentication, API endpoints, and cloud deployment, making EyeconAI a scalable SaaS product.
  - ○ **Context-Aware Models**: Built and fine-tuned machine learning models using TensorFlow and PyTorch, to enhance real-time adaptability to different environments.

- **Syslab.ai**   Onsite
  Research Intern   *Nov 2023 - Jun 2024*
  - ○ **Generate Large Dataset**: Employed data augmentation (rotation, scaling, noise addition) and preprocessing techniques to expand and refine datasets, enhancing model robustness and accuracy for AI applications.
  - ○ **Data Vizualization on WebApp**: Designed and implemented an intuitive dashboard using Next.js and TailwindCSS, allowing us to analyze trends, and evaluate model performance on the dataset mentioned above.

RESEARCH PUBLICATIONS

- **Context-Aware Detection of Mixed Critical Events using Video Classification**: Published on arxiv - arxiv.org/abs/2411.15773v1
  - ○ Developed a novel context-aware system for smart city safety, achieving 99.25% F1-score on traffic incident detection and 91.59% F1-score on fire detection.
  - ○ Utilized DenseNet-Transformer architectures to process spatial-temporal features, outperforming standard CNN baselines by 3% in prediction accuracy.

CERTIFICATIONS

- **Microsoft Azure Fundamentals**: Gained foundational knowledge of cloud concepts, core Azure services, plus Azure management and governance features and tools. (⊟ Verify)

- **Microsoft Azure Data Fundamentals**: Gained foundational knowledge of core data concepts related to Microsoft Azure data services. (⊟ Verify)

- **Microsoft Azure AI Fundamentals**: Gained fundamental AI concepts related to the development of software and services of Microsoft Azure to create AI solutions. (⊟ Verify)

- **Microsoft AI Business Professional**: Gained foundational knowledge of how artificial intelligence can be applied in business scenarios, including AI workloads, responsible AI principles, and identifying opportunities for AI-driven solutions. (🛡 Verify)

PROJECTS

- **VoltOS**: A linux distribution made solely with the vision of speed, fluidity, and customizable. It's lightweight and speedy characteristics come from being based on Arch Linux, with fluid animations that is meticulously built using Quickshell. Its premise is a highly customizable yet fast and fluid desktop environment.
  - ○ **Agentic OS Interaction**: Engineered a unified AI desktop environment featuring a built-in chatbot with Model Context Protocol (MCP) integration, allowing the LLM to execute system-level tool-calling and device management.
  - ○ **Multimodal AI Integration**: Implemented real-time SAM (Segment Anything Model) features for image segment layering and completion, alongside Speech-to-Text (STT) typing and Super Resolution.

- **System Control Architecture**: Developed a comprehensive suite of "Quickshell" store modules including a Network Manager (VPN/Firewall), Media Control, Workflow Builder (Cron-style), and real-time System Performance Stats, effectively eliminating any bloatware.
- **Integrated Services**: Feature-rich environment with LAN chat/sharing, Obs-based Recording Center, clipboard management, and dynamic Alt+Tab task switching.

- **LinguaLink**: An AI-powered chat platform that integrates real-time translation to enable seamless multilingual communication. Built with Next.js, Express.js, FastAPI, and TensorFlow, it uses Socket.io for real-time messaging and AI models for accurate translations.
  - **Real-time Synchronization**: Engineered a high-concurrency chat platform using Socket.io for bi-directional communication, achieving sub-1000ms translation latency between English and French.
  - **Model Optimization**: Implemented custom Seq-Seq and Transformer models to ensure high-accuracy performance even in resource-constrained environments.

- **CoinScope**: A real-time cryptocurrency tracker that provides users with up-to-date insights. Built with Next.js and TailwindCSS for a sleek UI, it leverages Kafka for efficient data streaming, ensuring instant updates on asset values and transactions.
  - **Event-Driven Architecture**: Engineered a high-throughput data pipeline using Apache Kafka to decouple market data producers from frontend consumers, ensuring scalable handling of high-velocity cryptocurrency price updates.
  - **Real-time Visualization**: Integrated Socket.io with a NodeJS backend to stream live asset values and transaction volumes to a sleek Next.js dashboard with zero-latency synchronization.

- **Memento**: A photo album application with an all in a user-friendly interface which includes, user sign-in and sign-out functionalities, photo uploads, and a favorites system. Built with NextJS, TailwindCSS, ShadCN/UI, Supabase.
  - **Frontend Performance**: Built an intuitive UI using ShadCN/UI components and Tailwind CSS, featuring a dynamic favorites system.

Honors and Awards

- Graduated Summa Cum Laude (Top 0.5%) - Jun 2025
- Placed in Rector's List (4 SGPA) x 3 - Jan 2022, Jan 2023, May 2023
- Placed in Dean's List (3.5+ SGPA) x 5 - May 2022, Jan 2024, May 2024, Jan 2025, May 2025
- First Place at Database Odyssey Competition - Mar 2025
- First Place at Openhouse Database Project Competition - Dec 2024
- First Place at Procom Math Olympiad Competition - Apr 2023
- Runner's Up at Coder's Cup Speed Programming Competition - Nov 2021
- Achieved a score of 8 in Academic IELTS - Nov 2025