# Muhammad Nouman Abbasi

+1 (765) 357 5928 | thenoumanabbasi@gmail.com | NoumanAbbasi.com | linkedin.com/in/MNoumanAbbasi

## Education

**Purdue University, West Lafayette, IN** May 2023
*MS Computer Science* – **GPA 3.93/4.0**
Coursework: Compilers, Computer Networks, Software Security, Algorithm Design, Data Mining

**Lahore University of Management Sciences (LUMS), Lahore** May 2021
*BS Computer Science* – **GPA 3.97/4.0 (Top 1%)**
Coursework: Data Structures, Operating Systems, Machine Learning, Deep Learning, Hardware Architectures for AI

## Technical Skills

**Languages:** Python, Java, Rust, C/C++, JavaScript, TypeScript, SQL
**Frameworks & Libraries:** Node.js, React, Django, GraphQL, RESTful APIs, NumPy, Tokio (Rust), JUnit
**Datastores & Messaging:** Redis, PostgreSQL, ScyllaDB, DynamoDB, RabbitMQ, Amazon SQS
**Cloud & Tools:** AWS, Docker, Jenkins, Marathon, Grafana, Kibana, GitHub, Linux, Jupyter

## Work Experience

**Amazon - AWS Bedrock** Feb 2025 – Present
*Software Engineer* *Seattle, WA*

- Built an end-user scoring feature, enabling customers to submit custom identifiers for aggregated analysis; leveraged DynamoDB, Lambda, CloudWatch, and a microservices architecture for high-volume requests.
- Architected the team's core infrastructure pipeline from scratch, automating production accounts and baseline service setup, reducing regional deployment time from **3–5 days to 2–3 hours**.
- Owned a critical billing integration, creating HLDs/LLDs, data contracts, and a new ingestion algorithm, delivering the solution in **under one month** for the AWS re:Invent launch.

**Hive AI** June 2023 – Feb 2025
*Software Engineer 2* *Seattle, WA*

- Launched "Portal," a self-serve, centralized hub for the company's AI/ML offerings, consolidating usage and billing while enabling direct customer engagement.
- Built a high-throughput, distributed sliding-window rate limiter in Redis and Rust, supporting per-IP, per-model, and per-API limits, and handling up to **100,000 RPS** with minimal latency overhead.
- Refactored the complete backend logging system with async multithreading, reducing **p90** latency by **40%** across customer API requests, from **4.6s** to **2.8s**.
- Led the migration of message queues (RMQ) from Classic to Quorum, yielding a **20-40%** throughput improvement and reduced memory usage.

**Meta** May 2022 – Aug. 2022
*Software Engineer Intern* *Menlo Park, CA*

- Developed a data library for the Family of Apps that handles more than **30 billion requests/day** to enable cross-app features.
- Implemented the library end-to-end with full-stack development, completing the project **2 weeks ahead of schedule**.
- Reduced the number of GraphQL queries by **5x**, lowering iOS app start-up times and server costs.

**Educative** June 2020 – Mar. 2021
*Software Engineer* *Lahore, PK*

- Increased web traffic and course registrations by **15%** and **11%** respectively by implementing features such as saving, sharing, and sending feedback on the Educative Answers platform.
- Helped modernize the tech stack by transitioning from JavaScript to TypeScript and Next.js to improve both code quality and website performance.

## Projects

**Kahaani** | *Python, Django, DRF, JavaScript, React* Jan. 2021 – June 2021

- Built a first-of-its-kind Urdu audio-storytelling website for the country's largest private school network, integrating an audio player with a Text-to-Speech ML model that achieved **91%** accuracy.

**Tripaze** | *JavaScript, React, Firebase, Cypress* Jan. 2020 – May 2020

- Co-founded a tech startup and developed a website that connected over **1,000** tourists with trip organizers.

**Federated Machine Learning** | *Python, TensorFlow, Numpy, Matplotlib* Jan. 2020 – May 2020

- Reduced ML training time by **4x** on low-end devices by splitting models into smaller subsets with parameter aggregation, and prepared a research paper on privacy-focused decentralized training.