

Abheesht Roy

SF, CA | (602) 807-8352 | abheeshtr11@gmail.com | [LinkedIn](#) | github.com/abheeshstroy

EDUCATION

Arizona State University | *Master of Science, Computer Science (GPA: 4.00)* | Tempe, Arizona, USA

Aug 2024 - May 2026

Manipal Institute of Technology | *BTech, Information Technology (GPA: 8.12/10)* | Manipal, India

Jul 2019 - Jul 2023

WORK EXPERIENCE

[Agent-Techs AI](#) | *Software Engineer Intern - Applied AI (Graduate internship)*

Jun 2025 - Aug 2025

- Architected and deployed a production LLM-driven data harmonization pipeline (FAISS, vector DBs, AWS S3 + Lambda) with a hybrid entity resolution system (field-level matching, dense embeddings, cosine similarity), achieving 95%+ accuracy across 3 domains and eliminating 70% of manual data cleanup. [\[GitHub\]](#)
- Shipped a modular multi-agent orchestration platform (LangChain, Ollama, GCP Cloud Run) with fault isolation, stateless scaling, and zero-downtime deployment, cutting cross-functional data integration turnaround by 60%.

[DeskNow GmbH](#) | *Software Engineer*

Sep 2023 - Jul 2024

- Owned full-stack development of a workspace booking and analytics platform (Node.js, Express.js, MongoDB, Angular, TypeScript, Next.js), integrated a Next.js micro-frontend that improved page load performance by 40% and deployed dashboards for 20+ enterprise clients.
- Reduced CI/CD build failures by 40% by implementing a GitLab pipeline with automated Docker builds, test gating (Jest), and environment-specific deployment configs.
- Optimized database schemas (Mongoose, PostgreSQL) and containerized deployments; reduced API response time ~25% via query optimization and connection pooling, diagnosed and resolved a latency regression affecting 3 enterprise clients within 4 hours via Azure Monitor.

[Samsung Semiconductor India Research](#) | *Software Intern*

Jan 2023 - Jul 2023

- Built a GUI-based diagnostic tool (Python, PyQt, C++) adopted across 3+ chip variants as the team's primary configuration interface, cutting chip setup time from days to hours and reducing misconfiguration errors by 50%.
- Diagnosed and resolved a real-time UI reliability issue caused by thread contention in the polling loop, applied a mutex-based fix and validated stability under sustained load with zero dropped register reads.
- Implemented I2C/SPI hardware interfacing and dynamic XML-driven clock configuration, enabling engineers to read/write registers and inspect waveforms live without recompiling or manual scripting.

[Atthah Infomedia Pvt Ltd](#) | *Software Intern*

May 2022 - Jul 2022

- Sustained 500+ concurrent users at sub-200ms latency, built chat, video conferencing, and polling modules for a Virtual Convention Center using MERN stack, Socket.io, WebRTC, and MongoDB with multi-room WebSocket management.
- Designed secure multi-room architecture with JWT-based authentication and RESTful APIs, supporting isolated session management across concurrent virtual spaces with zero auth bypass incidents
- Reduced live event feature latency ~10× (seconds → sub-200ms) by profiling and optimizing Socket.io event handling and WebSocket state synchronization under peak load.

[National Council of Science Museums](#) | *Research Intern*

May 2021 - Aug 2021

- Built Muso-Assist, an IoT-enabled chatbot (TensorFlow NLP, NodeMCU) deployed across museum galleries, boosted visitor engagement and led to two IEEE conference publications.
- Developed a real-time crowd counting system using CNN density estimation on existing CCTV feeds, avoiding expensive new hardware and cutting costs by 60–70%.

PROJECTS

Context Monitoring Android App [\[GitHub\]](#)

- Optimized Kotlin Android app on a Linux-based OS: captured health sensor streams via a foreground service, persisted data in SQLite, implemented battery-aware polling and WorkManager scheduling for 24/7 reliability

Text2SQL

- Built a multi-stage Text-to-SQL validation system (syntax + KG semantic check + TRC safety) improving execution accuracy by 7.8% over baseline FLAN-T5 and rejecting 100% of adversarial DROP/DELETE prompts on the Spider benchmark.

Scalable Graph Data Pipeline

- Deployed a two-phase containerized data pipeline, local Docker/Neo4j graph analytics in Phase 1, distributed real-time ingestion via Kafka and Minikube Kubernetes in Phase 2, with PageRank and BFS implemented via the Neo4j GDS library.

TECHNICAL SKILLS

- Programming Languages:** Python, JavaScript, C++, Java, TypeScript, Kotlin, SQL
- Machine Learning & AI:** LangChain, FAISS, PyTorch, TensorFlow, Prompt Engineering, Agent Orchestration, Vector Databases, RAG
- Web & Backend:** Node.js, Express.js, React, Angular, REST APIs, WebSockets, PostgreSQL, MongoDB, MySQL
- Systems & Infrastructure:** Linux/*nix, TCP/IP Networking, Docker, Load Balancing, CDNs, Thread Concurrency, I2C/SPI
- Cloud & DevOps:** AWS, S3, Lambda, Azure, Azure Monitor, GCP, Cloud Run, GitLab CI/CD, Kubernetes
- System Design & Architecture:** Design Patterns, OOP, DS & Algorithms, Capacity Planning, Distributed Systems, System Design

RESEARCH AND PUBLICATION

- Shatadal Ghosh; Abheesht Roy; Sriparna Saha. A Low-Cost Virtual Guide for Next Generation Smart Museum.. *6th IEEE International Conference (IEEE-RTEICT 2021)*; Link: <https://ieeexplore.ieee.org/document/9573753>
- Shatadal Ghosh; Abheesht Roy; et al.. Interactive Communication Robot Handling Crowd Management & Content Delivery in Museums Employing Crowd Counting.. *2nd IEEE-INCON 2023*; Link: <https://ieeexplore.ieee.org/document/10101373>