Gishnu Madhu

New York | 7162914219 | gishnum.work@gmail.com | Linkedin | Github | Portfolio

Education

University at Buffalo — Buffalo, NY
Master of Science in Computer Science and Engineering — *Aug. 2024 – May 2026*

Experience

AI Engineer — Jun 2025 – May 2026
PhysicianX — *Buffalo, NY*

- Architected an autonomous web crawler using Temporal for durable workflow orchestration, integrating backend distributed systems to execute 3 million daily activities and ingest data into a scalable Knowledge Graph.
- Developed a full-stack RAG-AI assistant equipped with RBAC data controls and OAuth/JWT authentication, served via a high-performance GraphQL and RESTful API to reduce internal document retrieval time by 20%.
- Deployed multi-agent conversational systems using LiveKit on internal GCP infrastructure utilizing cloud computing and microservices architecture to facilitate private, automating 100+ weekly interview simulations and saving 50 hours of manual screening time..
- Designed an SEO-compliant market data aggregation frontend in React, building a responsive Single Page Application (SPA) to funnel user engagement and accelerated customer acquisition by 25%.
- Constructed a CI/CD pipeline framework utilizing Git version control and containerization that reduced deployment latency by 75%, while triaging production incidents and driving technical direction across the full stack.

Graduate Researcher — Jan 2025 – May 2026
University at Buffalo — *Buffalo, NY*

- Architected Edge Field Graph Network (EFGN), which enables guidance to heuristic algorithms' navigational capabilities in mixed integer non-linear environments.
- Applied developed technique on satellite debris capture optimization problem, reducing fuel consumption by 80% via advanced data modeling.
- Applied Reinforcement Learning (DDPG & PPO) for GPS-denied drone-to-drone flight, building a virtual simulation environment in Unreal Engine that achieved a 90% target tracking accuracy.

Staff Software Engineer — Sep 2019 – July 2024
Credit Suisse AG — *Pune, India*

- Led zero-downtime production migrations for Neo4j (v3 to v5) and Java (8 to 17) across an 18-service microservice architecture deployed on OpenShift and Kubernetes.
- Optimized distributed ETL processes and complex Relational Database (RDBMS) and NoSQL queries, reducing latency and increasing overall data pipeline scalability and efficiency by 80%.
- Served as the SME for Neo4j graph modeling and Cypher optimization, architecting distributed systems to synchronize intelligence across 20 global locations cutting global data sync latency by 40%.
- Architected a scalable NLP ingestion engine using CoreNLP and Object-Oriented Programming (OOP), parsing 1 GB of unstructured market reports into a knowledge graph to accelerate automated risk insight generation by 10 hours per week.
- Developed event-driven Java Spring Boot microservices utilizing Kafka with partitioned, idempotent consumers, building resilient backend REST APIs to resolve historical SLA breaches and maintained 99.9% workflow uptime.
- Enforced high-availability and testability standards for production launches by leading Agile/Scrum evaluations, rigorous code reviews, and Unit Testing/Test-Driven Development (TDD) practices.

Technical Skills

Languages: Java, Python, TypeScript, JavaScript, C++, SQL, Cypher, GQL, HTML5/CSS3
Frameworks & Backend: Spring Boot, FastAPI, React, Next.js, Kafka, Temporal, Node.js, RESTful APIs, GraphQL
Databases & Storage: Neo4j, MySQL, PostgreSQL, MinIO, Supabase, Vector Databases, NoSQL, RDBMS, Redis
Cloud & DevOps: GCP, AWS, Docker, Kubernetes, OpenShift, Spark/PySpark, CI/CD, Git
Architecture & Systems: Microservices, Event-Driven Architecture, Distributed Systems, Multi-Tenant Systems, Query Optimization
AI & Machine Learning: GraphRAG, LangChain, Mastra, PyTorch, LLM SDKs (OpenAI/Gemini/Claude), Knowledge Graphs

Publications

- Madhu, G., Liu, F., Chowdhury, S. "Learning-Based Directed Graph Abstraction in Combinatorial Spaces for Order Preserving Search in Mixed Combinatorial Nonlinear Optimization." *ASME 2026 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC-CIE 2026), Houston, TX, Aug 23-26, 2026.* (Accepted)
- Liu, F., Boonrath, A., Madhu, G., Botta, EM., Chowdhury, S. "Designing Active Tether-Net Systems for Space Debris Capture with Graph Learning Aided Mixed Combinatorial Optimization." *AIAA AVIATION 2026, San Diego, CA, Jun 8-12, 2026.*

Projects

- **Malayalam Small Language Model(link)**: Pre-trained and fine-tuned a small language model on a native Malayalam corpus.
- **BlazegraphJVM Graph Database(link)**: Implemented an in-memory, GQL-compliant graph database in pure Java, including a custom ANTLR-based query parser and query planner, used for ephemeral analytics, zero-dependency embedded processing, and fast integration testing in CI/CD environments where spinning up Docker bogs down the deployment pipeline.
- **GraphRAG Document Retriever(link)**: Built a GraphRAG pipeline on Neo4j Aura combining GLiNER-based entity extraction with LLM-driven query routing, grounding retrieval in graph structure to enable accurate, cost-effective in-context reasoning for downstream agents.
- **Automated Knowledge Graph(link)**: Building a self-learning knowledge graph in Python/Neo4j with an evolving RDF ontology; given only seed details, the system autonomously extends the ontology and underlying data graph over time.
- **Graph Learning Framework for Active Tether-Net Systems**: Engineered scalable data structures and custom graph-mining algorithms to handle the high-dimensional combinatorial optimization space underlying the active tether-net co-design problem (companion implementation to the AIAA AVIATION 2026 publication).