

# Muhammad Muzzammil

[muzzabyte@gmail.com](mailto:muzzabyte@gmail.com) ❖ (312) 889-4606 ❖ Chicago, IL ❖ [Github](#) ❖ [LinkedIn](#) ❖ Eligible to work in Canada and U.S.

---

## EDUCATION

Georgia Institute of Technology — *MS. in Machine Learning*

Atlanta, GA | August 2025 – May 2027 | GPA: 4.0

University of Illinois at Chicago — *B.S. in Computer Science*

Chicago, IL | August 2021 – May 2024 | GPA: 3.62

## TECHNICAL SKILLS

- **Languages:** Python, TypeScript, JavaScript, SQL, C++
- **Machine Learning & AI:** Generative Models (GANs, VAEs, Diffusion), RAG Pipelines, Vector Databases, Fine-Tuning (LLMs & Vision Models), Model Deployment & Optimization
- **Frameworks & Libraries:** PyTorch, TensorFlow, Transformers, FastAPI, LangChain, scikit-learn, vLLM, NumPy, Pandas
- **Cloud & MLOps:** AWS (S3, Lambda, EC2), Docker, CI/CD, MLflow, Weights & Biases
- **Backend & Web:** Node.js, Django, REST APIs, React
- **Data Engineering:** Spark, ETL Pipelines, Data Warehousing, Structured/Unstructured Data Processing
- **Tools:** Git, Databricks, Linux, Jupyter, Redis

## WORK EXPERIENCE

Founding Machine Learning Engineer

May 2025 – Present

LeadEdge AI LLC

Chicago, IL

- Built autonomous AI-driven lead generation systems that continuously ingest, enrich, and qualify prospects using machine learning, NLP, and multi-source data pipelines to deliver CRM-ready, high-intent leads.
- Designed and implemented end-to-end full-stack infrastructure, including scalable ETL workflows, backend APIs, databases, and a modern web platform for discovering, filtering, and purchasing leads.
- Developed predictive lead-scoring and entity extraction models leveraging real-world engagement signals to automate prospect prioritization and improve lead relevance and conversion outcomes.

Machine Learning Engineer

August 2025 – November 2025

Georgia Institute of Technology

Atlanta, GA

- Collaborated with a team of four researchers on *DeepCura*, an AI assistant for dental clinicians that analyzes facial features, teeth conditions, and patient history to generate personalized treatment plans.
- Contributed to the design and training of deep learning models for anatomical and dental structure analysis, integrating them with a recommendation system for treatment and material selection.
- Assisted in system architecture planning, data pipeline development, and deployment workflows to support scalability and clinical reliability.

Software Engineer (Contract)

July 2024 – October 2024

LeadFuze

Phoenix, AZ

- Developed and deployed a high-performance Go-based billing API with Stripe integration (recurring + one-time payments), reducing payment discrepancies by 30% and improving system efficiency by 20%.
- Automated subscription lifecycles via Stripe webhooks, cutting billing errors by 25%, and optimized backend design to reduce API latency by 25%.
- Built CI/CD pipelines using Docker, Kubernetes, and GitHub Actions, improving deployment speed by 50% and ensuring scalable, reliable production releases.

## PERSONAL PROJECTS

- **Neural Generators Project** (*Python, PyTorch, Deep Learning, Computer Vision*)
  - Developed GAN, VAE, and Diffusion model architectures from scratch using PyTorch, implementing custom forward passes, loss functions, optimizers, and noise schedules, reducing framework overhead by ~20% compared to baseline tutorial implementations.
  - Implemented full generative training workflows, including latent-space reparameterization,

discriminator–generator adversarial loops, and noise-prediction denoising steps for diffusion models, achieving stable training with >95% loss convergence consistency across runs.

- Achieved high-quality MNIST synthesis, generating samples with over 92% classifier agreement accuracy, and improved diffusion denoising performance with a 30–40% reduction in reconstruction error, supported by real-time training visualizations using torchvision.

- **GraphMatch Pro** (*Apache Hadoop, AWS, Scala, MapReduce*)

- Designed a distributed graph-processing pipeline on Hadoop to perform large-scale similarity and connectivity analysis across millions of graph nodes and edges, enabling scalable computation on datasets too large for single-machine processing.
- Optimized MapReduce jobs by restructuring mappers/reducers, improving shuffle efficiency, and minimizing data skew, resulting in a 40% increase in processing throughput and a 60% reduction in end-to-end runtime for graph score computation.
- Deployed, tuned, and monitored the full workflow on AWS EMR, leveraging auto-scaling and instance grouping to maintain 99% job reliability and reduce cluster operating costs by ~25% during peak workloads.

- **Movie Orchestration Service** (*React, Golang, PostgreSQL, JWT, GraphQL, Docker, Kubernetes*)

- Built a scalable full-stack web application with secure JWT-based authentication, role-based access control, and complete CRUD operations for managing movie metadata, improving request handling throughput by ~40%.
- Integrated GraphQL APIs to streamline complex queries and reduce over-fetching, cutting data retrieval overhead by 35% and improving client-side load times by ~25%.
- Containerized the entire system using Docker and deployed on a Kubernetes cluster, configuring pods, nodes, and autoscaling policies to maintain high availability (>99% uptime) and support 2–3× traffic surges without performance degradation.