

Maithil Rupesh Mehta

maithil.mehta@wisc.edu | +1 (608) 509-3112 | LinkedIn | Github

EDUCATION

- **University of Wisconsin-Madison | CGPA: 4.00/4.00** Madison, Wisconsin
• **Master of Science in Computer Science** September 2025 – May 2027
Relevant Coursework: Machine Learning, Foundational Models, Distributed Systems, Big Data
Teaching Assistant: Data Management for Data Science
- **Birla Institute of Technology and Science, Pilani | CGPA: 8.29/10.00** Pilani, India
• **Bachelor of Engineering in Computer Science, Minor in Data Science** August 2018 – May 2022
Relevant Coursework: Data Structures and Algorithms, Foundations of Data Science, Information Retrieval, Graph Mining

TECHNICAL SKILLS

- **Languages:** Python, Java, JavaScript, TypeScript, SQL, C++, C, Ruby, HTML, CSS
- **Machine Learning & Data Science:** Pytorch, TensorFlow, Keras, Scikit-learn, Pandas, NumPy, Matplotlib, OpenCV
- **Generative AI & LLMs:** HuggingFace Transformers, PEFT, RAG, Quantization, Diffusion Models, Prompt Engineering
- **Big Data & Distributed Systems:** Spark, PyArrow, HBase, ChromaDB, Qdrant, Cassandra, Kafka, MySQL, MongoDB
- **DevOps & Tools:** Git, Docker, Kubernetes, Linux, Bash/Shell scripting, Node.js, gRPC
- **Software development practices:** Agile, REST APIs, CI/CD, Unit Testing

WORK EXPERIENCE

- **MathWorks** Natick, MA
• *Software Engineering Intern* May 2026 – August 2026
 - Built an autonomous **LLM agent** for a **50TB** infrastructure data warehouse, using **MCP tool-calling** and schema-driven APIs to translate natural language into validated queries across **45+ endpoints**.
 - Engineered the agent's reasoning framework for multi-step planning, parameter extraction, sequential/parallel API workflows, with **caching, validation, and enterprise SSO** for safe data access.
- **The D. E. Shaw Group** Hyderabad, India
• *Senior Member Technical, Investor Relations* July 2022 – June 2025
 - Architected a **data platform** with temporal modeling (**Python, SQL**) over **100M+ records**, delivering **2.5×** **growth** in active platform usage through optimized query pipelines and automated validation workflows.
 - Shipped a **React** full-stack **Language-to-Visualization** tool used by **100+ stakeholders**, enabling natural language-driven analytical plot generation to eliminate dependency on analyst intermediaries for ad-hoc reporting.
 - Reduced **data validation** turnaround from **days to hours** by engineering automated anomaly detection and reconciliation pipelines across multiple upstream data feeds.
- **Amazon** Hyderabad, India
• *Software Development Engineering Intern, Finance Automation* February 2022 – June 2022
 - Reduced manual finance review effort by **35%** by building **JavaScript** and **Ruby on Rails** tooling that automatically surfaced edge-case discrepancies in vendor invoice workflows across high-volume transaction data.
 - Cut operational maintenance costs by **25%** by leading migration of critical finance APIs from on-premises **Java** infrastructure to native **AWS**, improving reliability and reducing deployment overhead.
- **Salesforce** Hyderabad, India
• *Software Engineering Intern, Big Data Cloud* June 2021 – July 2021
 - Slashed incident resolution time by **30%** by building a multithreaded **Python** library for real-time deployment metric tracking across distributed data centers, with integrated alerting and anomaly detection.
 - **Won** the company-wide intern **hackathon** (competing against 200+ interns) by engineering a computer-vision based 2FA liveness verification system using **OpenCV** and **TensorFlow**.

PROJECTS

- **TALMAS: Training-Free Attention Suppression for Diffusion LMs:**
 - Invented a training-free inference intervention for **Masked Diffusion Models** that suppresses [MASK] token noise via dynamic, asymmetric logit biasing improving generation coherence on **GSM8K** and **HumanEval** with negligible $O(n)$ overhead and no loss in parallel decoding speed.
- **Course-Specific Exam Assistant - Fine-Tuning vs. RAG Pipeline:**
 - Built and benchmarked a hybrid **RAG pipeline** (**Elasticsearch** for lexical retrieval + **ChromaDB** for semantic search) against a **LoRA** fine-tuned Llama-3.2 model as a TA-led exam prep tool for **150 students**.
 - Quantitative analysis showed Hybrid RAG reduced hallucinations by a statistically significant margin for transcript-based Q&A, informing deployment decision.