

Amogh Sudhir Dixit

amogh.dixit@wisc.edu ◇  amoghdixit2010 ◇  fluorspar20

Education

University of Wisconsin–Madison (UW–Madison)

Sept. 2025 – May 2027

Master of Science (MS) in Computer Science

Overall GPA: 4.0/4.0

Key Courses: Advanced Reinforcement Learning, High Performance Computing, Computer Vision, Foundation Models, Statistical Learning Theory, Advanced Operating Systems

Teaching Assistant: Machine Organization and Programming

Indian Institute of Technology Indore (IIT Indore)

July 2019 – May 2023

Bachelor of Technology (B.Tech.) in Electrical & Computer Engineering

Overall GPA: 8.76/10

Key Courses: Probability & Random Processes, Digital Signal Processing, Linear Algebra, Calculus

Technical Skills

Languages: Python, Java, C/C++, TypeScript, SQL, HTML

CS Fundamentals: Data Structures, Algorithms, Operating Systems, Database Systems, Distributed Systems

Tools & Frameworks: ReactJS, Angular, NodeJS, Spring Boot, FastAPI, GraphQL, OpenMP, MySQL, DynamoDB, Redis, Django, Docker, Azure, Terraform, Kubernetes, Linux, Jenkins, GitHub Actions

Machine Learning & Data Science: PyTorch, TensorFlow, LangChain, LangGraph, Autogen, NumPy, Pandas

ML/AI Competencies: Transformers, Diffusion Models, Reinforcement Learning, Gaussian Splatting, LLMs, GPU Programming, Information Retrieval, RAG, Agentic AI

Software Development Practices: Agile, REST APIs, CI/CD, Unit Testing

Work Experience

Software Development Engineer Intern, Amazon.com

Seattle, WA | May 2026 – Aug. 2026

- Built an oncall investigation agent in **Python** that ingests oncall alarm tickets, correlates evidence from **AWS CloudWatch** and data from **Spark** processors running on **EMR** via an **MCP** server and recommends mitigations
- Achieved **88%** recommendation accuracy and reduced diagnosis time from 30–60 min to under **10 min**

Software Engineer II, UnitedHealth Group

Hyderabad, India | July 2023 – Aug. 2025

- Developed Coding Intelligence, a real-time medical coding platform with high-performance search and retrieval of medical codes and associated data, using **Angular** and **Spring Boot microservices**
- Led **Azure** migration of Simba, a **Java**-based PDF engine processing critical CMS medical data
- Reduced **Terraform** upgrade effort **90%** by building an **Autogen/LlamaIndex** AI agent for Azure infrastructure
- Built a real-time summarization pipeline for 7,000+ AHA articles using GPT-4o mini with Azure Cognitive Search, dense embeddings, and hybrid retrieval
- Migrated search APIs to **Apache Solr**, cutting latency to **300ms** and indexing time **15%** via multithreading
- Delivered Java features with **Mockito/JMeter** testing while managing **Docker, Terraform, and Azure** pipelines

Undergraduate Researcher, Indian Institute of Technology Indore

Indore, India | Aug. 2022 – Mar. 2023

Supervisor: *Dr. Nitya Tiwari*

DOI: 10.1007/s00034-025-03036-6

- Engineered novel noise-aware **variational autoencoders** for **self-supervised** single-channel speech enhancement
- Implemented a **PyTorch dual-VAE** with **CNN layers** with a cyclic loss to optimize latent space for robust denoising
- Outperformed SOTA baselines by **5%** in NISQA/STOI via noise-aware training to maximize speech quality

Research Intern, University College London

London, United Kingdom | May 2022 – Oct. 2022

Supervisors: *Dr. Philip Treleaven, Dr. Hirsh Pithadia*

DOI: 10.2139/ssrn.4475767

- Engineered a **decentralized** VC tokenization platform on **Polygon** enabling startup listing and investor funding
- Developed Solidity contracts and tokenomics, implementing bonding curves for automated liquidity and fee structures
- Built a **React** frontend with wallet auth, integrating smart contracts and IPFS via Hardhat and Ethers.js

Selected Projects

Pedestrian Intent Prediction for Autonomous Vehicles

[GitHub](#)

- Implemented offline **CLIP knowledge distillation** pipeline with trainable MLP with combined cross-entropy+MSE loss; **Grad-CAM++** analysis revealed partial pedestrian attention shift and background reinforcement post-distillation
- Designed spatial masking and scale ablations proving scene-reliance of EfficientPIE (IJCAI 2025) over pedestrian body language (F1: 0.064 vs 0.457) and identified resulting overconfidence failure at 76.5% accuracy on closest pedestrians

Timestep-Adaptive Attention Suppression on DLMs

[GitHub](#)

- Designed TALMAS, a **training-free** attention intervention on **LLaDA-8B** that suppresses masked-token key attention via adaptive logit biasing, improving GSM8K accuracy from **71%** to **73%**
- Proposed PPML-inspired **dependency-aware unmasking** for **masked diffusion models**, using cross-attention weights to identify load-bearing masked positions and replace the standard greedy confidence-first unmasking schedule