

# Pranav Mishra

773-280-4615 | pmishr23@uic.edu | LinkedIn/pranavgamedev | Github/PranavMishra17 | Portfolio | Metuchen, NJ

## Education

<b>Master of Science, Computer Science [ Graduate Assistant ]</b> University of Illinois at Chicago, Illinois, USA	<b>Aug 2023 - May 2025</b> GPA: 3.6/4.0
<b>Bachelor of Science, Computer Science and Engineering</b> Dayananda Sagar College of Engineering, Bangalore, Karnataka, India	<b>Aug 2019 - June 2023</b> GPA: 4.0/4.0

## Technical Experience

<b>AI/ML Intern</b>   <i>WheelPrice</i> ↗   Charlotte, NC	<b>July 2025 - Jan 2026</b>
<ul style="list-style-type: none"><li>Building end-to-end ML prototype for automotive part fitment prediction using PyTorch and computer vision models.</li><li>Deployed production-grade CMS blog system using React TypeScript with Node.js backend, MongoDB database, and RESTful APIs, scaling daily webapp viewership by 10-20k through enhanced content delivery and SEO implementation.</li></ul>	
<b>Research Software Engineer</b>   <i>UIC- VARE Lab</i> ↗   Chicago, IL	<b>Feb 2024 - May 2026*</b>
<ul style="list-style-type: none"><li>Built virtual patient system using Unreal Engine and C++ . Deployed REST APIs with Python backend for data analysis ↗</li><li>Deployed end-to-end audio ML pipeline with Flask backend comparing MFCC feature engineering, CNN architectures, and Transformer models achieving 98.52% accuracy; deployed real-time inference system with VAD &amp; robustness testing. ↗</li></ul>	

## Research & Publications

<b>TeamMedAgents: Medical Decision Making with SLMs through Teamwork</b>	<a href="#">GitHub</a>   <a href="#">Paper</a>   <i>SLMs/v2.0</i>
<ul style="list-style-type: none"><li>Engineered modular multi-agent system using Google ADK with prompt engineering for medical QA across 8 benchmarks.</li><li>Achieved 77.63% accuracy on LLMs and 3.1x inference speedup on 4B models through collaborative deliberation, trust-weighted voting, and structured multi-turn reasoning with comprehensive evaluation frameworks.</li></ul>	
<b>MetaRAG: A Systematic Framework for Enterprise Knowledge Retrieval:</b>	<a href="#">GitHub</a>   <a href="#">Paper</a>
<ul style="list-style-type: none"><li>Architected end-to-end metadata enrichment pipeline for RAG systems using LangChain, vector databases with Pinecone.</li><li>Engineered hybrid retrieval system with reranking and cross-encoder models, achieving 92.5% precision and 25% hallucination reduction; deployed on Azure with automated CI/CD pipelines and comprehensive evaluation metrics.</li></ul>	

## Projects

<b>MockFlow-AI: Real-Time Voice Interview Platform with Multi-Agent Architecture</b>	<a href="#">GitHub</a>   <a href="#">Live Website</a>
<ul style="list-style-type: none"><li>Engineered real-time voice AI system with sub-400ms response latency using microservices FSM architecture, streaming STT-LLM-TTS pipelines, and distributed session management for production ready conversational multi-agents.</li><li>Shipped end-to-end AI features including feedback loop, RAG-based context injection, and cloud-native deployment.</li></ul>	
<b>SnakeAI-MLOps: Multi-Agent Reinforcement Learning Snake Game</b>	<a href="#">GitHub</a>   <a href="#">Demo</a>
<ul style="list-style-type: none"><li>Engineered RL training framework in C++ with LibTorch, PyTorch integration, and CUDA GPU acceleration.</li><li>Built production MLOps pipeline with model comparison tools, Docker CI/CD, and real-time gameplay integration.</li></ul>	
<b>MedRAG Avatar Platform - Intelligent Healthcare Conversational AI System</b>	<a href="#">GitHub</a>   <a href="#">Demo</a>
<ul style="list-style-type: none"><li>Architected and deployed a production RAG system using Azure OpenAI, Cosmos DB vector search, and Pinecone for semantic document retrieval, enabling context-aware conversational avatars with 95% knowledge accuracy.</li><li>Built FastAPI backend with ETL pipelines supporting source attribution and inline citations for avatar responses</li></ul>	

## Skills

**TECHNICAL SKILLS:** Python, C++, JavaScript, C#, TensorFlow, PyTorch, MatLab, Pandas, LangChain, React, NumPy, Node.js, Flask, Next.js, Django, FastAPI, PostgreSQL, MongoDB, Docker, Git, GCP, Azure, AWS, DevOps, REST APIs

**APPLICATIONS:** Machine Learning, Deep Learning, NLP, Computer Vision, Neural Networks, Transformers, Data Science, MLOps, Cloud Computing, Agile Development, Linux, Fine tuning, Data structures, Mathematics, Version Control

## Extracurricular

**Winner of MIT XR Hackathon** | Built Meta Quest 3 app using Unity and Hugging Face for spatial data visualization ↗  
**INFORMS Analytics+** Presented MetaRAG research to 700+ professionals | First place HINT 5.0 Web3 virtual museum ↗  
Explore my portfolio for 25+ innovative projects spanning game design, AI/ML research, and in-development work ↗