

VARUN MATHUR

9044789900 | varunmathur1990@gmail.com | [LinkedIn](#) | [Github](#)

WORK EXPERIENCE

ACCENTURE

AI/ML Computational Science Senior Analyst

Bangalore, India | Nov 2025 - Present

- Architected and scaled an enterprise AI platform, implementing conversational memory, intelligent suggestion services, query correction, and prompt orchestration, reducing response latency by 40% and production defects by 57%.
- Designed and optimized a large-scale semantic search system indexing 167K+ enterprise documents using vector embeddings and NLP pipelines, improving search precision from 30% to 87%.
- Developed and deployed a generative AI analytics service for automated earnings-report processing, reducing analysis turnaround time by 93% through scalable backend workflows and LLM-powered pipelines.
- Built and delivered reusable AI-powered applications and services adopted across internal and client-facing products, significantly improving knowledge discovery and operational efficiency.
- Reengineered a production RAG platform using Azure Functions, Python, and Azure Blob Storage, improving system reliability by 35% and eliminating service downtime through resilient cloud-native architecture.
- Collaborated across product, engineering, and business teams to deliver scalable AI capabilities from design through production deployment.

AI/ML Computational Science Analyst

Bangalore, India | Sept 2023 – Nov 2025

- Migrated a production generative AI application to the Haystack framework, improving response accuracy from 66% to 92% while enhancing maintainability and deployment reliability.
- Built backend services integrating LLMs, retrieval systems, and external tools through function-calling architectures to support production AI applications.
- Reduced incorrect and hallucinated responses by 78% through systematic prompt optimization, retrieval improvements, evaluation frameworks, and tool integration.
- Presented technical architecture and implementation strategy for generative AI solutions to senior leadership, driving investment and delivery of 5 production-grade AI assets.

INTERNSHIP

Samsung R&D

Computer Vision Engineer

Bangalore, India | May 2021 – Feb 2022

- Trained GANs written with PyTorch to overlay realistic reflections on images, improving dataset robustness under glare conditions by 35%.
- Applied GAN-generated synthetic reflections to augment computer vision datasets, enhancing the accuracy of glare removal by up to 21%.

LEADERSHIP

AI Architect and Team Leader

- Mentored cross-functional teams of up to 19 engineers across 2 enterprise Generative AI initiatives.
- Launched 5 production AI assets; reduced dev cycle by 28% and improving cross-team delivery efficiency by 35%.
- Re-architected validation and testing workflows, raising accuracy to 95% and reducing release time by 30%.
- Spearheaded generative AI Proof of Concept to senior leadership, securing investment for dedicated asset development which led to the development of 5 assets.

SKILLS

- Languages: Python, Java, SQL
- Backend & Cloud: FastAPI, REST APIs, Microservices, Microsoft Azure, Azure Functions, Docker, CI/CD, Azure Blob Storage
- Databases & Software Engineering: PostgreSQL, SQL Server, Vector Databases, System Design, Distributed Systems, Scalability, Performance Optimization, Unit Testing, Git
- AI/ML & Frameworks: LLMs, RAG, Semantic Search, NLP, Embeddings, Prompt Engineering, LangChain, Haystack, CrewAI, OpenAI SDK, Multi-Agent Systems

EDUCATION

Vellore Institute of Technology

Bachelor of Technology in Electronics & Communications Engineering

Jul 2019 - Jul 2023

CGPA: 9.03

CERTIFICATIONS

- Microsoft Certified: AI Transformation Leader (AB-731)
- Claude Certified Architect – Foundations
- Claude Certified Architect – Early Adopter
- Developing Solutions for Microsoft Azure