

Academics

2018-2022 | B.Tech in Computer Science and Engineering, Presidency University
CGPA | 7.7/10.0

Experience

Present	Software Engineer - AI/ML: Microsoft COSMIC AI – Implemented multilayered remote MCP (Model Context Protocol) servers and client architecture with robust authentication and authorization, integrating Agent2Agent (A2A) protocol to enable interoperable communication between autonomous AI Agents and seamless orchestration of external services. – Finetuned LLMs for specific tasks, optimizing performance and reducing computational overhead through 16-bit and 8-bit quantization with fast evaluation. – Developed and deployed secure RAG -based REST Apps integrated with Azure OpenAI and Azure Cognitive Search for GenAI -powered responses, leveraging Azure AD, JWT, and Managed Identity for authentication, with centralized configuration management and comprehensive monitoring via Application Insights.	Tech Mahindra
Jan '24	Machine Learning Engineer: Evidence Extraction – Developed a Document Question Answering platform using Gen-AI to extract key evidence from audit and compliance reports, reducing manual effort by 60% and enhancing the accuracy of generated evidence. – Built and deployed a Question-Answering pipeline as a REST API endpoint using Django, leveraging Azure OpenAI for vectorization and GPT-4 for responses, and using pdfplumber and Azure Document Intelligence for data extraction from complex PDFs.	Freelance
Nov '23	Data Scientist: Customer Experience Intelligence – Developed scalable code to produce precise transcripts from medical audio data , achieving an accuracy exceeding 98% through OpenAI Whisper. – Seamlessly aligned transcriptions with audio files utilizing a large-scale neural network, yielding aligned transcription files suitable for speaker diarization, sentiment analysis, and language detection. – Implemented Speaker Diarization using NVIDIA NeMo, accurately mapping speakers to corresponding sentences. Additionally, ensured proper punctuation in the transcriptions for clarity and readability.	Kanini
	Data Scientist: RT Noise Cancellation – Conducted comprehensive training on diverse datasets to ensure the robustness and efficacy of the model. – Engineered Django backend infrastructure to streamline audio input processing and machine learning-based noise suppression, achieving a notable 60% enhancement in audio quality. Additionally, wrapped up the codebase as a Rest API endpoint for seamless integration with ML processes.	Kanini
	Data Scientist: Deloitte Neuroscience – Developed scalable code to analyze animations, fonts, images, word count, bullet points, and content from PowerPoint files, offering tailored improvement suggestions in collaboration with the Neuroscience Research team. – Utilized BERT for extractive summarization, OpenCV with OCR for text extraction from slide images, and Spacy and NLTK for Named Entity Recognition, improving presentation analysis efficiency by threefold and receiving positive feedback..	Kanini
Aug '22	Data Scientist: Deloitte HOAIML – Orchestrated automation of file uploading processes, facilitating seamless format conversion and restoration post-OCR processing, and elevated table detection accuracy from 60% to over 95% through innovative algorithms and improved documentation clarity.	Kanini
Feb '22	SDE Intern: Conversational Assistant – Led scalable backend development and seamless REST API integration with ReactJS, fine-tuned models for accurate user intent interpretation, enhancing responses.	Kanini

Technical Skills

Languages	Python C++ C SQL
Frameworks	Django Flask FastAPI
DBMS	MySQL PostgreSQL MongoDB
VectorDB	FAISS Pinecone ChromaDB
Cloud Services/VCS	AWS Azure GCP
Frontend	HTML CSS Javascript Typescript
LLMs	OpenAI Gemini Misrtal Phi llama Gemma Claude
Embedding	text-embedding-ada-003-large
RAG/Agentic	LangChain LangGraph LlamaIndex MCP A2A
Transformers	BERT T5
Python Packages	Pandas Numpy Matplotlib Scikit-learn OpenCV NLTK Spacy

Activities & Leadership

- **OpenSource Involvement:** Successfully contributed to Khargpur Winter of Code, IIT-KGP for two consecutive years 2018 and 2019. Participated in GirlScript Summer of Code (GSSoC) 2019.