

Amey Karan

☎ +91 79047 35718 · ✉ ameykaran2005@gmail.com · 🌐 portfolio · 🌐 ameykaran · 🌐 ameykaran

EDUCATION

International Institute of Information Technology
BTech & MS (by Research) in CSE | CGPA - 9.1

Hyderabad, TS
Oct '22 – Jul '27 (*Expected*)

EXPERIENCE

- MontyCloud (with SERC) | Research Assistant** May '24 – Present
- Developed a **novel multi-agent framework** - *MOYA* available as open-source package. Leveraged **LLMs** and used **Model Context Protocol** for tool access. Created various agents that specialise in particular tasks with dedicated data sources. Built an **autonomous CloudOps Copilot** to optimise **AWS accounts**. Solved various **AI and SE problems**, such as unstructured data, scalability and extensibility through RAG and hierarchical summarisation.
- Software Engineering Research Center | Undergraduate Researcher** May '24 – Present
- Researching in the intersection of Generative AI and Software Architecture. Currently exploring the possibility of extracting design decisions from source code using Large Language Models (LLMs).
- IIIT Hyderabad | Teaching Assistant** Jan '25 – May '25
- Assisted in conducting the *Introduction to Software Systems* course during Monsoon '25 by supporting tutorials and labs for hands-on learning. Helped create and evaluate assignments and exams.
- Student Alumni Connect Cell | Web Team Head** Nov '23 – Dec '24
- Led the development of the **Yearbook portal** to streamline the collection of data from graduating batch of **300+ students**. Coordinated with other teams to ensure seamless integration of features and a user-friendly interface. Designed the entire backend to handle different APIs and services.

PUBLICATIONS

- Engineering LLM Powered Multi-Agent Framework for Autonomous CloudOps** in International Conference on AI Engineering (CAIN) 2025– worked with different agentic frameworks and tools.
- DRAFT-ing Architectural Design Decisions using LLMs** in (work in progress) – exploring a novel mix of few-shot learning, RAG, and fine-tuning to enhance domain-specific decision-making.

PROJECTS

- LearnAI - Adaptive Agentic AI Assistant | Strands, RAG, OCR, FastAPI, NextJS** **GitHub**
- Architected a **self-adaptive document ingestion pipeline** using OCR and semantic chunking to extract, structure, and index content from textbooks, articles, and exam papers with rich metadata preservation.
 - Built a **multi-agent AI system** with five specialised agents powered by task-specific tools and a math fine-tuned local model. Implemented a **dynamic user profiling layer** to deliver personalised, learner-centric responses.
- DilLeiX - Multilingual SLM pretrained from scratch | Transformers, Python** **GitHub**
- Curated a multilingual text corpora of about **3B tokens** for **English, Hindi and Mizo** languages and implemented **Byte-level BPE tokeniser** for efficient multilingual tokenisation.
 - Pretrained a **150 M-parameter Deepseek-V3-like** model from scratch using **curriculum training** strategy and fine-tuned this SLM on **reasoning tasks** like humour detection and temporal reasoning.
- Legends of Stonks - Stock Market Learning Platform | Python, Docker, Nginx, PostgreSQL, NextJS, FastAPI**
- Created a working prototype of a stock market learning platform in a team of 5. Used service-oriented architecture to create **7 services** to manage different functionalities like user management, lessons and different levels. **Gamified** the platform to be more engaging and interactive. **Dockerised** the services and used **nginx** for reverse proxy.
- OCR using CRNNs | Python, PyTorch** **GitHub**
- Built and trained machine learning models for extracting text from images using **CNN encoder**. and **RNN as decoder** for auto-regressive modelling, achieving an accuracy of **77.16%**.
- Rudra's Subscription Service - Refactoring | Java, HTML/CSS, JavaScript** **GitHub**
- Worked on existing Sismics RSS Reader in a team of 5. Made **UML class diagrams** to identify important classes and their relationships. Used tools like **Designite** and **Sonarqube** to identify **design and code smells**. Derived **code metrics**, which helped in refactoring the code by implementing **design patterns**. Used LLM to perform **automated refactoring**.
- Tiny Tots - Baby Growth Tracker | ViteJS, User Research, Figma** **GitHub**
- Came up with a solution to help new mothers track their baby's growth effectively. Conducted **user research** to identify root cause and ideated on different solutions. Made **Figma wireframes** and built **prototype using ViteJS**.
- Recruitment Portal for IIIT Hyderabad | ReactJS, MongoDB, NodeJS** **GitHub**

- Built an end-to-end **online recruitment portal** for IIIT Hyderabad to streamline hiring process for staff and faculty with features like job posting, application submission, and interview scheduling. Used **CAS** for authentication.

AKonsole - Custom Shell | *C language, system calls, MakeFile, Linux IO redirection* **GitHub**

- Built a simple shell from scratch in C language with features like IO redirection, piping, and background processes. Implemented the shell using system calls and fork-exec model. It supports signal handling for handling signals.

Optimising Matrix Multiplication | *C, MakeFile, SIMD, CUDA* **GitHub**

- Used different GCC flags to optimise matrix multiplication. Implemented vectorisation to allow parallel processing utilising SIMD operations. Used multi-threading libraries like *omp*. Used CUDA for GPU optimisations. Achieved 100 times speed-up compared to baselines.

IndCity | *Android Studio, Java* **Google Play Store**

- Developed a wordle-like game for the Indian audience where players have to guess the name of the city based on the hints provided. The game was developed using Android Studio and Java and published on Google Play Store in Feb 22

TECHNICAL SKILLS

Languages: Python, Java, C/C++, JavaScript, TypeScript, Bash, HTML, CSS

Frameworks & Tools: NextJS, FastAPI, Docker, Nginx, AWS, PyTorch, Transformers, Linux (Archlinux)

Databases: PostgreSQL, MongoDB

Relevant Coursework: Distributed Systems; Software Engineering; Computer Vision; Language Models and Agents; Statistical Methods in AI; Operating Systems & Networks; Data Structures & Algorithms; Probability and Statistics; Linear Algebra

ACHIEVEMENTS

Dean's List: Top 10-20% of the batch in my second, fourth and fifth semester at IIIT Hyderabad

Research Award: Publicising at a A-tier conference during my undergraduate studies.

NTSE Scholar 2020: Among the 1000 students across the country to get selected for this fellowship

SQL Injection workshop: Took a session on SQL injection as part of the Hacking Club of IIIT Hyderabad

VOLUNTEERING

HackIIIT 2025 | *Organiser-cum-Judge* Feb '25

- Helped in conducting the *HackIIIT* Hackathon in collaboration with *MontyCloud*. Evaluated the submissions on their novelty and application of *MOYA* framework.

ICSA 2024 | *Student Volunteer* June '24

- Helped in organising the International Conference on Software Architecture '24. Responsible for ensuring the smooth transportation of attendees across various locations and managing the various technical equipment.

Fail? 2023 | *E-Cell, IIITH* Jul - Sep '23

- Core Organising Committee to conduct the *Fail?* talk held on the 9th Sep at IIIT Hyderabad. Responsibilities included cold-mailing and providing a warm experience to the speakers.