

Kashiful Haque

+91 82408 68544 | haque.kashiful7@gmail.com | [github/kashifulhaque](https://github.com/kashifulhaque) | [linkedin/kashifulhaque](https://www.linkedin.com/in/kashifulhaque) | ifkash.vercel.app

Software Engineer with 2 years of experience in Backend, Data Science, AI, ML and NLP. Expertise in fine-tuning LLMs and building RAG solutions. Studied Data Science at IIT Madras.

Work Experience

Fiery (formerly known as EFI) 07/2023 – Present
Associate Software Engineer, Data Science Bengaluru, India

- Built a solution to identify (watermarked) ads played on commercial stations like TV, etc.
- Worked on a solution using vector databases for similarity matching of images and text to enhance search experience.
- Build a Retrieval-Augmented Generation (RAG) based chat product using Mistral-7B LLM, trained on company data for Question and Answering.

Fiery (formerly known as EFI) 01/2023 – 07/2023
Data Scientist, Intern Bengaluru, India

- Improved efficiency by developing a pipeline to eliminate manual labour in text extraction.
- Created realistic conversational dataset to amplify accuracy of text classification model.
- Built cost-effective solution for creating product mockups using ImageMagick & Node.js/Python, surpassing results of expensive Photoshop APIs hence saving significant costs for the company.

Corteva AgriScience 07/2022 – 12/2022
Fullstack Developer, Intern Hyderabad, India

- Improved Flask codebase resulting in 30% faster AutoML job triggering on Kubernetes.
- Led migration of Flask monolith to scalable REST API microservices, reducing deployment time by 40%
- Worked under Corteva AgriScience incubation centre @ CIE, IITH

Education

Indian Institute of Technology Madras 12/2020 – 09/2024
Bachelor of Science, *Data Science and Applications*

University of Calcutta 08/2016 – 03/2019
Bachelor of Technology, *Electronics and Communications Engineering*

Projects

FieryGPT
Python, PyTorch, Ollama, llama.cpp, Mistral-7B, FastAPI, Node.js, Vue.js

- Engineered a retrieval-augmented generation LLM, integrating company data from vector databases for contextualized responses.
- Implemented innovative LLM architecture, optimizing information retrieval and generation for enhanced organizational knowledge utilization.

Skills

Language proficiency

- Python, JavaScript, TypeScript, C/C++, Rust

Web technologies

- Node.js, Express.js, Vue, Angular, FastAPI, Flask

Data Science & Machine Learning

- NumPy, Pandas, scikit-learn, Pytorch, HuggingFace