

SOHAM GHADGE

+919930227099 • sohamghadge0903@gmail.com • Maharashtra, India • [LinkedIn](#) • [GitHub](#)

ML and LLM engineer

Core Competencies

Programming Languages Python, C, C++, Java, HTML, CSS
Frameworks Flask, Django, FastAPI, LangChain, Jupyter
Libraries Pandas, OpenCV, Tensorflow, Sklearn, Keras, PyTorch, Llama.CPP
Databases MySQL, Oracle, PostGreSQL, MongoDB, ChromaDB, Weaviate DB
Version Control Git, Github, NodeJs
LLMs BERT, Llama, Gemma, Mistral, Qwen-VL, Pixtral
Cloud Platforms GCP, Azure, AzureOpenAI, OpenAI

Professional Experience

SuperKalam(YC W23), India

Aug 2024 - Present

AI Research Intern, Bengaluru

- Designed advanced prompt engineering techniques for evaluating and providing feedback on competitive examination responses (e.g., UPSC Mains) across 10+ subject domains, resulting in a 20% reduction in evaluation time and improving the accuracy of automated feedback.
- Fine-tuned Large Language Models (LLMs) specific to UPSC domains, optimizing parameters and employing Reinforcement Learning from Human Feedback (RLHF), which increased model accuracy by 15% and also increased the relevance of generated answers.
- Explored advanced Retrieval-Augmented Generation (RAG) pipelines, which improved the quality of answer generation by 25%, enhancing user satisfaction and reducing the need for manual corrections.
- Created an AI-based grading model for evaluating student answers, achieving a 30% improvement in grading accuracy compared to traditional methods.
- Developed custom scripts that systematized data preprocessing tasks in model training processes, minimizing manual input efforts by ten hours per week while ensuring cleaner datasets were used for analysis.

AllAdvanceAI, India

Feb 2024 - Jul 2024

AI and Backend intern, Mumbai

- Introduced a Python library for inference and management of Large Language Models (LLMs), accessible via Command Line Interface (CLI) and Web User Interface (Web UI), which reduced inference processing time by 30% and improved user accessibility.
- Implemented user management, version control, and workflow automation, which led to a 100% increase in the number of clients by improving system efficiency and user experience.
- Designed and implemented the entire backend of the company's product using FastAPI, resulting in a 40% improvement in system performance and scalability.

MSM-AI, India

Jun 2023 - Aug 2023

AI-LLM intern, Mumbai

- Optimized large language models for enhanced performance through strategic fine-tuning with PyTorch, increasing model accuracy by 20% and reducing training time by 25%.
- Diagnosed and resolved critical bugs in model training, resulting in a 15% reduction in error rates and improving response generation accuracy by 10% for the Python web application.

Education

University of Mumbai, Mumbai, India, Bachelor in Information Technology, (GPA: 9.41), Sep 2021 - Sep 2024

- Coursework:
Data Science, Computer Networks, Machine Learning, Operating Systems

Awards

Certifications

- Supervised Machine Learning: Regression and Classification, DeepLearning.AI
- Advanced Learning Algorithms, DeepLearning.AI
- Unsupervised Learning, Recommenders, Reinforcement Learning, DeepLearning.AI
- Machine Learning Specialisation, DeepLearning.AI
- Neural Networks and Deep Learning, DeepLearning.AI
- Structuring ML Projects, DeepLearning.AI
- Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization, DeepLearning.AI

Publications

- Ensemble Deep Neural Networks: An assessment for early Alzheimer's Disease diagnosis with structural MRI, International Conference on Image Processing and Capsule Networks

Jul 2024