

BHOOP SINGH GURJAR

☎ (+91) 7023337229 | ✉ bhoopsinghgurjar001@gmail.com |  [B h o o p S i n g h](#)

Machine Learning Engineer & AI Engineer with expertise in AI model optimization, cloud-based AI solutions, and scalable ML architectures. Proficient in Python, PyTorch, TensorFlow, FastAPI, LangChain, and Hugging Face Transformers, with hands-on experience in AutoGen, distributed computing, and LLM fine-tuning. Skilled in building retrieval-augmented generation (RAG) applications, AI agents, and autonomous systems. Experienced in databases (SQL, PostgreSQL, MongoDB) and cloud platforms (AWS, Colab). Passionate about advancing AI research, NLP innovations, and autonomous AI workflows to develop impactful solutions.

EDUCATION

Bachelor Of Technology
College of Technology
and engineering

Udaipur,rajasthan

CGPA- 6.6

July 2021 - 2025

Intermediate/+2

Ravindra Public school,jhunjhunu

R.B.S.C, Rajasthan

Percentage — **87.86 %**

2019

Matriculation

Ravindra Public school,
jhunjhunu

R.B.S.C, Rajasthan

Percentage — **88.89 %**

2017

SKILLS

Languages : Python

Database: SQL, MongoDB,
PostgreSQL

Framewroks: Django, Rest
Framework, Flask, Fastapi

Libraries: NumPy, Pandas,
Pytorch,Matplotlib,
Tensorflow,

Autogen,LangChain,
llamaIndex,JAX

Others:, Gitlab, HTML, CSS,
Docker, RestAPIs, Reactjs,
VsCode, AWS, Linux,Cursor,colab,

Language: English, Hindi

EXPERIENCE

Coding Pro

Machine learning Intern

Jun 2024 — Aug 2024

Intern

Unified Mentor

Machine learning Intern

Feb 2025 - Present

Intern

PROJECTS

LLaMa 2 Fine-Tuning

Sep 2024 - Dec 2024

Tech Stack : Python ,Hugging Face

Transformers,Pytorch,PEFT,TRL,Langchain, BitsAndBytes,W&B

- ❑ Designed and implemented a fine-tuning pipeline for LLaMA 2 using **parameter-efficient fine-tuning (PEFT)** techniques such as **LoRA** to optimize training efficiency.
- ❑ Leveraged **Hugging Face's** trl and transformers libraries to train a domain-specific language model with customized **prompts** and dataset **preprocessing**.
- ❑ Incorporated **LangChain** for dynamic **prompt engineering** and evaluation, enhancing model performance for downstream NLP tasks.
- ❑ Utilized BitsAndBytes for **quantization**, reducing memory footprint while maintaining high model accuracy.
- ❑ Integrated Weights & Biases (W&B) for **logging, monitoring**, and performance tracking to streamline experimentation and hyperparameter tuning.

Transformer Model Implementation

Feb 2024 - Mar 2024

Tech Stack : Python ,Pytorch, Distributed Data Parallel(DDP)

- ❑ Developed a **Transformer-based model from scratch** using PyTorch, implementing key components such as **self-attention**, positional encoding, and multi-head attention.
- ❑ Integrated **torchtext** for dataset preprocessing and vocabulary building, leveraging SpaCy for efficient **tokenization** and linguistic analysis.
- ❑ Designed and optimized training workflows with **Distributed Data Parallel (DDP)** to enable large-scale training across multiple GPUs.
- ❑ Utilized Altair for interactive visualization of training dynamics, including loss curves and attention weight distributions.
- ❑ Implemented efficient **learning rate scheduling** and optimizer strategies to enhance model convergence and performance.

POSITIONS OF RESPONSIBILITY

Author and Founder

ASI21 AI newsletter

Hand-on AI Paper newsletter

(Daily , weerkly AI cutting edge tech update)

ACHIEVEMENTS & CERTIFICATIONS

- ❑ Certificate of completion of the **Introduction To Transformer for Large lanuage Models** organized by The AiEdge.
- ❑ Certificate of completion of **NLP with end to end project**, Udemy.
- ❑ Reading **200 + AI** Research Paper
- ❑ 10+ **Research Paper** implement into **code**

AI Agent for Stock Report Jan 2025 – Mar 2025

Tech Stack : Python,AutoGen,GPT-4o,LLM APIs,Conversable Agents

- ❑ Developed an AI-powered autonomous agent using **autogen** to automate financial report generation based on real-time stock performance data.
- ❑ Designed a **multi-agent system** with specialized Conversable Agents (**Planner**, **Writer**, and **Admin**) to structure tasks, retrieve relevant data, and generate high-quality financial insights.
- ❑ Integrated **GPT-4o** for intelligent text generation, ensuring well-structured and insightful stock performance reports.
- ❑ Implemented dynamic **task planning** to refine and improve content iteratively based on predefined workflows and **user feedback**.
- ❑ Enhanced AI agent capabilities by incorporating context-aware **decision-making** and automated workflow execution for efficient financial analysis.

RAG Application

Mar 2025 - Current

Tech Stack : Python ,LangChain, Hugging Face Transformers,OpenAI API

- ❑ Developed a **RAG-based** AI system that enhances language model responses by dynamically retrieving relevant external knowledge from **vector databases**.
- ❑ Integrated **LangChain** for **seamless orchestration** of retrieval and generation workflows, ensuring high-quality and factually accurate AI-generated content.
- ❑ Optimized the model pipeline with Hugging Face Transformers and OpenAI API to improve response coherence and knowledge depth.
- ❑ Designed a scalable architecture for real-time retrieval and content augmentation, reducing AI hallucinations and enhancing response accuracy.