

# Muhammad Umer Bhutta

📍 Islamabad, 44000 | 📞 03354898079 | ✉ bhutta.umer65@gmail.com  
in [linkedin.com/in/bhutta-umer65](https://www.linkedin.com/in/bhutta-umer65) | 🐙 [github.com/umer-bhutta-1997](https://github.com/umer-bhutta-1997)

## Summary

Passionate NLP and AI enthusiast dedicated to leveraging technology for societal impact. Self-motivated, with a drive to innovate and improve language processing for a better world. Experienced in developing robust AI solutions, process automation, and custom ML/AI models to address unique business challenges.

## Work Experience

### Software Engineer

March 2024 - Present

*Autosphere.ai*

*Islamabad*

- Worked with software development and testing team members to design and develop robust AI solutions to meet client requirements for functionality, scalability, and performance
- Collaborated with cross-functional teams to identify opportunities for process automation using AI, driving operational efficiency
- Evaluated effectiveness of various machine learning algorithms, selecting best-suited models for specific tasks
- Developed custom ML/AI models to address unique business challenges, leading to improved decision-making capabilities
- Performed rigorous testing and validation of AI models to ensure robustness and reliability in real-world applications
- Maintained existing software systems by identifying and correcting software defects
- Developed scalable and maintainable code, ensuring long-term stability of software
- Collaborated with clients to define solution requirements

### ML Engineer

August 2022 - February 2024

*DiveDeepAI*

*Islamabad*

- Compiled, cleaned and manipulated data for proper handling
- Analyzed large datasets to identify trends and patterns in customer behaviors
- Developed polished visualizations to share results of data analyses
- Modeled predictions with feature selection algorithms
- Devised and deployed predictive models using machine learning algorithms to drive business decisions
- Collaborated with multidisciplinary team to develop machine learning solutions for real-world challenges
- Utilized tools including Numpy, TensorFlow, PyTorch, OpenAI, Flask, Django, FastAPI, Docker etc to build and deploy AI models
- Led and contributed to various projects, including: Custom CHATBOT, Outscore.com, Data visualization
- Gained expertise in data preprocessing, feature engineering, model validation, and optimization

## Skills

- Python Programming
- Neural Networks
- Machine Learning
- Natural Language Processing
- Retrieval augmented generation (RAG)
- Generative AI (GenAI)
- LangChain
- MultiAgents Systems
- LLMops
- Fine-tuning (LLMs)

## Projects

- **Attention is all you need:** Implemented the details described in the paper without using any external library in python. Trained a simple sentence completion model using the model built from scratch. Obtained accuracy of approximately 75% on the test dataset (custom dataset created using huggingface datasets)
- **Finetuning of BERT model:** Fine-tuned a BERT model (bert-base-uncased-en) for the classification of sentiments in the IMDBReviews dataset
- **Stock Price calculation:** Developed a model to predict stock price values based on indicators and news sentiments. Best results achieved using the BLS model
- **Custom hybrid model for data values prediction:** Designed a custom hybrid model using CNN and FCNN to predict 9 values from 49 values. Trained the custom-built model using PyTorch and achieved 90% accuracy

## Research Projects

- **Mitigating Hallucination in Large Language Models (LLMs):** Investigating underlying causes of hallucination in LLMs. Analyzing effects on model responses and impact on downstream applications. Developing techniques to validate LLM responses to enhance model reliability. Exploring methods to reduce need for human evaluation, enabling faster feedback and improvements
- **Compression of Pre-trained Language Models using Knowledge Distillation:** Investigating methods to distill knowledge from large pre-trained LLMs into smaller, more efficient models. Exploring techniques for optimizing inference speed and memory footprint while maintaining performance. Enhancing accessibility of advanced language models by reducing resource requirements

## Education

### Master's in Artificial Intelligence

2021 - 2023

*NUCES, FAST*

*Islamabad*

### Bachelor of Electrical Engineering (Telecom)

2015 - 2019

*NUST*

*Islamabad*