Antariksh Shreshthi

Mumbai, India me@antarikshshreshthi.com | LinkedIn | Medium

AI Researcher/Engineer

"Driven machine learning and AI enthusiast with a strong foundation in Natural Language Processing and software development. Currently pursuing a PhD in Computer Science, with research focused on Large Language Models. Passionate about building AI-driven solutions, contributing to cutting-edge research in AI, and optimizing machine learning models for real-world impact."

- **Programming Languages** – Python, C++, Matlab, SQL
- **ML Frameworks** – PyTorch, HuggingFace, scikit-learn, Keras, Surprise
- **Libraries & tools** – Pandas, NumPy, Matplotlib, NLTK, SciPy, MLflow, Streamlit
- **Model Deployment & MLOps**– Flask, FastAPI, Docker, AWS, Grafana, Git
- **Algorithms & Techniques** – Transformers, NLP, LLMs, Tokenization, Embeddings, Attentions Mechanisms, Text Summarization, Sentiment Analysis, Predictive Modelling, Forecasting
- **Vision** – Image Segmentation, Object Detection, Facial Recognition, Filtering, Feature Extraction, Open CV
- **IOT** – Edge Imp., Device Connectivity (MQTP, HTTP), Cloud Integration, Real-time Dataprocessing, MIT app inv
- **Languages** – Native (English, Marathi, Hindi), B1-German, A1-Italian

EXPERIENCE

Researcher (AI/ML LLMs)
University of Rome Tor Vergata - Electronics Department Rome, Italy
Conducted research on LLMs, using papers, articles and framework/library documentation that provide in-depth explanations, discussions, and experiments on the transformer model used in LLMs, especially the tokenization and embedding steps.

- **Tokenization** – Comprehensive study of the tokenization process and various tokenization algorithms and tools. *Byte Pair Encoding*, *WordPiece*, *Unigram* and *SentencePiece* were discussed in detail and implemented in code.
- **Embeddings** – Conducted an in-depth analysis of the embedding process, including word embeddings and positional encoding. Various research papers on different types of *positional encodings*, such as *sinusoidal*, *relative* and *rotary*, were investigated. *Word embeddings* like *Word2Vec* and *GloVe*, were studied and implemented in code.
- **Transformer** – Explored key components, including the attention mechanism, encoder/decoder architecture, and the roles of *queries, keys and values*. Studied innovations and implementations of prominent models like GPT, BERT, T5, etc. Learned how to build *GPT-1 from scratch* in PyTorch environment, gaining insights into its architecture and training process.

PUBLICATIONS

A.J. Shreshthi, S. Spanò, L. Di Nunzio, R. La Cesa, C. Valenti, M. Re, G.C. Cardarilli. *Parameter-Efficient Fine-Tuning of Large Language Models for Low-Power Sentiment Analysis Applications*. International Journal on Informatics Visualization, 2026.

PROJECTS

Fine-tuned BERT to Outperform GPT models
Fine-tuned BERT base using the SST-2 dataset for sentiment analysis

- Improved accuracy from 0.51 to 0.91 and F1 Score from 0.39 to 0.91
- Outperformed GPT-3.5(and older gpts) and models like Mistral-7B, Falcon-7B, etc. Matched GPT-4o-mini performance.

Personalized movie recommender system
Developed a collaborative filtering model using matrix factorization with Surprise to provide personalized movie recommendations.

- Reduced RMSE from 1.15 to 0.95 and improved Precision@10 from 0.65 to 0.77 by fine-tuning model parameters.

EDUCATION

PhD Computer Science (AIML NLP) – DY Patil University, India Mar 2026 – Present
Research focused on the transformer architecture and large language models, including LLM training and fine-tuning, evaluation and benchmarking, model interpretability, efficient training methods, and the safe deployment of large-scale AI systems.

MS Mechatronics Engineering - University of Rome Tor Vergata
*Recipient of the highly selective 'University Incentive', awarded to the **top 3** students of MS Mechatronics Engineering.*

B.E. Mechanical Engineering – University of Mumbai, India
Led the final year project group in the design and fabrication of a Mars rover, proposing an innovative self-extraction solution to help the rover navigate extra-terrestrial terrain effectively, showcasing my skills in engineering, project management and problem-solving.