SAMAR SINGH

COMPUTER SCIENCE STUDENT AND ML RESEARCHER

linkedin.com/Samar | bufferfis | samar.singh2023@vitstudent.ac.in | +91 972-455-5679 | Samar_Singh1

EDUCATION

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE
Vellore Institute of Technology
Aug 2023 – May 2027 **(CGPA: 9.07)**

EXPERIENCE

UNIVERSITY OF PRETORIA

ML INTERN

Pretoria, South Africa (Remote)
Feb 2024 – Feb 2025

Skills: Pytorch, Transformers

Responsibilities:

- Worked on project **MAGE: Multi-Head Attention Guided Embeddings for Low Resource Sentiment Classification**
- Work submitted as an abstract to the SACAIR 2025 conference.
- Achieved a boost of 2% accuracy and precision over classification tasks by a LSTM through novel language-agnostic data augmentation
- Demonstrated incremental improvements in challenging low-resource sentiment classification scenarios

HELIVERSE X GOOGLE DEVELOPERS GROUP ON CAMPUS

ML DEVELOPER

On campus
Jan 2025 – Apr 2025

Skills: Pytorch, FastAPI, Huggingface, Transformers

Responsibilities:

- Worked on vision-language description layer for a crime detection prototype.
- Collaborated in a multi team environment. Lead the development of scene understanding component of the pipeline.

PROJECTS

Project: Transcript Correction Pipeline — Open Source Project

Stack: Python, FastAPI, Pydantic, Google Gemini (structured output via google-genai), RapidFuzz, Uvicorn, Requests

- Designed and shipped a two-stage correction API with deterministic entity normalization and LLM refinement with a strict JSON schema, exposed as a single endpoint for easy integration.
- Logged human-in-the-loop CSVs for active learning: a review file for edge cases (“review” with reasons) and an accepted file for high-confidence outputs, enabling future replacement with a locally deployed end-to-end model trained on the “accepted” corpus.
- Results and speed (best observed): punctuation/flow cleanup improved readability on key indices (e.g., FRE +4.06, FKGL 1.56, ARI 2.00, CLI 3.70, LIX 4.00), with latency as low as 1.25 s per segment and typical 1.27–2.24 s across the evaluation set.

Project: Transformer from scratch — Open Source Project

Stack: Python, Pytorch

- Implemented a Transformer model from scratch using PyTorch, based on the ”Attention is All You Need” paper.
- Built complete encoder-decoder architecture with multi-head self-attention, positional encoding, and feed-forward networks.
- Implemented training features including learning rate scheduling, gradient accumulation, and checkpoint management.

TECHNICAL SKILLS

Languages	Python, Java, C/C++, HTML/CSS/JavaScript
Tools	Git, GitHub, Visual Studio, VS Code, Colab, Kaggle, Linux, LaTeX
ML Stack	PyTorch, TensorFlow, Scikit-learn, Transformers, google-genai, KaggleHub
Data Science	Pandas, NumPy, SeaBorn, Matplotlib
Backend	Oracle SQL, MySQL, FastAPI, Requests, Uvicorn

CERTIFICATIONS

Supervised Machine Learning: Regression and Classification

- **Stanford Online** [Certificate Link]
- Scikit-learn model building & evaluation
- Linear/Logistic Regression & Gradient Descent

Advanced Learning Algorithms

- **Stanford Online** [Certificate Link]
- Neural Networks with TensorFlow
- Decision Trees & Ensemble Methods

ACHIEVEMENTS

Pentathon 2024 Finalist
33rd nationwide rank (**Top 2%**) among **1557 teams** in national cybersecurity competition

Password 24 Top 10
Solved a broad range of cybersecurity challenges, leveraging Linux utilities.

FOSSIT Finalist
1st runner up in an open source themed hackathon