






# ANUSHA NANDY

DATA SCIENTIST / ML ENGINEER

 [Portfolio](#) |  [Linkedin](#) |  [anushacodes](#) |  (205) 401-6375 |  anusha.nandy1@gmail.com

## PROFESSIONAL SUMMARY

Enthusiastic and fast-learning Data Science student with a **strong foundation** in **Machine Learning**, **NLP**, and **Computer Vision**. Adaptable and solution-oriented, thrives in collaborative environments and is always ready to tackle new challenges. Seeking opportunities to contribute to cutting-edge projects at the intersection of AI research and impactful applications.

## EDUCATION

### [University of Alabama at Birmingham]

#### Masters in Data Science

Sep '24 - Mar '26

- **Relevant courses:** Machine learning, Deep learning, Data mining, Foundations of Data Science, Advanced algorithms, OOP (Java)

### [Mahindra University]

#### Bachelors in Artificial Intelligence

Aug '20 - Jun '24

- **Relevant courses:** NLP, Reinforcement learning, ML with Python, Image processing, Big Data, Computation theory, DBMS, OS

## PROJECTS

### Transformers de zéro (from scratch)

- Built a **Transformer model from scratch in PyTorch**, implementing self-attention, positional encoding, and layer normalization per the "Attention Is All You Need" paper.
- Trained a **22M**-parameter model on a **6M+** row *English-French dataset* to demonstrate **scalability** and **performance**.

### PaperSage

- Developed a **RAG** based PDF assistant using **LangChain**, **ChromaDB**, and **Ollama** (Mistral) to retrieve and summarize research documents.
- Optimized text chunking, retrieval, and response generation using the **Mistral LLM** with semantic search and vector embeddings, enabling fast, context-aware querying for accurate and efficient document-based Q&A.
- Designed a **Flask-based web interface** for user-friendly interaction.

### SpamSense: YouTube Comment Spam Detection

- Built and evaluated ML/DL models to classify spam comments; **fine-tuned BERT** achieving **96.94% accuracy**, outperforming traditional models like SVM and XGBoost.

### Book Recommendation Engine

- Built a full pipeline on the GoodReads10K books dataset and benchmarked popularity baselines, user-user/item-item CF, TF-IDF content filtering, SVD/SVD++ matrix factorization, and a two-tower neural model; achieved a best RMSE of **0.84** and Precision@5 of **77%**.

### Grokking Optimizers

- Implemented and visualized **gradient-based optimization algorithms** (GD, SGD, RMSProp, and Adam) using **Numpy**, with 2D/3D animations to analyze convergence behavior, and wrote a detailed report.

### Cohesive Group Emotion Recognition

- **Published at SNPD 2023**. Collaborated in a team of 6 and developed a deep learning model to predict emotions in group images by analyzing individual expressions.
- Optimized face detection models (**YOLOv3**, **HaarCascade**, **SSD**) and pre-trained emotion recognition models (DeepFace, FER) on a custom dataset, achieving **~90% top-3 accuracy**.

## WORK EXPERIENCE

### [Indian Oil Company]

#### ML Intern

Jun '23 - Jul '23

Technologies: TensorFlow, YOLO, OpenCV, MediaPipe, BeautifulSoup

- Developed a **Python web scraper** to automate data collection, acquiring 10k+ labeled images of Indian vehicles.
- Enhanced internal datasets and improved research efficiency by **50%**, enabling better model training for vehicle classification.
- **Fine-tuned YOLOv7 via transfer learning**, increasing precision by **20%** for diverse vehicle detection
- Optimized and converted the model to **TFLite**, enabling real-time monitoring on Android with **30% faster inference**.

## SKILLS

**Languages:** Python | MySQL | Java | R | C/C++

**Libraries:** PyTorch | Scikit-Learn | Transformers (Hugging Face) | TensorFlow | Ollama | LangChain | OpenCV (CV2)

**Specializations:** NLP | Computer Vision | GenAI (LLMs)

**Tools:** Git | GitHub | ChromaDB | Docker | Visual Studio | Jupyter Notebook

## CERTIFICATIONS AND COURSES

- **Certification:** Problem Solving (Intermediate), SQL (Intermediate), Software Intern – [HackerRank](#)
- **Courses:** Mathematics for Machine Learning Specialization (Coursera), CS224N (Stanford NLP), CS229 (Stanford Machine Learning)