# Anurag Mishra

+91-9911210461
officiallyanurag1@gmail.com
https://www.github.com/OPanurag
https://www.linkedin.com/in/anuragmishra02/

## PROFESSIONAL SUMMARY

Data Scientist and ML Engineer with hands-on experience developing and deploying scalable LLM based solutions & ML models in production environments. Skilled in designing sustainable ML pipelines, optimizing model performance, & translating complex data into strategic insights. Adept with modern NLP frameworks, cloud platforms (GCP) & MLOps best practices to drive data-driven decision-making & automation at scale.

## TECHNICAL STACK AND SKILLS

**Programming Language**: Python, R, Bash, Git, HTML
**Data Science**: Pandas, Scikit-Learn (sk-learn), Transformers, LLM, Flask, PyTorch, Tensorflow, MLOps
**Tools**: JIRA, Notion, Jupyter, Power BI, Tableau, Git, Docker, Kubernetes, NLTK, ZenML
**Cloud Platform**: Google Cloud Platform (GCP), Amazon Web Services (AWS)
**Database**: SQL, Milvus Vector DB
**Data Analytics**: Seaborn, Dask, Matplotlib, Plotly, Bokeh, SciPy, Spacy

## PROFESSIONAL EXPERIENCE

- **Kounsel – https://kounsel.io**                                                 *Oct, 2024 – Present*
  *Graduate Data Scientist*                                                          California, US
  – Utilized GCP and specialized NLP libraries to scale and optimize the LLM development process
  – Developing a Large Language Model (LLM) focused on medical benefits, especially in recipe and diet generation
  – Extracting and structuring nutritional data from diverse sources to build a large-scale ingredient nutrition dataset
  – Collaborated with medical professionals & researchers ensuring accuracy & relevance of dietary recommendations

- **Omdena – https://www.omdena.com**                                              *Jul, 2024 – Sep, 2024*
  *Machine Learning Engineer* – Intern                                               California, US
  – Worked on "AudioShield" project to distinguish deepfake audio from original audio samples
  – Utilized XGBoost to optimize audio threat detection, reducing false positives by 25%
  – Enhanced audio classification model with a 20% accuracy improvement through feature engineering and tuning

## HANDS-ON PROJECTS

- **AudioShield: Deepfake Audio Detection – https://huggingface.co/spaces/savinshynu/audioshield-hugg**
  *Developed an audio classification model to detect deepfake audio, deployed on Hugging Face*
  – Deepfake audio classification model through feature engineering and model tuning
  – Fine-tuned XGBoost for better threat detection
- **Singapore: Recycled Energy Saved – https://github.com/OPanurag/Singapore_Recycled_Energy_Saved.git**
  *Analyzed 18+ years of recycling and waste data to quantify energy savings and trends.*
  – Revealed energy savings of up to 500 GWh annually from five waste types, supporting sustainability decisions
  – Recommended a strategy that could reduce landfill dependency by 30%
- **Generative AI Chat Bot – https://github.com/OPanurag/AI_ChatBot_System**
  *Developed a chatbot using NLP that handles 1,000+ monthly interactions with environment-personalized responses.*
  – Boosted engagement by 35% through personalization based on 5 environmental and historical data factors
  – Integrated chatbot into web platforms for seamless deployment
- **12+ Projects of AI Engineer, ML Engineer, Data Analyst & Data Scientist roles**

## EDUCATION

- **Vellore Institute of Technology**
  *Bachelor's of Technology - Computer Science Engineering Spl. in Artificial Intelligence and Machine Learning*    83.5%

## CERTIFICATIONS AND TRAININGS

- **Data Science** – IBM
- **Python** – Google
- **Applied Machine Learning in Python** – University of Michigan
- **SQL** – Kaggle