

Vikrmaditya Mishra

+91 6394514394 | vikrmadityamishra@gmail.com | [Portfolio](#) | [LinkedIn](#) | [GitHub](#)

EXPERIENCE

Data Science Intern | VRCYN

Dec 2024 – May 2025

- Assisted in preparing, cleaning, and structuring customer feedback datasets (chat logs and surveys) using **Python**, **NLTK**, **spaCy**, and **SQL**.
- Explored and implemented **NLP techniques**, including **LSTM-based sentiment analysis**, keyword extraction, and basic topic modeling to identify customer pain points and satisfaction drivers.
- Contributed to insights that supported data-driven discussions for improving product features, customer experience, and internal decision-making.

TECHNICAL PROJECTS

ClaimLens - Clause-Aware RAG System | [LangChain](#), [FAISS](#), [LLM](#) | [GitHub](#)

Jan 2026 – Mar 2026

- Built a production-grade **Retrieval-Augmented Generation (RAG)** pipeline for insurance policy analysis across multi-insurer documents.
- Implemented a **deterministic clause splitter** parsing legal PDFs into atomic **clause chunks** across 5 structural formats with canonical clause IDs and duplicate detection.
- Designed a two-stage retrieval architecture using **embeddings + FAISS (Top-40 retrieval)** with **cross-encoder reranking (Top-5)** achieving **Recall@20: 0.93**, **MRR: 0.89** on single-clause queries.
- Developed a multi-clause evaluation framework with stage-wise diagnostics, achieving **Coverage@20: 0.87**, **Full Recall@20: 0.60**, and **MRR: 0.77** on composite multi-hop queries.
- Developed an **LLM reasoning engine** with strict **Pydantic** schema validation, citation grounding, and JSON retry logic returning structured answers with confidence and **clause citations**.

GPT From Scratch | [Python](#), [PyTorch](#) | [GitHub](#)

Oct 2025 – Nov 2025

- Built a decoder-only transformer language model from scratch, implementing multi-head self-attention, positional embeddings, feed-forward layers, and layer normalization across **6 transformer blocks** with **384-dimensional embeddings**.
- Trained a character-level transformer on 200K tokens of structured Shakespearean dialogue (20K+ lines) with speaker annotations, modeling long-context dependencies (256-token window) and generating coherent multi-turn text using the **AdamW optimizer**, achieving **training loss 1.5** and **validation loss 1.8**, generating coherent text with structured dialogue and thematic consistency.
- Optimized the training pipeline with **gradient clipping**, **dropout (0.2)**, and **efficient batching (64 sequences, 256-token context)**, reducing overfitting while improving validation stability.

TECHNICAL SKILLS

Languages:	Python, SQL, C++.
Data Tools:	Pandas, NumPy, Matplotlib, Seaborn.
ML/DL Frameworks:	PyTorch, TensorFlow, Scikit-learn, OpenCV, YOLO.
LLM / GenAI:	RAG, LangChain, Vector Databases.
APIs & Deployment:	FastAPI, Docker, Git, GitHub.
Databases:	MySQL, PostgreSQL.
Soft Skills:	Resiliency, Communication, Adaptive Learning, Interpersonal Skills.

ACHIEVEMENTS

IFERP | Conference | [GitHub](#)

- Paper accepted at IFERP International Conference on "Real-Time Sign Language Detection Using CNN and OpenCV" (2024).

GDSC | AI/ML Lead

- Executive member at Google Developer Student Clubs.
- Helped organize multiple technical and ML-focused events under GDSC.

EDUCATION

Indian Institute of Technology Patna

Patna, Bihar

Master of Technology in AI and DSE

Jan 2026 – Present

Lovely Professional University

Jalandhar, Punjab

Bachelor of Technology in Computer Science and Engineering

Aug 2020 – Aug 2024