

EISHAAN KHATRI

+91-8171380034

eishaankhatri@gmail.com

LinkedIn/Eishaan Khatri

Github/Eishaan Khatri

Research Interests

Natural Language Processing, Multilingual & Low-Resource Language Modeling, Conversational AI, LLM Evaluation & Safety, Explainable AI for Fairness

Professional Summary

Final-year Mathematics & Computing student with published research in NLP for low-resource and code-mixed languages, and hallucination reduction in zero-cost LLMs. Experienced in building multilingual conversational AI systems, multi-agent pipelines, and explainable AI for ethical decision-making. Brings rigorous mathematical foundations in optimization and statistics toward advancing inclusive and equitable language technologies.

Education

Rajiv Gandhi Institute of Petroleum Technology, Amethi

May 2026 (Expected)

B.Tech in Mathematics & Computing | Minor in Business Consulting

CPI: 8.60 | 8.14

- **1st Rank** in the Mathematics Department
- **Relevant Coursework:** Natural Language Processing, Machine Learning, Deep Learning, Probability & Statistics, Linear Algebra, Optimization, Numerical Methods

Delhi Public School, Hapur

Apr 2021 - Jun 2022

Class 12

90.67%

Delhi Public School, Hapur

Feb 2019 - Mar 2020

Class 10

96.80%

Publications

Target Based Offensive Language Detection in Hinglish [Link](#) | Dr. Santosh K. Mishra

- Accepted at **Expert Systems with Applications** Volume 308, 1 May 2026, 131033
- Constructed and released a **6.8k-instance Hinglish offensive language dataset**, addressing data scarcity in **low-resource, code-mixed language** processing and enabling cross-lingual robustness studies
- Proposed a **novel token-level annotation framework** improving offensive language detection across pre-trained multilingual models (mBERT, XLM-R, MuRIL)

Bridging the Gap: Enhancing and Evaluating Zero-Cost LLMs for Scientific Question Answering [Link 1](#) [Link 2](#)

- Accepted for **Poster Presentation** at the 2nd Conference on Applied AI and Scientific ML (**CASML 2025**)
- Engineered a neuro-symbolic pipeline integrating **Dense Retrieval** and **Knowledge Graph Grounding** to democratize AI access by optimizing free-tier LLMs for researchers with limited compute
- Achieved a **68% reduction in hallucinations** for Physics/Materials Science tasks and introduced novel fairness-oriented validation metrics (**FC-PS**)

Experience

Indian Institute of Management Shillong

May 2025 - Present

Research Analyst Intern (Mentor: Dr. Shrawan Kumar Trivedi)

Remote

- Researched and implemented **explainable AI techniques** (LIME, SHAP, attention mechanisms) to detect and mitigate **bias and fairness** issues in financial decision systems
- Developed and evaluated predictive models using Python and R, improving accuracy by **15%** through XAI-driven interpretability assessments and comparative evaluation across model families
- Authoring a **research paper** on explainable AI for financial modeling with applications to model transparency and equitable decision-making, under Dr. Shrawan Kumar Trivedi
- Created **visualization dashboards** in Tableau translating model explanations into interpretable stakeholder insights

Polka Labs Pte. Ltd. (Polkasassembly, Singapore)

May 2025 - November 2025

AI Developer Intern

Remote

- Built **multilingual-ready chatbot platform** supporting Discord, Slack, and Telegram with shared conversation history and autonomous response generation
- Designed **multi-agent pipeline** with specialized agents for task decomposition, content generation, and quality assurance; conducted **empirical evaluation** of response quality and persona consistency across platforms
- Implemented **hub-and-spoke architecture** reducing platform integration time from 1 week to under 5 hours
- Developed conversation engine with **persistent memory (Mem0)** and real-time orchestration across multiple LLM services (GPT-4, Claude, Gemini)

Projects

Real-Time Transcription System with Speaker Identification [Code](#)

Jan 2024 - Apr 2024

- Built a **multimodal speech processing system** achieving **85% speaker identification accuracy** using Siamese networks and SVM, improving accessibility for multi-speaker conversational scenarios
- Enhanced transcription quality via **BERT-based embeddings**, achieving **90% accuracy** in multi-speaker settings across diverse acoustic environments
- Improved speaker turn detection with **context-aware NLP**, reaching **80% accuracy** in noisy environments, with emphasis on inclusivity in low-resource audio conditions

Multi-Agent AI for Subsurface Decision Support [Code](#)

October 2024 - January 2025

- Architected a **multi-agent system** with four specialized agents (Petrophysical, Seismic, Risk/ESG) using auditable **JSON artifacts** to ensure transparency in automated decision-making
- Engineered a **fully deterministic pipeline** with input hashing for **100% reproducibility** and a RAG-based chat interface that cites sources to eliminate hallucination
- Validated system logic across **four scenarios**, identifying prime targets and flagging geological or ESG-based risks with interpretable, traceable outputs

Neural Approaches to Pseudo-Random Generation [Code](#)

January 2025 - April 2025

- Explored **Transformer-LSTM hybrid models** combined with reinforcement learning for pseudo-random number generation, with reward functions based on **entropy and unpredictability measures**
- Evaluated outputs against **NIST statistical test suites**; conducted ablation studies on architectural design choices and reward signal formulations

Technical Skills

- **Core Areas:** Multilingual NLP, Conversational AI, Explainable AI, Deep Learning, LLM Evaluation & Safety
- **Programming Languages:** Python, R, C++, JavaScript, SQL, Rust
- **ML/NLP Frameworks:** PyTorch, TensorFlow, HuggingFace Transformers, spaCy, NLTK, fastText, Keras, Scikit-learn
- **Tools & Platforms:** Docker, AWS, Apache Spark, Tableau, Power BI, Git, Snowflake, CircleCI
- **Evaluation & Metrics:** ROUGE, BLEU, F1-Score, NIST Statistical Tests, SHAP/LIME interpretability

Achievements

- **Finalist** at the National Policy Hackathon 2024 by **IIM Calcutta & IIFT Kolkata**
- **Finalist** of FloodAI Hackathon conducted by **IIEC IIT Gandhinagar**
- **3 Star** coder on multiple platforms in **Python** and **C**
- **JEE Advanced 2022:** Qualified (Top 1.5 percentile) amongst **1.2 million candidates**

Leadership & Service

- **Overall Management Head** – OWASP, RGIPT: Led AI security and blockchain initiatives; guided **500+ freshmen** through orientation, emphasizing ethical technology practices
- **Teaching Volunteer** – Gyanarpan, RGIPT: Mentored underprivileged students in STEM subjects and career guidance, promoting inclusive access to education

Languages

- **English** (Fluent), **Hindi** (Native), **Hinglish** (Native proficiency — directly relevant to code-mixed NLP research)