Shubham Kumar

Email: shubham24101@iiitnr.edu.in
Linkedin: linkedin.com/in/cosmicshubham
Mobile: +91-6378051416
Github: github.com/CosmicShubham1

Education

- **International Institute Of Information Technology** — Naya Raipur, India
  B.Tech in Electronics and Communication Engineering — *August 2024 – Present*
  - **Current Semester**: 4th
  - **Current CGPA**: 8.32/10.0
  - **Relevant Coursework**: Data Structures, Python Programming, Probability & Statistics, Linear Algebra.

Research Experience

- **Low-Resource Neural Machine Translation (NMT)** — IIIT Naya Raipur
  Undergraduate Researcher (Advisor: Dr. Aruna Shukla) — *July 2025 – Present*
  - **Text Mining & Data Alignment**: Curated and **manually aligned** a parallel dataset for a Himachali dialect, developing a custom cleaning pipeline using **Regex** to ensure data quality and handle unstructured text.
  - **Architectural Pattern Analysis**: Implemented and compared three distinct architectures to identify **linguistic pattern convergence**: a custom **Baseline Transformer** (trained from scratch) versus fine-tuned multilingual models **NLLB-200** and **M2M-100**. Also experimented with IndicTrans2 and mBART.
  - **Optimization & Evaluation**: Performed **hyperparameter optimization** to minimize loss on sparse data. Benchmarked model performance using standard metrics **(BLEU, chrF, TER)** and semantic scorers **(Perplexity, BERTScore, F1)**.
  - **Visualization**: Utilized **Graphviz** to visualize model architectures and attention mechanisms.
  - **Publication**: Authored a research paper detailing the methodology, benchmarking, and error analysis, targeting submission to **LoResMT 2026**.

Projects

- **Language Preservation App (NLP Integration)**: Mobile application (Himachaliverse) for digitizing local languages. Integrated the custom-trained NMT models to provide real-time translation. Optimized model inference for performance.
- **Automated Pen Plotter (CNC Machine)**: Built a complex hardware-software system to replicate handwriting. Developed **Python algorithms** to convert vector graphics into coordinate-based G-code for precise motor control.
- **Cerebellum (AI Multiplayer Game)**: Built a real-time Pictionary-style game using **Flask** and WebSockets. Implemented backend logic for scoring and room management. (Demonstrates API & Deployment skills).
- **Electronic Voting Machine (Face Recognition)**: Secure voting system using Python, OpenCV, and Arduino for facial recognition based authentication.

Research Interests

Natural Language Processing (NLP), Low-Resource Text Mining, Neural Machine Translation (NMT), Machine Learning, Deep Learning Optimization, Pattern Recognition in Language Models.

Technical Skills

- **Languages**: Python, C++, SQL
- **NLP & AI**: Hugging Face Transformers, PyTorch, Spacy, NLTK, Regex (Text Extraction)
- **Concepts**: NMT, LLM Fine-tuning, Tokenization, Model Evaluation, Optimization
- **Tools**: Git, Google Colab Pro, LaTeX, Graphviz
- **Web & APIs**: Flask, REST APIs, HTML/CSS

Achievements

- **Qualified IOQM**: National-level math olympiad organized by MTAI and HBCSE (Strong Mathematical Foundation). (January 2021) **[View Certificate]**
- **National Symposium on Digital Humanities (DREAM)**: Participated in the national symposium focused on NLP advancements and digital research. Received Certificate of Participation. (November 2025) **[View Certificate]**

Certifications

- **Introduction to Artificial Intelligence (AI)**: IBM via Coursera (February 2025) **[View Certificate]**
- **Deep Learning for Developers**: Infosys Springboard (February 2025) **[View Certificate]**

Extra-Curricular Activities

- **3rd Place – IPL Auction** — IIIT-NR
  E-Summit analytics and strategy event — *February 2025*
- **Event Execution Team – Ganesh Chaturthi** — IIIT-NR
  Coordinated logistics for annual cultural celebration. — *September 2024*