

# Shivraj Anand

Researcher (Language Analyst) @ IIT Kharagpur

**Contact:** 8092776242 **Email:** shivrajanand022002@gmail.com **Location:** Kharagpur, India  
**Linkedin:** linkedin.com/in/shivrajanand **Github:** github.com/shivrajanand **Portfolio:** shivrajanand.github.io

## Profile

Language Analyst (Contractual Researcher) at IIT Kharagpur under the MeitY-funded NLTM: BHASHINI project, working on interpretable frameworks for multilingual text-to-text translation using Sanskrit-based interlingua representations. B.Tech. in Computer Science Engineering (2021–2025) with a CGPA of 8.8. Experienced in Python, machine learning, NLP, and data analytics, with 9+ end-to-end projects. Passionate about research in language models, machine translation, and interpretable NLP systems.

## Areas of Expertise

• Python Programming • Machine Learning • Natural Language Processing (NLP) • Data Analysis • Visualization • Deep Learning • PyTorch • Scikit-learn • Transformers • Web Development (HTML, CSS, JavaScript, React, Flutter, Django) • Git • GitHub • Algorithm Design • Problem Solving • Research • Technical Writing

## Education

**B.Tech in Computer Science and Engineering** 2021–2025  
B.P.M.C.E., Madhepura — CGPA: 8.80  
Relevant Subjects: Data Structures & Algorithms, DBMS, ML, NLP, OS, Computer Networks  
Achievements: Branch Topper in all academic years; GATE'24 CS, GATE'25 CS, DA Qualified

**Higher Secondary Certificate (12<sup>th</sup> Grade)** 2018–2020  
Kendriya Vidyalaya, Jamalpur — 94.4%  
Subjects: Physics, Chemistry, Mathematics, Computer Science  
Achievements: 2<sup>nd</sup> rank in school in 12<sup>th</sup> board exams

**Secondary Certificate (10<sup>th</sup> Grade)** 2017–2018  
Kendriya Vidyalaya, Jamalpur — 90%  
Subjects: General Science, Math, English, Social Science, Hindi  
Achievements: 2<sup>nd</sup> rank in school board exams

## Projects [View all](#)

- **NLLB Finetuning for Sanskrit-Hindi Translation** — NLP (B.Tech. Major Project) [\[Github\]](#) [\[Blog\]](#)
  - Using model nllb-200-distilled-600M large parallel corpus of sanskrit-hindi dataset, from Gitapress, consisting of 16,330 training sentence pairs.
  - Attempting to improve Sanskrit-Hindi translation accuracy over existing models (e.g., NLLB, mT5, mBART, IndicTrans2), which currently show limited performance on this language pair.
  - Under supervision of Prof. Sujeet Kumar.
  - Got a BLEU score of 11 and CHRF++ score of 37 compared to models like mt5, Mbart for which scores are less than 1 in zero-shot-environment.
  - Tech-Stack: Transformer, HuggingFace, NLLB, mT5.
- **Autocorrect System** — NLP, Language/Channel Models (B.Tech. Minor Project) [\[Blog\]](#) [\[Kaggle\]](#)
  - Implemented spelling correction using probabilistic models for text decoding.
  - Tech-Stack: Pandas, Numpy, Re, Spacy.
- **Next Word Recommender System** — NLP (2024) [\[Blog\]](#) [\[Kaggle\]](#)
  - Built a next-word prediction engine using bigram and trigram probabilistic models.

- Tech-Stack: Pandas, Numpy, Re, Spacy.
- **Machine Learning Course Website** — HTML, CSS, JavaScript (2024) [\[LIVE\]](#)
  - Designed and deployed a beginner-friendly machine learning tutorial website.
  - Covered concepts like supervised learning, regression, and classification through blog-style lessons.
  - Tech-Stack: HTML, CSS, JS
- **Housing Sales Data Analysis** — Machine Learning (May 2023) [\[Kaggle\]](#)
  - Performed data preprocessing, visualization, and model building on housing datasets.
  - Tech-Stack: Pandas, Numpy, Scikit-learn, Seaborn .
- **Smart Task Manager** — Python, Tkinter, GUI [\[Github\]](#)
  - A desktop-based task management system built with Python and Tkinter.
  - Tech-Stack: Tkinter, Python, GUI
- **Random Password Generator** — HTML, CSS, JavaScript [\[LIVE\]](#)
  - Web app that generates passwords using randomness derived from the Collatz Conjecture.
  - Tech-Stack: HTML, CSS, JS
- **Cognify Data Science Internship Project** — Python, Pandas, Matplotlib [\[LIVE\]](#)
  - This repository contains the complete solutions to the tasks assigned during my internship at Cognifyz Technologies under the Data Science Internship Program.
  - Through this internship, I gained hands-on experience in data preprocessing, analysis, visualization, and predictive modeling using real-world data
  - Tech-stack : Python 3.10.11, Pandas, Numpy, Matplotlib, Seaborn e, Geoplot.

## Professional Experience

---

### Language Analyst (Contractual Researcher)

*IIT Kharagpur*

**Kharagpur, West Bengal**

*July 2025–present*

- Selected as a Language Analyst (Contractual Researcher) at IIT Kharagpur under the National Language Translation Mission (NLTM): BHASHINI, funded by MeitY, Government of India.
- Working under Prof. Pawan Goyal (CSE, IIT Kharagpur) on developing an interpretable, unified framework for multilingual text-to-text translation across Indian languages using a Sanskrit-based interlingua.
- Contributing to research in Natural Language Processing (NLP), focusing on machine translation, low-resource language modeling, and explainability in transformer-based architectures.
- Engaged in data curation, linguistic analysis, and the design of interpretable intermediate representations to bridge structural divergences among Indian languages.
- Supporting open-source and policy-aligned NLP solutions as part of India's multilingual AI infrastructure under the BHASHINI initiative.

### Student Representative

*Startup Cell, BPMCE, Madhepura*

**Madhepura, Bihar**

*Oct 2023–May 2025*

- Helped plan and conduct 30+ startup events, workshops, and seminars to encourage entrepreneurship on campus.
- Directed logistics and communication with internal teams as well as external parties to facilitate the smooth flow of events.
- Created detailed records, documentation, and reports for Startup Cell operations in line with institutional objectives.
- Fostered innovation and startup culture among students through outreach and engagement efforts.
- Served as a bridge of communication between student entrepreneurs and Startup Cell leadership.
- Helped to maintain the Cell's social media and online communications.

## Online Courses & Certifications [View all](#)

---

- **Experience Certificate**

- Student Representative — Startup Cell, BPMCE (May 2025)

- **Machine Learning & NLP**

- Machine Learning Specialization – DeepLearning.AI & Stanford University (Coursera)
- Supervised Machine Learning: Regression and Classification – DeepLearning.AI (Coursera)
- Advanced Learning Algorithms – DeepLearning.AI (Coursera)
- Unsupervised Learning, Recommenders, Reinforcement Learning – DeepLearning.AI (Coursera)
- Natural Language Processing — Internshala (Mar 2025) — Scored 90%
- NLP Training Completion — NSDC (Mar 2025) — Grade: A
- Data Science Internship — Innovate Interns (Nov 2024 – Jan 2025)
- Data Science — Internshala (Ongoing)
- Machine Learning — Internshala (Mar–May 2023)

- **NPTEL & Core CS Subjects**

- Programming, DSA using Python — NPTEL (Jan–Mar 2024) — Top 2%
- Database Management Systems — NPTEL (Jul–Sep 2024) — Top 2%, Score: 79%

- **Programming Languages & Tools**

- C, C++, Java, Python — Spoken Tutorial (2023–2024)
- Programming in C and C++, DSA — Internshala (2023)

- **Cybersecurity & Systems**

- Introduction to Cybersecurity — Cisco (Apr 2023)
- NDG Linux Unhatched — Cisco (Jul 2023)

- **Other Certifications**

- Salesforce Developer Virtual Internship — Salesforce (Jan 2024)
- Digital Productivity — NIIT Foundation (Dec 2024)

## Skills

---

- **Programming Languages:** Python, C, C++, JavaScript
- **Web & App Development:** HTML, CSS, React.js, Flutter, Django
- **Machine Learning:** Scikit-learn, Pandas, Numpy, Matplotlib
- **NLP:** Transformers, SpaCy, Regex, PyTorch, Tensorflow
- **Data Visualization:** Power BI, Excel, Matplotlib, Seaborn
- **Tools:** Git, GitHub, Jupyter, Google Colab, VS Code
- **Soft Skills:** Communication, Teamwork, Event Management, Research Mindset

## Honors & Distinctions

---

GATE 2024 AIR 8532 • Top 2% in NPTEL DBMS & DSA • Branch Topper in B.Tech. (1st–4th Year) • GATE 2025 AIR 4883(DA), AIR 8357(CS)

## Languages

---

- Hindi (Native), •English (Fluent)