

# Mansi Budamagunta

DATA SCIENTIST

✉ [b.mansi12@gmail.com](mailto:b.mansi12@gmail.com) | [in mansi-budamagunta](https://www.linkedin.com/in/mansi-budamagunta) | [github mansi-budamagunta](https://github.com/mansi-budamagunta)

## Summary

---

I'm a data scientist with a research background. I specialize in using computational techniques to study problems across a variety of domains — including AI, policy, environment, and healthcare.

## Education

---

### Utrecht University

MSc Applied Data Science, GPA (so far): 8.5

July 2026 (expected)

Utrecht, The Netherlands

- Thesis: Computational Policy Analysis of EU Budgets

### Indian Institute of Science Education and Research (IISER) Pune

BS-MS Dual Degree, Interdisciplinary Major, GPA: 8.9

Sept 2021

Pune, India

- STEM degree with coursework spanning mathematics, data science, natural sciences, and humanities
- Thesis: Dynamical Processes on Transportation Networks: Effects on Infectious Disease Spreading

## Experience

---

### AI Trainer

Jan 2023 – Mar 2025

Turing

- Worked as a Python programmer on LLM training pipelines for one of the world's leading AI labs
- Contributed to the design of evaluation environments and guidelines across different LLM capability domains — including code generation, reasoning, tool use, web browsing, and computer usage
- Evaluated LLM outputs, identified model weaknesses, and generated targeted feedback within the RLHF loop to improve frontier models
- Led teams of varying sizes, optimising workflows and establishing quality standards

### Research Intern

July 2022 – Jan 2023

Data & Energy in Buildings & Cities Lab, IIT Bombay

Mumbai, India

- Clustered Indian urban districts using remote sensing, climate, census, and demographic data
- Collected and collated diverse datasets for 96 districts; performed k-means clustering to identify patterns
- Revealed that geographic proximity doesn't always determine similarity of districts
- Presented this research at [BuildSys '22: Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation](#)

### Research Intern

May 2019 – July 2019

Theoretical Neuroscience Group, Aix-Marseille Université

Marseille, France

- Analysed connectome-based whole-brain models via numerical simulations
- Contributed to *The Virtual Brain* software (European Human Brain Project)

## Skills

---

**Programming:** Python, R, SQL, HTML, CSS, JavaScript

**ML, NLP & Networks:** PyTorch, Hugging Face Transformers, scikit-learn, NetworkX

**Data & Viz:** Pandas, NumPy, Matplotlib, Seaborn, Plotly

**Tools:** Git, Docker, Quarto, LaTeX

**Languages:** English, Hindi, Telugu (fluent) · French, Dutch (beginner)

## Selected Projects

---

### EU Funding & Technological Sovereignty (MSc Thesis, ongoing)

- Investigating EU research funding alignment with technological sovereignty goals (openness and value alignment) using computational text analysis
- Extracting portal data and extending the [sedia-api-fetchers](#) library to build a reproducible NLP pipeline
- Developing a public-facing dashboard to visualize funding allocation and participant metadata

### Multi-Model vs Single LLM for Code Generation

- Evaluated performance trade-offs between modular multi-model pipelines (reasoner + coder) and single large LLMs for code generation tasks
- Implemented local inference pipelines on CUDA-enabled GPUs using PyTorch and Hugging Face Transformers
- Benchmarked structured reasoning capabilities across specialized models using the LiveCodeBench and APPS datasets

### Predicting Type 2 Diabetes Risk with Fairness Auditing

- Built an interpretable logistic regression classifier as a case study in responsible AI development and bias mitigation
- Conducted comprehensive fairness audits across demographic groups using Equal Opportunity and Predictive Rate Parity metrics
- Applied EU AI Act, DEDA, and FRAIA frameworks to propose governance protocols, risk assessments, and accountability structures

### Infection Spread on Transportation Networks

- Collected Indian census + rail/air transport data for 446 cities
- Simulated SIR metapopulation model to construct an infectious disease hazard map for India
- Contributed to the resulting publication in [Current Science \(2021\)](#) and created the [project website](#)

## Publications

---

**Budamagunta M.**, A. Gaur, A. Verma, A. Anand, C. Deb (2022), Clustering Indian Districts Based on Multidimensional Demographic and Climate Data, *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '22)*, <https://doi.org/10.1145/3563357.3566146> 2022

O. Sadekar, **Budamagunta M.**, G.J. Sreejith, S. Jain, M.S. Santhanam (2021), An Infectious Diseases Hazard Map for India Based on Mobility and Transportation Networks, *Current Science*, 121(9), 1208–1215., <https://doi.org/10.18520/cs/v121/i9/1208-1215> 2021

## Achievements & Extracurriculars

---

- Member of the Young Innovators honours programme, Utrecht University (2025)
- Part of a 9-member national student delegation to Vietnam for a cultural and academic exchange with universities and government officials, selected by the Ministry of Youth Affairs & Sports, Government of India (2017)
- Volunteer teacher and programme coordinator at Disha, IISER Pune's social outreach club — taught children in slums and villages, and helped design and run some of the club's programmes (2016–2021)
- INSPIRE Scholarship holder (2016–2021), awarded by DST, Government of India