

# ANIRUDDHA SAHAY VARMA

+919910233530

New Delhi, India

[aniruddhasvarma@gmail.com](mailto:aniruddhasvarma@gmail.com)

[Website](#) | [GitHub](#) | [LinkedIn](#)

Current projects explore grammar induction via substitutability patterns, and adversarial vulnerabilities in text embedding spaces.

## EDUCATION

|  |      |   |           |
|--|------|---|-----------|
| <b>B. Tech. (Computer Science and Engineering)</b> | 2027 | Delhi Technological University, New Delhi | 8.87 CGPA |
| <b>CBSE (Class XII)</b>                            | 2023 | Amity International School - Pushp Vihar  | 97.4%     |
| <b>CBSE (Class X)</b>                              | 2021 | Amity International School - Pushp Vihar  | 95.2%     |

Relevant Coursework: Information Theory, Machine Learning, Probability & Statistics, Linear Algebra\*, Deep Learning\*, Parallel Computer Architecture\* (\*In progress)

## RESEARCH PROJECTS

### Substitutability-Based Grammar Induction | [GitHub](#) | Python, Rust

- Developed a grammar induction algorithm that discovers structural equivalences (e.g., “eat” ≈ “ate” in shared contexts) rather than relying on token frequency.
- Formalized using Kolmogorov conditional complexity:  $K(Y|X)$  measures how well learned patterns compress held-out data.
- Implemented hierarchical pattern matcher with parallel processing (Rayon); outputs BNF grammars.
- Substitutability captures grammatical categories that frequency-based methods (BPE, WordPiece) miss.

### Black-Box Adversarial Attacks on Text Embeddings | [Colab Notebook](#) | PyTorch, EvoTorch

- Investigated whether evolutionary algorithms can exploit embedding spaces without gradient access.
- Evolved ASCII sequences using genetic algorithms and SNES to maximize cosine similarity against target queries.
- Finding: On a top-ranked MTEB embedding model (1.5B parameters), evolved gibberish achieves higher similarity scores than semantically correct text.
- Direct application to RAG security – adversarial documents can hijack retrieval pipelines without model access.

### Transformer Implementation from Scratch | JAX, einops

- Implemented decoder-only transformer in pure JAX without high-level frameworks (Flax, Haiku).
- Components: multi-head causal self-attention, sinusoidal positional encoding, layer normalization, residual MLPs.
- Used JAX’s functional paradigm: explicit parameter threading, JIT-compiled training loop, manual gradient computation via `jax.grad`.

## EXPERIENCE

### Data Science Intern

July 2025 – August 2025

Municipal Corporation of Delhi (MCD)

- Built an anomaly detection pipeline (Pandas) over five years of property tax records to flag discrepancies and potential tax evasion patterns.

## OTHER PROJECTS

### Sororia - Women Safety App | [GitHub](#) | Flutter, Gemini, Firebase

- **Top 14 out of 3700+ teams** in Google Developer Group on Campus competition (Team Leader).
- Safety app with SOS alerts, safest route planning, and Gemini-powered support chatbot.

### Declarative Function Agent | [GitHub](#) | Python, Gemini API

- Framework translating pseudo-code into executable programs with automatic LLM delegation.
- Agent handles abstract functions (e.g., `is_funny()`) via LLM while generating standard code for defined logic.

## TECHNICAL SKILLS

**ML & Research:** Python, PyTorch, JAX, Hugging Face, Evolutionary Algorithms, Text Embeddings, LLMs (Gemini, OpenAI)

**Systems & Languages:** Rust, C/C++, Git, Linux

**Other:** Flutter, Firebase, Node.js, PostgreSQL

## ACHIEVEMENTS

- **Finalist (Top 14/3700+ teams)**, Google Developer Group on Campus Competition – Team Leader for Sororia.
- **Smart India Hackathon 2024** – Led team to university finals (Top 50 teams).