

# Aarush Sinha

Chennai, India

aarush.sinha@gmail.com — LinkedIn — GitHub — Website

## RESEARCH INTERESTS

---

Natural Language Processing, Information Retrieval, QA-Systems, MultiModal Systems

## EDUCATION

---

**Vellore Institute of Technology**, Chennai, India Sep 2021 — Jun 2025  
Bachelor of Technology in Computer Science and Engineering  
Member: IEEE CS Society, IEEE RAS, University Editorial Team, University Design Team  
**Kendriya Vidyalaya, IIT Madras**, Chennai, India Aug 2016 — May 2021  
Schooling Final Percentage: 92.5

## PROFESSIONAL & ACADEMIC EXPERIENCE

---

**Artificial Intelligence Institute of UofSC (AIISC)** Remote  
*Research Intern* May 2024 — Present  
Worked on creating the **ViBe** dataset. Currently working on alignment for mitigation of hallucinations in Text-to-Video models.  
This research is done under the supervision of Prof. Amitava Das.  
**Dept. of Data Science and AI, IIT Madras** Chennai, India  
*Research Intern* Aug 2023 — Present  
Working on efficiently implementing generative pseudo labelling for effective retrieval of PubMed articles.  
This research is being done under the supervision of Prof. Nirav Bhatt.  
**Olam International** Chennai, India  
*Intern* Jun 2022 — Oct 2022  
Worked with the company's security team to build a web application to generate weekly user reports. Successfully built and deployed the app for the company's intranet worldwide.

## PROJECTS

---

**DocQA** (GitHub)  
Custom **13.2M** parameter cross-encoder QA model for symptom-based disease diagnosis (**94.5% accuracy**).  
**Learning to Forget** (GitHub)  
Explored weight-randomization in NLP models; accuracy dropped from **90%** to **50%**.  
**PoincareEmbedding** (GitHub)  
Implemented hierarchical embeddings using hyperbolic space for lexical entailment.

## PUBLICATIONS

---

**Predicting Liquidity-Aware Bond Yields using Causal GANs and Deep Reinforcement Learning with LLM Evaluation**  
*Arxiv*  
**ArxEval: Evaluating Retrieval and Generation in Language Models for Scientific Literature**  
*Arxiv*  
**ViBe: A Text-to-Video Benchmark for Evaluating Hallucination in Large Multimodal Models**  
*TrustNLP @ NAACL2025*  
**Graph Masked Language Models**  
*Under review AI Open*  
**Efficient Biomedical Information Retrieval via Bio-Informed Generative Pseudo Labelling**  
*Under review ACL ARR*

## CERTIFICATIONS

---

Natural Language Processing with Attention Models Certificate  
Natural Language Processing with Classification and Vector Spaces Certificate  
Mathematics for Machine Learning: Linear Algebra Certificate  
Natural Language Processing in TensorFlow Certificate  
Building Transformer-Based Natural Language Processing Applications Certificate