

# Tarun Gupta

B.Tech., Computer Science and Engineering  
Indian Institute Of Technology (IIT), Indore

[tarungupta360@gmail.com](mailto:tarungupta360@gmail.com)  
[tarun360.github.io](https://tarun360.github.io) (personal website)  
[linkedin.com/in/tarun-gupta-116860178](https://linkedin.com/in/tarun-gupta-116860178)

## EDUCATION

**B.Tech., Computer Science and Engineering**  
Indian Institute of Technology (IIT), Indore  
*Department rank 1, Institute rank 2*

**2018-2022**  
**CGPA: 9.81/10**

## EXPERIENCE

### Research associate at IISc

July 2024 - Present

*LLMs, NLP*

[Paper](#)

- Studied plagiarism in AI-generated research and found 24% of LLM outputs contain unattributed content. Showed that current detection tools miss deliberately plagiarized ideas. Paper accepted to ACL 2025 and shared widely on X, including by Stanford NLP and former Mozilla CEO.
- Investigated position bias and lost-in-the-middle problems in RAG systems. Key concepts included continual pre-training and co-reference resolution.
- Analyzed philosophical arguments for and against “world-models” research direction in AI research. Paper submitted to EMNLP 2025.

### Senior Member Technical at D. E. Shaw (Hedge Fund)

July 2022 - June 2024

*Software Development, High-Performance Computing*

- Developed High Performance Computing (HPC) job scheduler to handle D. E. Shaw’s large-scale data processing demands. [Featured in Business Insider](#).
- Served as primary full-stack developer for LLM-powered employee review assistance software.
- Contributed to the open-source container technology tool Podman by adding a feature to encrypt container-images ([GitHub PR link](#)).
- Awarded **super-star promotion** to Senior Member Technical (L2) in 1.5 years.

### Software Development Intern at D. E. Shaw (Hedge Fund)

May 2021 - July 2021

*Full-stack software development, React, TypeScript, Java*

- Designed and implemented a graph visualization tool using GoJS to enhance traders’ workflow efficiency with an intuitive representation of their complex workflows.

### Research Intern at NTU Singapore and Skit.ai

Dec 2021 - June 2022

*Speech Processing, Transformers*

[Paper 1](#), [Paper 2](#)

- Built a bi-encoder transformer mixture model to estimate speaker age and height from speech signal, leading to a relative improvement of 18.5% in male age estimation and 8.6% in female age estimation over the state-of-the-art benchmarks on the TIMIT dataset (refer to Paper 1).
- Developed a new data augmentation method for spoken language identification, Map-Mix, that leverages model training dynamics of individual data points to improve sampling for latent mixup. Map-Mix improves weighted F1 scores by 2% compared to the random mixup baseline (refer to Paper 2).

### Research Intern at Carnegie Mellon University

Jan 2021 - May 2021

*Computer vision, Self-supervised learning*

[Paper](#)

- Built a contrastive self-supervised learning (CSSL) pipeline for macromolecular structure classification using cryo-electron tomography (cryo-ET) data.
- Extended techniques such as SimCLR, MoCo, and SwAV to the cryo-ET domain, achieving a significant improvement over non-self-supervised learning methods.

### Research Assistant at IIT Indore

Jan 2020 - May 2020

*SVM, Convex programming*

[Paper](#)

- Designed a novel plane-based clustering algorithm based on the principles of SVM, utilizing pinball loss to improve generalization performance on noise-corrupted datasets. Used Concave-Convex Procedure (CCCP) for optimization.
- The proposed algorithm improved clustering accuracy by ~1% on noise-corrupted UCI datasets compared to existing plane-based clustering algorithms.

## PUBLICATIONS

---

**Gupta, T.** and Pruthi, D., 2025. Beyond World Models: Rethinking Understanding in AI models. In submission to EMNLP 2025.

**Gupta, T.** and Pruthi, D., 2025. All That Glitters is Not Novel: Plagiarism in AI Generated Research. ACL 2025.

Rajaa, S., Anandan, K., Dalmia, S., **Gupta, T.** and Chng, E.S., 2023, June. Improving Spoken Language Identification with Map-Mix. International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2023.

**Gupta, T.**, Truong, T.D., Anh, T.T., Chng, E.S. (2022) Estimation of speaker age and height from speech signal using bi-encoder transformer mixture model. Interspeech 2022.

**Gupta, T.**, He, X., Uddin, M.R., Zeng, X., Zhou, A., Zhang, J., Freyberg, Z. and Xu, M., 2022. Self-supervised learning for macromolecular structure classification based on cryo-electron tomograms. Frontiers in Physiology.

Tanveer, M., **Gupta, T.**, Shah, M. and Richhariya, B., 2021. Sparse twin support vector clustering using pinball loss. IEEE Journal of Biomedical and Health Informatics (JBHI).

## ACHIEVEMENTS

---

- **CSE Department Rank 1** and **Institute Rank 2** at IIT Indore (based on CGPA).
- **Best Undergraduate Researcher Award, 2021.**
- **Awarded AP grades:** for exceptional performance in **5** courses.
- **Inter IIT Tech Meet 9.0 (2021):** Awarded **Silver medal** in Bridgei2i's NLP competition.
- **Inter IIT Tech Meet 8.0 (2020):** Awarded **Bronze medal** in BITGRIT's data-science competition.
- **JEE Advanced 2018:** Secured All India Rank 1055 among 150,000 candidates (top 0.7%).

## KEY PROJECTS

---

### Satellite video prediction (in collaboration with [REint.co](#))

August 2023 - September 2023

*Computer Vision, Spatio-Temporal Predictive Learning*

- Implemented and tested various spatio-temporal prediction models, such as ConvLSTM and SimVP, on real-world satellite data. The predicted frames can be fed to models such as MetNet-2 for weather prediction.
- Explored and employed PyTorch DDP for multi-node and multi-GPU training.

### Parallelizing Red Deer Algorithm (RDA) – A Nature Inspired Meta-heuristic Algorithm

April 2021

*Parallel Programming, MPI*

[GitHub Link](#)

- Implemented Red Deer Algorithm (RDA) for solving the Traveling Salesman Problem (TSP). RDA is a meta-heuristic algorithm inspired by the unique mating process of Scottish red deer.
- Used Message Passing Interface (MPI) for parallelizing the RDA algorithm, achieving a speed-up factor up to 4.

### Bridgei2i's NLP competition at Inter IIT Tech Meet 9.0 (2021)

March 2021

*Awarded Silver Medal*

[GitHub Link](#)

- *Theme identification task:* Applied DistilBERT, a distilled version of the BERT model, for fast and scalable binary classification of tweets and articles, achieving an F1 score of 0.89.
- *Sentiment analysis task:* Utilized Ada-BERT transformer model fine-tuned for aspect-based sentiment analysis to identify the sentiments associated with a particular brand.
- *Headline generation task:* Utilized Pegasus, T5, and BART models, achieving 37% average similarity score.

### Adversarial Attacks on Brain Tumor Segmentation Models

April 2020 - June 2020

*AI for Medicine*

[GitHub Link](#)

- Implemented a 3D-UNet model using the TensorFlow library and trained it on BraTS brain MRI data for brain tumor segmentation.
- Tested the model's robustness by implementing adversarial attacks such as Fast-Gradient Sign Method (FGSM), Iterative Fast-Gradient Sign Method (iFGSM), and Carlini & Wagner (CW) attack.

### Cache-Oblivious Algorithms

May 2020 - July 2020

*Design and Analysis of Algorithms*

[GitHub Link](#)

- Implemented various cache-oblivious algorithms such as Van-Emde-Boas search tree, Funnel-sort, and Median of medians algorithm.
- Analyzed the memory-transfer complexity of the above algorithms. Used Valgrind to calculate the cache hit-miss ratio of cache oblivious algorithms and compared it with their cache ignorant counterparts.

## RELEVANT COURSEWORK

---

- **Computer Science:** Computer Vision, Cryptography & Network Security, Advanced Algorithms, Machine Learning, Compiler Techniques, Computational Intelligence, Computer Networks, Computer Graphics & Visualization, Optimization Algorithms & Techniques, Operating Systems, Computer Architecture, Parallel Computing, Design & Analysis of Algorithms, Software Engineering, Automata Theory & Logic, Logic Design, Data Structures & Algorithms, Database & Information Systems<sup>†</sup>, Discrete Mathematical Structures.
- **Mathematics:** Numerical Methods<sup>†</sup>, Complex Analysis & Differential Equations II, Linear Algebra & Differential Equations I, Calculus.
- **Other:** Psychology<sup>†</sup>, Physics<sup>†</sup>, Chemistry<sup>†</sup>.
- **AI for Medicine (Coursera):** AI For Medical Diagnosis, AI For Medical Prognosis & AI For Medical Treatment.
- **Quantum Computation (David Deutsch).**

(<sup>†</sup> indicates an exceptional grade (AP) achieved in the course.)

## TECHNICAL SKILLS

---

- **Programming:** Python, C++, C, POSIX C, Go, Java, SQL, VerilogHDL, HTML, CSS.
- **Tools/Libraries:** PyTorch, TensorFlow, Container technology (Podman, Docker), LaTeX.

## POSITIONS OF RESPONSIBILITY / EXTRA CURRICULAR ACTIVITIES

---

- **Head, Cynaptics Club (AI and ML student club of IIT Indore):** Led, mentored, and taught club members and students. Organized competitions and administered tutorials. *Aug 2020 - Jun 2021*
  - **Department Undergraduate Representative,** Department of Mathematics. *Feb 2020 - Oct 2020*
-