

ARYAN SOOD

Gurugram, Haryana, India – 122018

📞 +91-7838668881 ✉ aryan_s2@ee.iitr.ac.in 🌐 Aryan Sood 📄 zer0-data 🎓 Google Scholar

Education

Indian Institute of Technology Roorkee
Bachelor of Technology in Electrical Engineering
Experience

CGPA: 8.175/10 (3.538/4)*
Aug. 2024 – 2028

Auric AI Labs

March 2026 – May 2026

Computer Vision Intern

- Worked on generating 2D renders from 3D meshes of military objects using **Blenderproc** and further enhancing their quality using EO data.
- Created a Semantic Suggestion Pipeline via **re-utilizing the methodology proposed in the paper Seeing The Unseen for EO data**, and further implementing deterministic placement rules for military objects in satellite data.
- Trained a **Remote-CLIP-UNet** (frozen CLIP text/image encoders with a UNet decoder) to predict where unseen objects could be placed in an image, with **LLaVA**.

Publications

More Than a Quick Glance: Overcoming the Greedy Bias in KV-Cache Compression

ICLR 2026

[Preprint](#) | [GitHub](#) | [ICBINB Workshop](#)

- *Authors:* **Aryan Sood**, Tanvi Sharma, Vansh Agrawal
- Developed LASER-KV, a KV-cache compression method achieving up to **10% higher accuracy at 128k context** over greedy pruning baselines, via block-wise accumulation and **Exact-LSH sampling**

Dropping the Anchor: Statistical Summarization for Distributed Systems via Pulsar Attention

ICML 2026

[Preprint](#) | [GitHub](#) | [AdaptFM Workshop](#)

- *Authors:* **Aryan Sood**, Shantanu Acharya
- Collaborated with a researcher at NVIDIA, to further develop StarAttention (a framework for distributed systems). Pulsar Attention enables **cross block communication using lightweight summaries** which lead to a **4.97%** improvement at long context over the full attention baseline.

Projects

ICLR Blogpost on Sparsity | Accepted at ICLR 2026 | [Blog](#)

Dec. 2025

- Wrote on various efficiency methods based on sparsity and categorized these efficiency techniques into a cohesive taxonomy based on their reliance on static structural motifs versus dynamic, data-driven attention behaviors
- Analyzed emergent attention patterns and their interplay with training stability, demonstrating how structural regularities in the attention matrix can be leveraged to design sparse approximations that preserve model fidelity without sacrificing convergence behavior

Sparse Attention Pre-Selection | [GitHub](#)

Sep. 2025

- Designed a neural framework that reduces autoregressive inference overhead in LLMs by dynamically selecting relevant context tokens per query
- Incorporated a layer-wise sparse attention mask by adapting TidalDecode principles, substantially **reducing per-step computational complexity**

GATE-IR: Gated Adaptive Thermal Enhancement for Infrared Detection | [GitHub](#)

Jan. 2026

- Developed a weather-adaptive thermal object detection pipeline combining condition-aware enhancement, **CycleGAN-based IR translation**, and a modified **YOLOv8** architecture with **ViT neck** and **P2 head** for robust small-object detection in adverse weather.

Adaptive Momentum-based PD Controller for Diffusion LLMs | [GitHub](#)

Dec. 2025

- Built a training-free **monotonic masking framework** for Diffusion LLMs, applying momentum to project discrete token spaces into continuous domains
- Implemented a momentum heuristic that drives embeddings off-manifold, achieving **earlier convergence** on GSM8K and characterizing hallucination lock-in on MATH-500

GNSS Anti-Spoofing Detection | [GitHub](#)

Mar. 2026

- Built a Transformer encoder with 8-channel early fusion over 14 features per satellite, detecting GPS spoofing by learning cross-satellite signal inconsistencies; achieved **Val F1 of 0.9798** across 5-fold CV
- Engineered paper-inspired features (CNO spatial/temporal std, EML discriminator) capturing the physical signature of single-source spoofing; trained with AdamW, BCELoss, and per-fold early stopping

Achievements

- **All India Rank 1519, JEE Advanced 2024** | *Top exam for IIT admission, taken by 150k+ students nationally*
- **All India Rank 2912, JEE Main 2024** | *Qualifying exam for JEE Advanced, taken by 1M+ students nationally*
- **Top 20 of 1000+ Projects** | *Recognized for exceptional technical implementation among IIT Roorkee peers*
- **2nd Place, PixelPlay 2025** | *Week-long computer vision hackathon among top AI talent at IIT Roorkee*

Extracurricular

Vision and Language Group, IIT Roorkee

Feb. 2025 – Present

Core Member | [VLG Website](#) | [GitHub](#)

IIT Roorkee

- Active member of a deep learning research group focused on computer vision and language modeling, contributing to collaborative projects and technical discussions