Soumyajit Karmakar

soumyajit.karmakar.cse@gmail.com | GitHub | LinkedIn | Google Scholar
soumyajitkarmakar.github.io

Education

University of California San Diego 2024 - 2026
M.S. in CS | **GPA: 3.92/4.0**
Advisor: Mikhail Belkin
Thesis: Learning Relations Between Steering Vectors

Indian Institute of Information Technology, Guwahati 2019 - 2023
B.Tech. in CSE | **CGPA: 9.35/10** | **Silver medalist**
Co-advisor: C. V. Jawahar, IIIT Hyderabad

Research Experience

Belkin Lab, UC San Diego Sep 2024 - Present
Graduate Research Assistant
Advisor: Mikhail Belkin

- Developed linear transformations for **LLM activation steering** across semantic opposites, layers, and translation tasks, generalizing to held-out concepts.
- Showed that on Llama 3.1 8B a single learned linear map converts a steering vector for one concept (e.g., cheerful) into one for its antonym (gloomy), scoring 2.76/3 vs. 2.73/3 for directly extracted vectors, as compared to 1.98/3 for simple negation. Consistent results on Gemma 2 9B and Mistral NeMo 12B.

University of North Carolina at Charlotte Aug 2022 - Mar 2023
Research Intern
Advisors: Srijan Das and Michael S. Ryoo

- Co-developed **Self-Supervised Auxiliary Task (SSAT)** training, jointly optimizing classification and masked autoencoder reconstruction with a shared ViT encoder and **no added inference overhead**.
- Benchmarked ViT, CvT, and Swin on **10 datasets**; raised ViT-S CIFAR-100 top-1 accuracy from **54.08% to 73.37% (+19.29 percentage points)** under equal 100-epoch training.

CSIR-CEERI Pilani, India May 2022 - July 2022
Research Intern at the Advanced Information Technologies Group (AITG)
Advisor: Sanjay Singh

- Developed a few-shot classification framework that fuses features from frozen pre-trained CNNs (ResNet-50, EfficientNet-B5, DenseNet-201) through a lightweight convolutional ensembling block with 2.1M trainable parameters.
- Achieved 92.30% 5-way 5-shot accuracy on a drone-captured power-line defect dataset, and outperformed the library-learner baselines on 5 of 6 standard benchmarks (e.g., FC100: 91.4% vs. 80.5%).

Service: Reviewer at ICVGIP 2022, AAAI 2027.

Industry Experience

Fujitsu Research, Bengaluru Sep 2023 - Aug 2024
ML Engineer | Collaborator: R. Venkatesh Babu, IISc Bengaluru

- Co-developed **Prompt-ReID**, a CLIP-based person and vehicle re-identification framework using learned identity-specific text prompts; achieved **91.1% Rank-1 accuracy** on MSMT17 (**+2.7% points** over CLIP-ReID).
- Evaluated robustness to domain shifts via leave-one-camera-out testing; improved over CLIP-ReID by **10.1 percentage points** on average across multi-source person re-identification benchmarks.

Open-source contribution: CompilerGym.

Publications

- **Soumyajit Karmakar**, Neil M., Mikhail B. Relational Structure Between Steering Vectors in Large Language Models. (under review at AAAI 2027)
- **Soumyajit Karmakar***, Shrinivas R.*, Harsh R.*, Yasuhiro A., Genta S., Venkatesh B. Prompt-ReID: Prompt Based Inference for Robust Re-Identification. (Technical Report)
- Srijan D., Tanmay J., Dominick R., Pranav B., **Soumyajit Karmakar**, Shyam M., Xiang L., Abhijit D., Michael R. Limited Data, Unlimited Potential: A Study on ViTs Augmented by Masked Autoencoders. **WACV 2024**.
- **Soumyajit Karmakar**, Abeer B., Prashant G., Sumeet S., Sanjay S. Convolutional Ensembling based Few-Shot Defect Detection Technique. **ICVGIP 2022**.
- Aman J., Sachin A., Nimish A., **Soumyajit Karmakar**, Srinibas S. Feedback vertex set using Edge Density and REmove Redundant (FEDRER): A heuristic solver for finding a feedback vertex set in a directed graph. **PACE 2022 solver description**.

Achievements

Best student submission award, Parameterized Algorithms and Computational Experiments (PACE) 2022 Heuristic Track (Directed Feedback Vertex Set), with solver FEDRER.

Projects

Model Insurance
Undoing Edits to Trained Models with Little Protected Information
Independent research project
Github

- Formulated the cost of guaranteeing that any edit from a given class can be undone, and showed it is set by the coloring number of a graph of confusable models, which error-correcting codes achieve up to constant factors. Any 64 changed weights in Llama 3 8B's 525M-weight output layer can be undone with a 464-byte certificate, versus about 1 GB for a clean copy.
- Showed that a record of the model's outputs lowers this cost. It drops to zero for sparse bias edits in one-hidden-layer ReLU networks once the record covers every neuron's isolating input region. For rounded final-layer outputs, it drops by about 2 bits each time the record doubles, with matching lower bounds.

Meaningless Tags, Meaningful Impact
Formatting Decides Context-Memory Conflicts in LLMs
Independent research project
Github

- Showed that wrapping a false, source-attributed claim in <document> tags flips Gemma 4 12B from 0% to 100% adoption of the false answer (log-prob margin shift of +41). That matches an explicit instruction to adopt it, and even a meaningless tag (<qzx_block>) reaches 75–79%.
- Designed 16 controlled wrappers over fixed content to separate delimiter syntax from tag semantics. The wrapper rankings replicate on Qwen 3.6 27B (Spearman $\rho = 0.79$–0.88), though Qwen's effects are about 5× smaller.

Technical Skills

Languages & Tools: Python • C++ • PyTorch • JAX • Git • LaTeX.
ML & AI: LLM Interpretability • Activation Steering • Vision Transformers • CLIP • Masked Autoencoders • Self-Supervised Learning.
Methods: Linear Probing • Recursive Feature Machines • PCA / SVD • Low-Rank Approximation • Hessian Spectral Analysis.