Armaan Sandhu

apsandhu@umass.edu · +1 (413) 425-3387 · linkedin.com/in/armaansandhu26 · github.com/armaansandhu26

EDUCATION

University of Massachusetts Amherst Aug 2025 – May 2027 (Expected)
MS in Computer Science, GPA 3.88/4.0
Relevant Coursework: Reinforcement Learning, AI Alignment, Deep Learning Systems, Machine Learning

Indian Institute of Technology Madras Aug 2016 – May 2021
BTech + MTech in Biological Engineering, Computational Biology specialization, GPA 8.99/10
Master's Thesis Guide: Dr. Hamsa Priya, BioSim Lab | Teaching Assistant, Life Sciences (Fall 2020 and Spring 2021)
Semester Exchange, Instituto Superior Técnico, Lisbon (Spring 2019)

PUBLICATIONS

Armaan Sandhu, Suyash Maniyar, Abhishek Mishra. "Measuring Reward Hacking and Reasoning-Answer Decoupling Under Position-Confounded Optimization." AI Measurement Science (AIMS) Workshop, COLM 2026.

Armaan Sandhu, Abhilasha Senapati, Hima J Kammachi. "Targeting the Attention Heads Behind Object Hallucination in LLaVA." Actionable Interpretability (AIW) Workshop, COLM 2026.

RESEARCH EXPERIENCE

Reducing Mode Collapse in Reinforcement Learning May 2026 – Present
Research Internship: Artificial Scientific Intelligence Lab, National University of Singapore *Advisor: Dr. Dianbo Liu*

- Designed an inverse probability scaling correction for GRPO that prevents RL fine-tuning from collapsing onto a few high-reward modes, especially for small group sizes.
- Validated on phylogenetic tree inference and reaction synthesis (many-to-one DAGs), matching target reward distributions near-exactly: Pearson 0.98 (vs. 0.32 for GRPO, 0.88 for GFlowNet) with ∼1M unique 10-taxon trees.

Localizing and Mitigating Object Hallucination in Vision-Language Models Feb 2026 – May 2026
Course Project: Neural Networks - A Modern Introduction *Advisor: Dr. Subhransu Maji*

- Investigated whether object hallucination in LLaVA-1.5-7B is concentrated in specific attention heads, using one-head zero-ablation to causally identify a sparse set of 32 out of 1,024 heads most responsible for it.
- Fine-tuned the identified heads using LoRA and DPO and added an inference-time grounding controller; reduced caption-level object hallucination by 38% on COCO, compared with 28% from LoRA alone and 16% from grounding alone.

Measuring Reward Hacking and CoT Faithfulness Under Confounded Rewards Feb 2026 – Jun 2026
Course Project: AI Alignment *Advisor: Dr. Scott Niekum*

- Fine-tuned Qwen and Llama models with GRPO on a positionally confounded MCQ dataset where option A was always correct; models learned the shortcut, with A-answer rate rising from ∼0.25 to ∼0.95 as held-out accuracy fell to chance.
- Measured reasoning–answer decoupling under reward pressure: reasoning traces were often correct while final answers followed the shortcut, reaching up to 66% decoupling on Qwen2.5-3B and indicating degraded CoT faithfulness.

Anchoring Dynamics in Multi-Agent LLM Systems May 2026 – Present
Research Assistantship: Resource Bounded Reasoning Lab, UMass Amherst *Mentor: Mason Nakamura*

- Designing a multi-agent experiment to test whether giving one LLM agent domain expertise makes it dominate a collaboration. A moderator picks each round's leader from both agents' pitches, with rewards tied to task outcomes.
- In a 6-task pilot, one agent won 5 of 6 leader roles because the moderator kept re-selecting it on "recent calibration" and over-penalized the other agent's single failure, surfacing that influence can track reputation momentum.

FlashAttention: Implementation and Roofline Analysis Feb 2026 – May 2026
Course Project: Systems for Deep Learning *Advisor: Dr. Shiqing Ma*

- Implemented FlashAttention from scratch in Triton (GPU) and Pallas (TPU). Verified correctness against PyTorch autograd and achieved 6.6× forward speedup with 176× lower memory use over naive attention.
- Diagnosed naive attention as memory-bound using roofline analysis on an A100 GPU. Integrated the kernel into GPT-2-124M JAX training and achieved 1.6× end-to-end throughput over the XLA baseline.

INDUSTRY EXPERIENCE

Google— AI Acceleration Team: LLM Training-Data Compliance Infrastructure
Software Engineer Jul 2024 – Jul 2025

- Built and operated systems to enforce data compliance for Google-scale LLM training pipelines, ensuring models consume only policy-compliant user data. These systems scan 120T+ rows and issue 300B+ actions daily across 9,000+ clients.
- Unified onboarding into a storage-agnostic framework, standardizing metadata and policy-spec processing across Spanner, Colossus, Bigtable, BigQuery, and Blobstore.
- Launched Colossus Table and Bigtable support on this layer, reducing onboarding time by 85% (2 weeks → 3 days) and expanding compliant LLM training-data coverage.

Goldman Sachs— Reconciliation Engineering Team: Large-Scale Trade Data Processing & Distributed Systems
Associate / Analyst Jul 2021 – Jun 2024

Pre-placement offer from a summer internship; earned accelerated promotion to Associate (1.5 years vs. typical 2.5)

- Owned a Java-based distributed reconciliation system processing daily trade data across Americas, EMEA, and APAC (300+ downstream users); mentored 2 junior engineers.
- Re-architected a two-stage batch pipeline into a single stage by pushing join logic upstream, cutting ∼20 min per run across 50+ daily runs (∼25% lower end-to-end latency).
- Root-caused recurring record-matching failures in the pipeline and shipped targeted fixes that cut unmatched daily volume by 37%.

Curations (link)
Co-Founder Apr 2024 – Present

- Building an AI-native social platform that converts user-curated items into structured taste graphs, using embeddings and semantic retrieval to move discovery beyond keyword search and follower-based feeds.
- Led product from 0→1 to 600+ users and 1,500+ curated items; reached #2 Product of the Day on Product Hunt.

TECHNICAL SKILLS

Programming Languages: Python, C++, Java, JavaScript, SQL
Deep Learning Frameworks: PyTorch, TensorFlow, Triton, Pallas, JAX
Libraries & Tools: NumPy, Pandas, Scikit-learn, Git, Docker, Apache Spark, Apache Kafka, Flume, Alloy/Legend
Research Methods: PPO, GRPO, RLHF/DPO, LoRA fine-tuning, mechanistic interpretability, evals & benchmarking