Nishant Bhargava

github.com/nishant9709 linkedin.com/in/nishant-bhargava97 bharga57@purdue.edu (646) 239-1100

EDUCATION

Purdue University Expected May 2028
B.S. in Computer Engineering + (Stats, Math) minor GPA: 3.69/4.0
Relevant Coursework: Multivariate Calculus, Linear Algebra, Statistics, Python for Data Science

Publication

Bhargava, N., et al. "Probe–Rewrite–Evaluate: Quantifying Evaluation Awareness in Large Language Models." **NeurIPS** 2025 Workshop on Evaluation for Language Models & Wordplay Workshop @ **EMNLP** 2025.
arXiv: 2509.00591 - https://arxiv.org/abs/2509.00591

EXPERIENCE

Undergraduate Research Assistant @ CYNICS Lab Purdue Sep 2025 – Present

- Developed a transformer-based linguistic steganography pipeline that perturbs next-token distributions with <2.0 PPL impact while embedding encrypted payloads.
- Built a custom Chrome extension and forked messaging client that decodes hidden watermarks in real-time, achieving 100% bit recovery rate across 300+ test conversations.
- Investigating steganographic techniques on transformer-based semantic search, demonstrating persistence of hidden text through vector embedding transformations.

Project Lead & Mentor @ The LAB (let all build) Company Oct 2025 – Present

- Mentoring 4 junior researchers on investigating "Attention Entropy in Non-Standard Dialects," analyzing how non-native syntax causes attention-head dispersion in Transformer middle-layers and triggers "reasoning collapse" in complex logic tasks.
- Developing behavioral analysis models on AWS SageMaker by processing chat logs from the lab's internal collaboration platform to quantify team dynamics and predict research bottlenecks.
- Established a structured publication roadmap for the cohort, instituting weekly technical review sessions to guide researchers from hypothesis formulation to final LaTeX drafting.

AI Safety & Interpretability Researcher @ Algoverse May 2025 – Nov 2025

- Developed Probe-Rewrite-Evaluate methodology to quantify evaluation awareness in LLMs, revealing how models alter behavior between test-like and deployment contexts, forming the basis of a NeurIPS'25 & EMNLP'25 workshop paper.
- Introduced Awareness Elasticity metric to quantify LLM sensitivity to evaluation interventions, providing first systematic framework for measuring benchmark-deployment performance gaps.

Computer Vision Engineer @ RoBoat: Autonomous Maritime Maneuvers Jan 2025 – May 2025

- Led the AI and perception team for Purdue Indy's autonomous maritime robotics project in the AIMM ICC competition.
- Implemented deep learning–based object detection models (YOLOv5) to identify buoys, docks, and obstacles, achieving 93% detection accuracy in simulation and 88%+ in real-world lake tests.
- Reduced false positives by 34% through dataset curation, augmentation, and transfer learning on maritime-specific image data.

SKILLS

Languages: Python, R, C, SQL, Bash, MATLAB
ML Stack: PyTorch, sklearn, HuggingFace, transformer-lens, NumPy, pandas, OpenCV
Systems/Tools: Linux, AWS SageMaker, Git, Docker, CUDA