# Harsh Nishant Lalai

*Incoming PhD Student at Johns Hopkins University*


hlalai1@jh.edu | +91 9765477123 | harsh-nishant-lalai
harshlalai | Harsh_N_Lalai10 | Harsh-Lalai


## About me

I'm interested in trustworthy AI, focusing on language models under uncertainty and multi-agent social interactions. My goal is to develop rigorous evaluations and robust methodologies that enable increasingly capable AI systems to be deployed safely and reliably in real-world settings.

## Education

Aug 2026 – Present
**Johns Hopkins University**, Baltimore, MD
*PhD in Computer Science*
Advisor: Kristina Gligorić, Co-advisor: Benjamin Van Durme

Oct 2021 – Jul 2026
**Birla Institute of Technology and Science, Pilani, Goa Campus**, India
*Bachelor of Engineering in Computer Science, Master of Science in Economics*
GPA: 9.40 / 10
Undergraduate Thesis at Georgia Institute of Technology: *LLMs' Intermediate Thinking Tokens Are Weak Predictors of Human Cognitive Effort in Scientific Misconception Judgments*

## Research Experience

Jul 2025 – May 2026
**Georgia Institute of Technology**, Atlanta, GA
*Undergraduate Thesis Researcher*
Mentors: Sashank Varma, Raj Sanjay Shah
Paper: *LLMs' Intermediate Thinking Tokens Are Weak Predictors of Human Cognitive Effort in Scientific Misconception Judgments*
Investigating whether LLM reasoning-effort signals (thinking tokens) correspond to human cognitive effort on scientific misconceptions.
Submitted to ACL Rolling Review (May 2026).

Jun 2025 – Mar 2026
**Harvard University**, Cambridge, MA
*Research Collaborator, Visual Computing Group*
Mentors: Hanspeter Pfister, Grace Guo
Paper: *When Visuals Aren't the Problem: Evaluating Vision-Language Models on Misleading Data Visualizations*
Studying VLM robustness to misleading chart and caption pairs and the ability to detect fine-grained reasoning and visualization errors.
Submitted to ACL Rolling Review (March 2026).

May 2024 – Mar 2025
**Emory University / Brock University**, Atlanta, GA
*Visiting Scholar*
Mentors: Ali Emami, Yi-Chia Wang
Paper: *The World According to LLMs: How Geographic Origin Influences LLMs' Entity Deduction Capabilities*
Evaluating implicit geographic bias in LLMs through a multi-turn entity deduction game. Used notable people and culturally significant objects from diverse regions as entities, and investigated the effect of pre-training frequency and popularity on deduction.
Presented at COLM 2025.

Dec 2023 – May 2024 | **Pennsylvania State University**, University Park, PA
*Research Intern, PIKE Group*
Mentor: Dongwon Lee
Paper: *From Intentions to Techniques: A Comprehensive Taxonomy and Challenges in Text Watermarking for Large Language Models*
Analyzed text-watermarking research by intentions, evaluation datasets, and addition / removal methods to construct a unified taxonomy, and highlighted gaps and open challenges in protecting text authorship.
Presented at NAACL 2025 Findings.

## Employment

Jan 2026 – Jun 2026 | **SingleStore**, Hyderabad, India
*Software Engineering Intern, AI Team*
Building an experimental semantic retrieval benchmark for SingleStoreDB, comparing transformer-based encoders (e.g., MiniLM, E5, Cohere Embeddings) and vector-index configurations.

May 2023 – Jul 2023 | **Jio Brain (formerly Jio Platforms)**, Mumbai, India
*Software Engineering Intern, Automated Transcription Team*
Developed preprocessing utilities and evaluation scripts for JioBrain's Speech-to-Text module, automating dataset ingestion and inference-quality testing across multilingual pilot datasets for Hindi, Gujarati, Marwari, Marathi, and Tamil.

## Honors and Awards

2025 | **K. M. Birla Foundation Student Travel Grant**
Awarded for travel to COLM 2025. Given to selected students with an accepted first-author paper at a top-tier international conference outside India.

2024 | **MITACS Globalink Research Internship Scholarship**
Selected from over 30,000 applicants for top Canadian research universities. Fully funded research internship at Brock University under Professor Ali Emami.

2023 – 2026 | **Merit Scholarship, BITS Pilani**
Awarded for six consecutive semesters to the top 1% of approximately 800 students for academic performance.

## Publications

2026
1. **Lalai, H. N.**, Shah, R. S., Pfister, H., Varma, S. & Guo, G. When Visuals Aren't the Problem: Evaluating Vision-Language Models on Misleading Data Visualizations. *arXiv preprint* (2026).

2025
2. **Lalai, H. N.**, Ramakrishnan, A. A., Shah, R. S. & Lee, D. *From Intentions to Techniques: A Comprehensive Taxonomy and Challenges in Text Watermarking for Large Language Models* in *Findings of the Association for Computational Linguistics: NAACL* (2025).
3. **Lalai, H. N.** *et al.* The World According to LLMs: How Geographic Origin Influences LLMs' Entity Deduction Capabilities. *Second Conference on Language Modeling (COLM)* (2025).

## Open Source Contributions

I believe in open-source development for reproducibility and impact.

**MisVisBench Dataset**
MisVisBench is a dataset of carefully curated visualizations spanning diverse visualization and reasoning errors, enabling systematic assessment of model robustness to deceptive visual communication and analytical mistakes.

**Geo20Q+ Dataset**
Geo20Q+ is a geographically balanced dataset designed to evaluate implicit geographic biases in LLMs through a multi-turn deduction game inspired by 20 Questions. It includes entities of two types, Notable People and Culturally Significant Things, from diverse regions across the globe.

**TwentyQ: The 20 Questions Game Engine (Python Package)**
A pip package that lets users simulate and evaluate a 20 Questions-style guessing game using LLMs such as OpenAI GPT, Anthropic Claude, Google Gemini, and Huggingface-hosted models.

**A Taxonomy of Text Watermarking for LLMs (NAACL 2025)**
An open-source, continuously updated repository that organizes the text-watermarking literature by intentions, methods, and evaluation datasets, serving as a community reference for researchers and practitioners.

## Teaching

Fall 2024 | **Undergraduate Teaching Assistant**
*Birla Institute of Technology and Science, Pilani, Goa Campus*, India
International Economics (ECON F311). Created a term-paper-based assignment to give students hands-on research exposure.

Spring 2024 | **Undergraduate Teaching Assistant**
*Birla Institute of Technology and Science, Pilani, Goa Campus*, India
Econometric Methods (ECON F241). Designed coding exercises to give students hands-on experience with R and Stata.

Fall 2023 | **Undergraduate Teaching Assistant**
*Birla Institute of Technology and Science, Pilani, Goa Campus*, India
Financial Management (ECON F315). Created homework, assignments, and assisted students with their capstone projects.

Spring 2023 | **Undergraduate Teaching Assistant**
*Birla Institute of Technology and Science, Pilani, Goa Campus*, India
Principles of Economics (ECON F211). Created homework and assignments for the course.

## Service

**Reviewing:** ACL Rolling Review (**ARR**) 2026, 2025; Conference on Language Modeling (**COLM**) 2026; International Conference on Mathematics and Computing (**ICMC**) 2023.

**Volunteering:** International Conference on Mathematics and Computing (**ICMC**) 2023; International Conference on Data-Driven Computing and Intelligent Systems 2023; International Conference on Advances in Data-Driven Computing and Intelligent Systems 2022.

## Technical Skills

**Proficient:** Python, C, C++, PyTorch, TensorFlow, Scikit-learn, Pandas
**Comfortable:** Java, LaTeX, Git, NLTK
**Familiar:** R, MySQL, HTML, CUDA

## Languages

**Native Proficiency:** English, Hindi, Gujarati, Kutchi
**Expert Proficiency:** Marathi

## Hobbies

**Sports & Fitness:** Football, Cricket, Mixed Martial Arts, Gym Training, Badminton, Table Tennis
**Reading:** Autobiographies
**Media & Entertainment:** Sports, Sports Documentaries, Action & Thriller Movies and TV Series, e-Sports

## References

**Dr. Kristina Gligorić**, Assistant Professor
Department of Computer Science, *Johns Hopkins University*

**Dr. Benjamin Van Durme**, Professor
Department of Computer Science, *Johns Hopkins University*

**Dr. Sashank Varma**, Professor
School of Interactive Computing, *Georgia Institute of Technology*

**Dr. Yi-Chia Wang**, Principal Research Scientist
*Carnegie Mellon University*. Visiting Scholar, *Stanford University*