Yash Saxena

✉ ysaxena1@umbc.edu | 🏠 yashsaxena21.github.io/Portfolio/ | ⬡ YashSaxena21 | in yash-saxena-a18b251bb | Scholar

Research Focus

Building trustworthy LLM agents through interpretable, robust knowledge retrieval and verifiable source attribution in generated responses.

Skills

RAG/IR	Chunking & Query Reformulation; Dense/Sparse Retrieval; Selection/Re-ranking; FAISS/HNSW
LLM Tuning	Instruction Tuning (SFT); Preference Tuning (PPO, DPO, GRPO); PEFT (LoRA/QLoRA)
Tools/Libraries	PyTorch; Hugging Face (Transformers, PEFT); LlamaIndex; LangChain; FAISS; SQL; Git

Education

University of Maryland, Baltimore County — *Baltimore, Maryland, USA*

PHD, COMPUTER SCIENCE — *Jan 2025 - Present*

- Research Assistant | GPA: 4.0/4.0

Galgotias University — *Greater Noida, Uttar Pradesh, India*

B.TECH, COMPUTER SCIENCE AND ENGINEERING WITH SPECIALIZATION IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING — *Nov 2020 - May 2024*

- CGPA: 8.75/10

Bishop Conrad Sr. Sec. School — *Bareilly, Uttar Pradesh, India*

SENIOR SECONDARY (XII) — *May 2018 - May 2019*

- Percentage: 83.20%

Professional Experience

Knowledge Infused AI And Inference Lab, University Of Maryland Baltimore County — *Baltimore, MD, USA*

REMOTE RESEARCH INTERN — *Aug 2024 - Jan 2025*

- Fine-tuned Llama-3-70b using DPO algorithm as well as built a RAG pipeline
- Designed the experiments for the ongoing research
- Worked with Slurm to use the UMBC's GPU cluster

Stareout Games — *Remote*

AI ENGINEER INTERN — *Jan 2024 - Mar 2024*

- Built a pipeline incorporating Large Language and Image Generation Model to bring game ideas to life
- Fine-Tuned Large Language and Image Generation Models
- Worked with python libraries like Langchain, Ntscraper, Streamlit , and many more

Artificial Intelligence Institute, University of South Carolina (AIISC) *Columbia, SC, USA*

REMOTE RESEARCH INTERN *Aug 2023 - Jul 2024*

- Worked on LLMs using libraries like Langchain and LLAMA Index
- Built a web scraper to construct the main dataset of the ongoing research

Celebal Technologies *Jaipur, Rajasthan, India*

REMOTE DATA SCIENCE INTERN *May 2023 - July 2023*

- Worked with large datasets, performed data pre-processing and extracted relevant features tailored to specific use cases
- Worked on NLP (Natural Language Processing) problems, built client specific applications using libraries like spacy

HCLTech *Noida, Uttar Pradesh, India*

INTERN *Nov 2022 - Jan 2023*

- Developed a Python-based Web app using spacy and streamlit libraries to extract text from resumes
- Created a web app that allowed users to filter resumes according to their specific requirements

Publications

PUBLISHED/ACCEPTED

Yash Saxena, Raviteja Bommireddy, Ankur Padia, Manas Gaur. "Generation-Time vs. Post-hoc Citation: A Holistic Evaluation of LLM Attribution". Accepted at **NeurIPS 2025 LLM Evaluation Workshop**. Link

Yash Saxena, Sarthak Chopra, Arunendra Mani Tripathi. "Evaluating Consistency and Reasoning Capabilities of Large Language Models". **2024 Second International Conference on Data Science and Information System (ICDSIS)**, Hassan, India, 2024, pp. 1–5. Link

Seyedreza Mohseni, Seyedali Mohammadi, Deepa Tilwani, **Yash Saxena**, Gerald Ketu Ndawula, Sriram Vema, Edward Raff, Manas Gaur . "Can LLMs Obfuscate Code? A Systematic Analysis of Large Language Models into Assembly Code Obfuscation". **Proceedings of the AAAI Conference on Artificial Intelligence**, 39(23), 24893-24901. Link

Yash Saxena, Aman Kumar Mishra, Daksh Arora, Runumi Devi. 2023. "Emotion Based Mental Health Classifier for NCR Based Engineering Students". IEEE 6th **International Conference on Contemporary Computing and Informatics (IC3I)**, Page(s):285-290. Link

UNDER REVIEW

Saxena et al. "Ranking Free RAG: Replacing Reranking with Selection in RAG for Sensitive Domains". (Communicated to **ICLR 2026**) Link

Saxena et al. "IMRNNs: Efficient Embedding Modulation for Interpretable Dense Retrieval". (Communicated to **EACL 2026**)

IN PREP

Saxena et al. "Attribution in Scientific Literature: New Benchmark and Methods".

Presentations

RECENT

Yash Saxena, Ankur Padia, Swati Padhee, Manas Gaur and Srinivasan Parthasarathy. June 2025. Title: "RASOR: Contextual Legal Intelligence via Rationalized Selection and Refinement in RAG". Venue: **Bloomberg Law, Language, and AI Symposium.**

Yash Saxena and Manas Gaur. October 2025. Title: "Building Trustworthy LLM Agents for Academia through Structured, Interpretable Knowledge Retrieval and Source Attribution". Venue: **UMBC Library 2025 AI Symposia.**

Recognitions, Fellowships, & Grants

2022   **Winner UNESCO India–Africa Hackathon**, Ministry of Education (India).

2022   **Winner Smart India Hackathon (Team Lead)**, Ministry of Education (India).