

SIDDHANT GUPTA

 sidworks01.github.io |  writetosiddhant@gmail.com |  +91-7836015759 |  huggingface.co/minemaster01

OBJECTIVE

AI researcher focused on LLMs, agent systems, alignment, and efficient modeling. Seeking to build scalable, reliable AI systems and contribute to research advancing the boundaries of modern NLP and multimodal reasoning.

EDUCATION

Indian Institute of Technology (IIT) Roorkee | *Bachelors of Technology* **Oct 2022 - July 2026**

WORK EXPERIENCE

Machine Learning Intern | **Observe.ai** **May 2025 – Nov 2025**

- Engineered a bias detection framework for contact center LLMs, benchmarking 18 models across 3,000+ transcripts using novel fairness metrics to quantify operational bias and failures and mitigating them increasing reliability.
- Designed confidence-aware scoring pipelines to enhance Auto-QA evaluation, reducing misclassification risk by 12–18% in pilot test and expanding the methodology to quantify uncertainty and optimize LLM outputs for accuracy.

NLP & ML-Agents Community Lead | **Cohere Labs** **Jun 2023 – Present**

- Led multiple research projects under **Expedition Aya (2024–25)** on reasoning evaluations, cultural bias in VLMs, speech synthesis & ASR, machine-generated text detection using fine-grained methods, and multilingual LLMs.
- Designed a DAG-based open-source agentic framework for **Cohere Labs** and collaborated globally on multi-agent evaluations and A2A orchestration and memory sharing modules as part of **ML-Agent** Program.
- Spoke at **Cohere ML Summer School** on Transformers and LLM evolution; drove 50+ speaker and paper reading sessions on NLP, agents, interpretability, LLM Optimization methods and synthetic data as part of the **NLP** Program.

Research Fellow | **Traversaal.ai** **Jan 2025 – Apr 2025**

- Benchmarking data-science agents - automated graph generation with ReAct based agent workflow and design

ACADEMIC EXPERIENCE

Researcher | **Massachusetts Institute of Technology (MIT)** **Nov 2025 - Present**

- Contributing to Project Mantis by developing a JSON-based workflow framework that enables agents to translate natural-language descriptions into complex computation graphs for creating task specific agentic workflows.
- Building the Mantis subsystem to collect and expand user-generated insights into structured datasets, forming a benchmarking suite for evaluating individual LLMs and multi-agent workflows used across the application.

Software Head | **Artificial Intelligence and Electronic Society (ArIES)** **May 2023 - Present**

- Headed teams representing my college society in AI hackathons, mentoring CV and NLP research alignment, leading to 10+ successful projects including leading Inter-IIT Tech teams and organizing workshops on deep learning.

PUBLICATIONS

Uncovering Cultural Representation Disparities in Vision-Language Models

Siddhant Gupta, Ram Mohan Rao Kadiyala, Jebish Purbey, Srishti Yadav, Alejandro Salamanca, Desmond Elliott

arxiv.org/abs/2505.14729 - Accepted at **AACL 2025 - Findings**

Multilingual Hate Speech Detection and Target Identification in Devanagari-Scripted Languages

Siddhant Gupta, Siddh Singhal, Azmine Toushik Wasi

arxiv.org/abs/2412.17947 - Accepted at **Chipsal @ COLING 2025**

SeQwen at the Financial Misinformation Detection Challenge

Jebish Purbey, Siddhant Gupta, Nikhil Manali, Siddartha Pullakhandam, Drishti Sharma, Ashay Srivastava, Ram Mohan Rao Kadiyala

arxiv.org/abs/2412.00549 - Accepted at **FinNLP-FNP-LLMFinLegal @ COLING 2025 (3rd position)**

Spot the BlindSpots: Quantifying Fine-Grained LLM Biases in Contact Center Summaries

Kawin Mayilvaghanan, Siddhant Gupta, Ayush Kumar

aclanthology.org/2025.emnlp-industry.91/ - Accepted at **EMNLP Industry Track 2025 – Work done at Observe.ai**

DSBC : Data Science task Benchmarking with Context engineering

Ram Mohan Rao Kadiyala, Siddhant Gupta, Jebish Purbey, Giulio Martini, Suman Debnath, Hamza Farooq

arxiv.org/abs/2507.23336 - Accepted at **AACL 2025 - Main**

If the Yardstick is Bent: Counterfactual Fairness Evaluation of LLM-Based Contact Center Agent Quality Assurance System

Kawin Mayilvaghanan, Siddhant Gupta, Ayush Kumar

arxiv.org/abs/TBD - Under Review at NAACL 2025 – Work done at Observe.ai

Robust and Fine-Grained Detection of AI Generated Texts

Ram Mohan Rao Kadiyala, Siddartha Pullakhandam, Siddhant Gupta, Jebish Purbey, Kanwal Mehreen, Drishti Sharma, et al.

arxiv.org/abs/2504.11952 - Under review at TMLR

Lexical Reranking of Semantic Retrieval (LeSeR) for Regulatory Question Answering

Jebish Purbey, Drishti Sharma, Siddhant Gupta, Khawaja Murad, Siddartha Pullakhandam, Ram Mohan Rao Kadiyala

arxiv.org/abs/2412.06009 - Accepted at RegNLP @ COLING 2025 (3rd position)

Improving Multilingual Capabilities in LLMs with Cultural and Local Knowledge

Ram Mohan Rao Kadiyala, Siddartha Pullakhandam, Siddhant Gupta, et al.

arxiv.org/abs/2504.09753 - Accepted at ACL 2025 - Main

Evaluating Generalization Capabilities of LLM-Based Agents in Mixed-Motive Scenarios Using Concordia

{20+ authors}, Siddhant Gupta

neurips.cc/virtual/2025/loc/san-diego/poster/121382 - Accepted at NeurIPS 2025

PROJECTS

Design Automation for Mechanical Workflows

B.Tech Thesis

Jul 2025 – Present

- Developed software automating mechanical design workflows through a single prompt and human feedbacks, including custom simulation models and generating product design & manufacturing and assembly documentation
- Built workflow modules integrated for engineering pipelines, enabling copilots to streamline iterative design tasks.
- Integrated MCP server to communicate workflow executions with engineering tools like AutoCAD, Ansys for analysis.

SpeechAya: Multilingual Speech Synthesis via Unified LLM Architecture

Open-Source Project – Cohere Labs Hackathon

- Engineered a multilingual LLM pipeline integrating speech and text modalities using 1000+ hours of audio.
- Optimized tokenization using MMS, mHuBERT and XEUS; achieved 0.112 WER on PolyAI/minds14 dataset with Qwen2-1.5b model; used CommonVoice and LibriSpeech for ASR, TTS, voice cloning and speech translation tasks

Advanced Attribute Extraction & Classification Pipeline

Amazon ML Hackathon 2024

- Applied OCR on 400K+ product images with 88% accuracy. Fine-tuned DistilBERT & LLaMA 3.2 for entity extraction.
- Optimized LayoutLM for attribute classification (weight, height, width), reducing misclassification by 10–15%.

DocAI: Smart Lab Report Generator

Hackathon Project

- Built full-stack Django app for doctors/patients; integrated traditional NLP-based report suggestions.
- Embedded chatbot and Plotly-based dashboard for real-time analytics and UX improvements.

TECHNICAL SKILLS

Languages: Python, C++, CUDA, Julia, JavaScript

Libraries and Frameworks: PyTorch, TensorFlow, HuggingFace Transformers, PEFT, TRL, BitsAndBytes, FlashAttention, xFormers, Triton, Sklearn, Librosa, NLTK, OpenCV, NumPy, Pandas, Matplotlib, Gradio, ONNX Runtime

Databases: PostgreSQL, SQL, SQLite

REFERENCES

Suman Debnath Principal Developer Advocate, Amazon, USA
suman.san14@yahoo.in

Ram Mohan Rao Kadiyala NLP Research Head, TraversaAI
<https://www.rkadiyala.com/>