Siddhant **Gupta**

Undergrad Researcher, Indian Institute of Technology Roorkee

sidworks01.github.io @ siddhant_g@me.iitr.ac.in github.com/SidWorks01 Google Scholar

Education

Oct 2022 – July 2026 | **Indian Institute of Technology Roorkee [🌍]** **Roorkee, India**
B.Tech student in Industrial Engineering
Coursework: Data Mining, Probability and Statistics, Calculus, C++

Experience

Observe.ai June 2023 – Present
ML Intern

- Co-authored a research paper submitted to EMNLP 2025 on bias in large language models (LLMs), proposing a taxonomy of bias dimensions and an evaluation framework for systematically measuring and addressing these biases.
- Conducted extensive research on autonomous agent systems, exploring orchestration strategies and practical use cases for integrating agents in real-world enterprise NLP workflows.
- Improved the evaluation pipeline of Automatic Quality-Assurance (Auto-QA).

Cohere Labs June 2023 – Present
Lead - NLP & ML-Agents | Research Lab and Open Science Community

- Engaged in 50+ technical discussions and speaker workshops on topics such as NLP, multi-agent systems, contextual learning, synthetic data generation, and interpretability, contributing to the community's knowledge base [NLP]
- Led efforts to introduce an open-source DAG-based agentic framework as part of ML-Agents program , where I collaborate with researchers globally to evaluate these agents as well [ML-Agents]
- Speaker at Cohere ML Summer School – *Introduction to Transformers and Evolution of LLMs* [Talk]
- Worked on a 8-week long hackathon Expedition Aya - 2024 where I developed speech synthesis method using ASR data.
- Led 5 research projects for the Expedition Aya - 2025 where I lead my team to develop reasoning evaluations, machine generated short text detection using fine grained methods, evaluation of cultural biases in VLMs, image caption dataset collection and improving LLMs for multilingual data across tasks, some of which reach publication.

Artificial Intelligence and Electronic Society (ArIES)) May 2023 – Present
Indian Institute of Technology, Roorkee | ML Executive

- Spearheaded teams in AI hackathons, providing mentorship in CV and NLP research alignment, leading to the successful implementation of 10+ innovative projects including participation in Inter-IIT competitions.
- Organized and conducted workshops and talks for 100+ participants, focusing on deep learning and image processing concepts such as edge detection, depth estimation, object detection and character recognition boosting technical proficiency across attendees.

M2ai February 2025 – Present
Research Head

- Led and coauthored multiple research projects and leading to publications in top conferences , prominent works being machine generated text detection and LLMs/VLMs for south-asian languages
- Working on improvement of STT & TTS models along with efficient audio tokenizers along with vertical agent tools and applications
- The group's projects run on several grants obtained over the past year totalling over 20000$ from OpenAI, Anthropic, Cohere etc..

Traversaal.ai April 2025 – Present
Research Fellow under Hamza Farooq

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

Computational Intelligence and Operations Lab (CIOL) September 2024 – November 2024
Research Collaborator

- Conducted research on hate speech detection across multilingual datasets using RoBERTa, addressing model bias and improving classification metrics and designed advanced solutions for Retrieval-Augmented Generation (RAG) tasks, enabling seamless integration of external knowledge retrieval improving F1@k, MRR, precision and recall.

Publications

[1] **Lexical Reranking of Semantic Retrieval (LeSeR) for Regulatory Question Answering** [🔗]
Jebish Purbey, Drishti Sharma, Siddhant Gupta, Khawaja Murad, Siddartha Pullakhandam, Ram Mohan Rao Kadiyala
[Accepted at RegNLP @ COLING 2025] [**4th position in workshop**]

[2] **SeQwen at the Financial Misinformation Detection Challenge Task: Sequential Learning for Claim Verification and Explanation Generation in Financial Domains** [🔗]
Jebish Purbey, Siddhant Gupta, Nikhil Manali, Siddartha Pullakhandam, Drishti Sharma, Ashay Srivastava, Ram Mohan Rao Kadiyala
[Accepted at FinNLP-FNP-LLMFinLegal @ COLING 2025] [**3rd position in workshop**]

[3] **Multilingual Hate Speech Detection and Target Identification in Devanagari-Scripted Languages** [🔗]
Siddhant Gupta, Siddh Singhal, Azmine Toushik Wasi
[Accepted at Chipsal @ COLING 2025]

[4] **Robust and Fine-Grained Detection of AI Generated Texts** [🔗]
Ram Mohan Rao Kadiyala, Siddartha Pullakhandam, Kanwal Mehreen, Drishti Sharma, Siddhant Gupta, Jebish Purbey, Ashay Srivastava, Subhasya TippaReddy, Arvind Reddy Bobbili, Suraj Telugara Chandrashekhar, Modabbir Adeeb, Srinadh Vura, Hamza Farooq
[EMNLP 2025 - Under Review] [—]

[5] **Uncovering Cultural Representation Disparities in Vision-Language Models** [🔗]
Siddhant Gupta, Ram Mohan Rao Kadiyala, Jebish Purbey, Srishti Yadav, Alejandro Salamanca, Desmond Elliott
[EMNLP 2025 - Under Review] [—]

[6] **Improving Multilingual Capabilities with Cultural and Local Knowledge in Large Language Models While Enhancing Native Performance** [🔗]
Ram Mohan Rao Kadiyala, Siddartha Pullakhandam, Siddhant Gupta, Drishti Sharma, Jebish Purbey, Kanwal Mehreen, Muhammad Arham, Hamza Farooq
[EMNLP 2025 - Under Review] [—]

[7] **Evaluating Generalization Capabilities of LLM-Based Agents in Mixed-Motive Scenarios Using Concordia** [🔗]
List too long
[NeurIPS 2025 - Under Review] [—]

[8] **Spot the BlindSpots: Systematic Identification and Quantification of Fine-Grained LLM Biases in Contact Center Summaries** [🔗]
Kawin Mayilvaghanan, Siddhant Gupta, Ayush Kumar
[EMNLP Industry Track 2025 - Under Review] [—]

[9] **NepX-Hate: A Nepali Hate Speech Corpus with Fine-Grained Sociocultural Annotations** [🔗]
Jebish Purbey, Sanjeeb Prasad Panday, Sudhan Shrestha, Rajan Kusi, Manoj Rokaya, Bikash Balami, Sajan Maharjan, Prabhat Bhatta, Ayush Purbey, Siddhant Gupta, Vatsal, Ram Mohan Rao Kadiyala
[EMNLP 2025 - Under Review] [—]

Projects

SpeechAya : Speech Synthesis August 2024 - September 2024
Open-Source 8-week long Hackathon Project by Cohere4AI

- Engineered a novel multilingual LLM pipeline integrating speech and text modalities, processing over 1000 hours of audio data from LibriSpeech and Mozilla CommonVoice datasets across 5 languages
- Implemented and optimized speech tokenization using state-of-the-art models (MMS, mHuBERT, XEUS), reducing processing time by 32% through efficient batching and parallel processing
- Achieved a score of 0.112 in Word Error Rate (WER) on the PolyAI/minds14 benchmark dataset by fine-tuning a Qwen2-1.5b model architecture with custom speech embeddings
- Developed a modular training pipeline supporting multiple speech tasks (ASR, TTS, voice cloning, translation) through a unified model architecture.

SpeechAya : Speech Synthesis August 2024 - September 2024

Open-Source 8-week long Hackathon Project by Cohere4AI

- Engineered a novel multilingual LLM pipeline integrating speech and text modalities, processing over 1000 hours of audio data from LibriSpeech and Mozilla CommonVoice datasets across 5 languages
- Implemented and optimized speech tokenization using state-of-the-art models (MMS, mHuBERT, XEUS), reducing processing time by 32% through efficient batching and parallel processing
- Achieved a score of 0.112 in Word Error Rate (WER) on the PolyAI/minds14 benchmark dataset by fine-tuning a Qwen2-1.5b model architecture with custom speech embeddings
- Developed a modular training pipeline supporting multiple speech tasks (ASR, TTS, voice cloning, translation) through a unified model architecture.

Advanced Attribute Extraction and Classification Pipeline July 2024 – August 2024

Amazon ML Hackathon 2024

- Applied advanced OCR techniques with pre-trained models to extract text from over 400,000 product images, achieving a 88% text recognition accuracy and significantly enhancing data extraction efficiency.
- Fine-tuned DistilBERT and LLaMA 3.2 for Named Entity Recognition (NER) tasks, using proper metrics for optimization, which resulted in an improvement in entity extraction precision and recall.
- Optimized LayoutLM for attribute classification tasks, such as identifying product dimensions (e.g., weight, height, width), reducing misclassification rates (False Positives) by 10-15 % and streamlining attribute extraction workflows.

Carbon Footprint Detector February 2023

Full-Stack Scalable extension made in 3 days

- Created an innovative Chrome extension that analyzed carbon emissions generated by 200+ websites, resulting in a 30% increase in user engagement with sustainability metrics.
- Developed a PostgreSQL database to manage 100,000+ user records and emission metrics, integrated with a Node.js backend for real-time data analysis.
- Engineered a robust CI/CD pipeline that streamlined testing and deployment processes, resulting in an acceleration of development cycles by 30% while ensuring consistent application scalability through containerized Docker components.

DocAI March 2024

Full-Stack Application made in 3 days

- Developed a full-stack Django web application to streamline medical test report operations, enabling seamless interactions between two distinct user roles (e.g., doctors and patients).
- Integrated NLP-based suggestion features for automated report generation, reducing manual input by 40%.
- Built a data analytics dashboard using Plotly for real-time insights and trends, and embedded a chatbot widget to assist users with suggestive use cases, improving user satisfaction by 30%.

Music Genre Classifier April 2023 – May 2023

Audio Classification Model

- Engineered a model to classify music genres using Librosa for signal processing, achieving an 91.2% accuracy rate across a 500+ hours and 6 genres dataset of music samples.
- Enhanced a CNN model with advanced techniques such as early stopping, weight decay, dropout, and batch normalization, resulting in a 38% reduction in overfitting and boost in accuracy.
- Implemented ensemble learning methods, including bagging, boosting, and voting, to improve prediction robustness and generalization.
- Optimized hyperparameters using GridCVSearch, for a better selection of models.

Deep Space Image Classifier March 2023

Celestial Object Classification Pipeline

- Developed a pipeline for classifying celestial objects using deep learning techniques, focusing on high-resolution image data.
- Conducted data preprocessing, augmentation, and multiclass labeling to handle imbalanced datasets effectively.
- Designed a multiclass classifier to predict black hole types, achieving 78% accuracy on astrophysical datasets.
- Conducted extensive Exploratory Data Analysis (EDA) and implemented imputation techniques such as KNN imputation and mean imputation, comparing their impact on model performance.
- Evaluated and deployed multiple algorithms, including Support Vector Machines, Random Forest Classifier, Logistic Regression, Artificial Neural Networks, LightGBM, CatBoost, and XGBoost, to ensure optimal performance.

Technical Skills

Languages	Python, C++, Julia, JavaScript
Libraries and Frameworks	Pytorch, Django, Tensorflow, Sklearn, Librosa, NLTK, Trl, Transformers, LoRA, OpenCV, Numpy, Pandas, Matplotlib , Gradio, BitsandBytes
Databases	PostgreSQL, SQL, SQLite

References

› Suman Debnath . . . . . *Principal Developer Advocate, Amazon, USA* [🌐]
› Ram Mohan Rao Kadiyala . . . . . *MS in Machine Learning, University of Maryland, College Park* [🌐]