


Siddharth Tumre

(+91) 96668-20295  siddharth.tumre@outlook.com

siddharthtumre.github.io

 siddharthtumre  /in/siddharth-tumre

Education

Indian Institute of Technology Delhi (IITD)

2021 - 2023

M.Tech. in Computer Technology (Dept. of Electrical Engineering)

CGPA: 7.979/10

Chaitanya Bharathi Institute of Technology (CBIT)

2017 - 2021

B.E. in Computer Science and Engineering

CGPA: 7.28/10

Research Interests

Natural Language Processing, Large Language models (LLMs)–Multimodality and Domain-specific applications, Information Extraction and Reasoning, LLM Alignment.

Work Experience

Researcher, Data and Decision Sciences, TCS Research

Sept. 2023 –

Theme: NLP for Business News

- Curated and benchmarked a dataset of business news snippets annotated for near-duplicate detection, tackling challenges posed by paywalls. [1]
- Designed a novel unsupervised near-duplicate detection technique combining metadata and dense-vector embeddings of news content, with a patent filed at the Indian Patent Office. [2]
- Designed algorithms and trained models for generating *thematic and customizable newsletters* tailored to enterprise needs, leveraging parameter-efficient fine-tuned BERT models, Zero-shot classification, and Topic Modeling, with human-in-the-loop feedback for continual learning and adaptation. [3]

Theme: NLP for Industrial Safety

- Aligned open-source LLMs through Supervised Fine-Tuning to improve specificity of preventive action recommendations for industrial safety. [5] [4]
- Employed LLM-as-a-Judge to evaluate LLM generated safety recommendations. [6]

Bio-NER (M.Tech. Project), IITD

Jul. 2022 – Jun. 2023

- Analyzed BioBERT and PubMedBERT models for Biomedical NER, fine-tuning using various modeling choices on the JNLPBA dataset.
- Conducted qualitative analysis, identifying failures in recognizing generic terms, compound entities, and IOB2 tagging violations.
- Integrated BioWordVec with BERT embeddings in a BERT-BiLSTM-CRF architecture, achieving improvements in specific error cases.

Web Developer Intern, IITD

Jan. 2023 – May 2023

- Developed a web application for ISRO Automatic Weather Stations, a smart system providing real-time flood warnings. Users can access, view, and download sensor data on temperature, humidity, soil moisture, and rainfall from multiple locations.

Teaching Assistant, IITD

Aug. 2022 - Dec. 2022

- Introduction to Machine Learning (ELL784)

Application Developer Intern, ThoughtWorks

May 2021 – July 2021

- Gained hands-on experience in the complete software development lifecycle, including coding, testing, and deploying production-ready software while adhering to agile methodologies.

Core Competencies

| | |
|---|---|
| Machine Learning | Supervised/Unsupervised Learning, Neural Networks, Transformers, PyTorch, TensorFlow, Pandas, NumPy, sci-kit-learn, MLflow, Weights & Biases (W&B), Generative AI (GenAI) |
| Natural Language Processing (NLP) | NLTK, spaCy, Hugging Face, Gensim, Large Language Models (LLM), Fine-tuning LLMs (e.g., GPT, BERT, T5, Llama), Prompt Engineering, Retrieval-Augmented Generation (RAG), LangChain, Reinforcement Learning with Human Feedback (RLHF) |
| Data Analysis | Data Preprocessing/Visualization, Exploratory Data Analysis (EDA) |
| Programming and Software Development | Python, C, C++, Java, SpringBoot, React.js, Node.js, Express.js, PostgreSQL, Git, GitHub, Agile, Docker, Flask |

Positions of responsibility

Covener, IITD

July 2022 – May 2023

Co-curricular and Academic Interaction Council (CAIC)

Deputy Executive (Design), CBIT

July 2020 – May 2021

Chaitanya Bharathi Cultural Club (CBCC)

Publications

Archival Conference Papers

- [1] Alok Kumar, Siddharth Tumre, and Sangameshwar Patil. “Benchmarking Near-Duplicate Detection in the era of Pay-walled News”. In: *Companion Proceedings of the ACM Web Conference 2025*. WWW '25. Sydney, Australia: Association for Computing Machinery, 2025. ISBN: 9798400701726. DOI: 10.1145/3701716.3715303. URL: <https://doi.org/10.1145/3701716.3715303>.
- [2] Siddharth Tumre, Sangameshwar Patil, and Alok Kumar. “Improved Near-Duplicate Detection for Aggregated and Paywalled News-Feeds”. In: *NAACL (Industry Track)*. 2025.
- [3] Siddharth Tumre, Alok Kumar, Ajay Phade, Nihar Riswadkar, and Sangameshwar Patil. “Newsletter-Factory: A Thematic Newsletter Generation Tool for Curating Business Insights”. In: *Text2Story@ECIR*. 2025.
- [4] Siddharth Tumre, Shubham Kumbhar, Sumit Koundanya, and Sangameshwar Patil. “Aligning LLMs to Improve Specificity of Preventive Action Recommendations for Industrial Safety”. In: *The International FLAIRS Conference Proceedings*. Vol. 38. 2025.

Non-Archival Posters & Workshop Papers

- [5] Shubham Kumbhar, Siddharth Tumre, Sumit Koundanya, and Sangameshwar Patil. “Employing Open-Source LLMs for Preventive Action Recommendations to Avoid Industrial Safety Incidents”. Paper presented at the AAAI '25 Workshop on Open-Source AI for Mainstream Use. 2025.
- [6] Sumit Koundanya, Shubham Kumbhar, Siddharth Tumre, and Sangameshwar Patil. “LLM-as-a-Judge: Evaluating LLM Recommendations for Industrial Safety”. Paper presented at the LLM4Eval @ WSDM 2025 Workshop. 2025.