

Aman Sinha

Website: <https://amansinha09.github.io/>

Github: <https://github.com/amansinha09/>

Email : aman.sinha@atilf.univ-lorraine.fr

Address : Bureau 003, 44 Avenue de la Libération,
ATILF, Université de Lorraine, 54000 Nancy

RESEARCH INTEREST

Medical Language Understanding, Natural Language Processing

EDUCATION

- **Université de Lorraine - Institut de Cancérologie Strasbourg** Nancy, France
PhD in Natural Language Processing *October 2021 –*
Advisor: Marianne Clausel, Matthieu Constant, Xavier Coubez
- **Indian Institute of Technology(Indian School of Mines), Dhanbad** Jharkhand, India
Integrated Master of Technology in Mathematics and Computing; *2015 – 2020*

RELEVANT PUBLICATIONS

1. I Buhnla , [A Sinha](#), M Constant. (2025). TrackList: Tracing Back Linguistic Diversity of Head and Tail Knowledge in Open Language Models (*Under Review*)
2. T Mickus, [A Sinha](#), R Vazquez. (2025). Data uncertainty is not data complexity (*Under Review*)
3. [A Sinha](#), T Mickus , M Clausel, M Constant, X Coubez. (2024). As Simple as 1, 2, 3: Quantifying the Simplicity of Interpretations (*Under Review*)
4. [A Sinha](#), T Mickus , M Clausel, M Constant, X Coubez. (2024). Domain-specific or Uncertainty-aware models: Does it really make a difference for biomedical text classification? *BioNLP Workshop, ACL 2024*
5. I Buhnla*, [A Sinha](#)*, M Constant. (2024). Retrieve, Generate, Evaluate: A Case Study for Medical Paraphrases Generation with Small Language Models. *KnowledgeableLM Workshop, ACL 2024*
6. R Agarwal*, [A Sinha](#)*, D Prasad, M Clausel, A Horsh, M Constant, X Coubez. (2023). Modelling Irregularly Sampled Time Series Without Imputation. (*Under Review*)
7. [A Sinha](#), CG Holgado, M Clausel, M Constant X Coubez. (2023). Can LLMs be used to understand clinical notes better?. *Actes des 5mes journées du Gdr LIFT 2023*.
8. [A Sinha](#), S Bigeard, M Clausel, M Constant. (2023). What shall we read: the article or the citations?-A case study on scientific language understanding. *Actes de l'atelier "Analyse et Recherche de Textes Scientifiques"(ARTS), TALN-CORIA 2023* .
9. [A Sinha](#), A Vishwakarma, M Clausel, M Constant. (2022). CME2 Net: Contextual Medical Event Extraction Network for clinical notes. *WNLPe-Heath Workshop, ICON 2022*
10. [A Sinha](#)*, CG Holgado*, M Clausel, M Constant. (2022). Catch me if you can! Capturing complex disease mentions in tweets. *Social Media Mining for Health 2022, COLING 2022*
11. [A Sinha](#), S Ollinger, M Constant. (2022). Word Sense Disambiguation of French Lexicographical Examples Using Lexical Networks. *TextGraphs-16 Workshop, COLING 2022*
12. AK Jena, [A Sinha](#)*, R Agarwal*. (2020). C-net: Contextual network for sarcasm detection. In *Proceedings of the Second Workshop on Figurative Language Processing, ACL 2020*

RESEARCH EXPERIENCE

- **University of Helsinki** Helsinki, Finland
Visiting Researcher July 2024 – October 2024
 - **Uncertainty paradigm for Medical domain:** Exploring *difficulty* in data/models; Hallucination in LLMs;
- **ATILF, University of Lorraine** Nancy, France
Research Engineer November 2020 – June 2021
 - **Word Sense Disambiguation:** Working on qualitative experiments for word disambiguation on text resources by incorporating lexical networks.
- **IECL, University of Lorraine** Nancy, France
Research Intern (Mentor: Prof. Marianne Clausel and Prof. Matthieu Constant) February 2020 – October 2020
 - **Information retrieval in linguistic networks:** Develop graph matching techniques well adapted to lexical data.
- **Indian Institute of Technology Kanpur** Kanpur, India
Master Thesis (Mentor: Dr. Ashutosh Modi) February 2020 – July 2020
 - **Multimodal Emotion Recognition:** Worked on developing a multimodal emotion recognition system.[Link]
- **IBM India Research Labs** Bangalore, India
Research Intern (Mentor: Parag Jain and Priyanka Agarwal) May 2019 – July 2019
 - **Multi-task learning under heterogeneous supervision for semantic parsing:** Leveraging heterogeneous supervision in multitask learning to boost weak supervised learning based semantic parsing.
- **Language Technologies Research Centre, IIIT Hyderabad** Hyderabad, India
Summer Research Intern (Advisor: Dr. Radhika Mamidi) May 2018 – July 2018
 - **Evaluation of Machine Translation Systems for Sarcasm Transfer:** Worked on evaluation of translation systems for transfer of sarcasm from source to translated language. Built a Sarcasm Detection System for Hindi language using various machine and deep learning models.
- **Oriserve** Mumbai, India
Data Science Intern May 2017 – July 2017
 - **Pre-processing Method for Text Correction:** Worked on character-based RNN model to pre-process user responses of DishTV chatbot conversation logs to change the corrupted text to English language.
 - **Intent Classification for User Responses:** Implemented a Multi-class Naive Bayes Classifier based on Bag-of-Words algorithm to classify the client responses based on the intent of the conversation.

ACHIEVEMENTS

- Recipient of PhD DREAM mobility funding of 5500 euros at University of Lorraine.
- Recipient of Diversity & Inclusion Award for attending ACL 2020 conference.
- Received Student Grant for the FIRE 2019 conference.
- Secured 2nd prize in IBM IRL Hackathon 2019.
- Mentor at IIIT-Hyderabad Advanced Summer School on Natural Language Processing 2018
- Secured 1801th rank in IIT-JEE ADVANCE-2015 from among 124741 students

POSITION OF RESPONSIBILITY

- **Teaching Assistant:** Probability and Statistics; Machine Learning TAL M1, IDMC, Université de Lorraine
- **Co-/Organizer:** MuSHROOM Shared Task 2025; UncertainNLP 2025; CaféTAL@ATILF, Université de Lorraine;
- **Reviewer:** ARR Oct 2024, COLING 2024, Neurips Interpretable Workshop 2024, ICANN 2021, WNUT 2020
- **Volunteer:** NAACL 2021, AISTATS 2021
- **Senior Mentor:** Data Science Club, IIT ISM Dhanbad (2017-20)
- **Event Coordination Member:** Society for Industrial and Applied Mathematics, IIT ISM Student Chapter
- **Teacher:** Taught underprivileged children at Kartavya, Student Run NGO (2015-17)

PRACTICAL SKILLS

- **Computer Skills:** Python3(Pytorch, Keras, Tensorflow, NLTK), C++, Unix, Latex
- **Languages:** English (Professional), French (A2), Hindi (native)