Ishani Mondal
Research Fellow| Microsoft Research India

🌐 https://ishani-mondal.github.io/   in linkedin.com/in/ishani-mondal-90ba39b0   ⌂ github.com/Ishani-Mondal
📱 +91 9748120449   @ ishani340@gmail.com
📍 4/3K/340, Sarsuna Satellite Township, Muchipara, Kolkata-700061

EDUCATION

March 2021 July 2017	**Indian Institute of Technology, Kharagpur, ,** Master of Science (MS) in Computer Science and Engineering Advisors : *Prof. Pawan Goyal*, *Prof. Sudeshna Sarkar* **CGPA : 9.53/10 or 3.91/4** (Thesis Submitted : February, 2020) Thesis Title : *Predicting similar and interacting entities from structured and unstructured biomedical resources*
July 2015 July 2011	**Institute of Engineering and Management, ,** Affiliated to Maulana Abul Kalam Azad University of Technology, West Bengal Bachelor of Technology (B.Tech) in Computer Science and Engineering **CGPA : 9.44/10 or 3.96/4**(Department Rank : 1 out of 100 students approx)

RESEARCH EXPERIENCE

Present July 2020	**Research Fellow, | MICROSOFT RESEARCH INDIA, | Speech and Natural Language Processing** Advisors : *Dr. Kalika Bali*, *Dr. Monojit Choudhury* > Developing an end-to-end multilingual conversational agent 'WhatsInMyReport' which helps the patients understand their medical reports better (**Honorable Mention Award in MS Global Hackathon**). > Exploring the gaps between the healthcare chatbots laid out as prototypes in the NLP based literature and the ones actually being deployed to use. > Developing a pipeline for grapheme to phoneme converter for the low-resource Indian languages. > Developing a linguistic annotation schema and annotation interface for the multilingual HIV living youth in Kenya, participating in an online peer support forum and performing sociolinguistic studies on their engagement patterns.
March 2020 May 2017	**Project Officer, | IIT KHARAGPUR, | Computer Science and Engineering Department** Advisors : *Prof. Sudeshna Sarkar*, *Prof. Pawan Goyal* Industrial Partner : *Excelra Knowledge Solutions, Hyderabad* > Worked on the main challenges of coreference resoution in mining the information about the biological entities like genes, drugs, diseases from the patents and journals. > Proposed novel methods for Drug-Drug Interaction Prediction and entity linking on medical entities.
June 2020 April 2020	**Summer Research Intern, | IBM RESEARCH UK, | Speech and Natural Language Processing** Advisors : *Dr. Yufang Hou*, *Dr. Charles Jochim*, *Prof. Debasis Ganguly* > Developed an end-to-end framework for the construction of NLP-based Knowledge Graph. > Proposed Active-Learning Based methods for enhancing explainability of classification models.
July 2019 May 2019	**Summer Research Intern, | TCS RESEARCH INDIA, | Speech and Natural Language Processing** Advisors : *Dr. Tirthankar Dasgupta*, *Dr. Lipika Dey* > Worked on implementing attention aware CNN-Bi-LSTM model for inclusion and exclusion criteria classification for the Clinical Trials and proposed a novel weakly-supervised co-training based method which can exploit a large pool of unlabeled criteria sentences to augment the limited supervised training data.
May 2017 November 2016	**Research Assistant, | IIT BOMBAY, | Systems and Network Lab** Advisors : *Prof. Abhay Karandikar* > Worked in a 5G Project where my role was to prepare a simulation of the RF planning tool aiming to dynamically update the OpenStreetMaps based on the link budget of the distributed nodes.

SOFTWARE DEVELOPMENT EXPERIENCE

October 2016 August 2015	Associate Software Engineer,	Nᴏᴍᴜʀᴀ Rᴇsᴇᴀʀᴄʜ Iɴsᴛɪᴛᴜᴛᴇ Fɪɴᴀɴᴄɪᴀʟ Tᴇᴄʜɴᴏʟᴏɢɪᴇs Iɴᴅɪᴀ Pᴠᴛ. Lᴛᴅ,
	› Developed web-application using JAVA based Spring Framework and Restful Web Services.

Sᴇʟᴇᴄᴛᴇᴅ Pᴜʙʟɪᴄᴀᴛɪᴏɴs

S=Iɴ Sᴜʙᴍɪssɪᴏɴ/Iɴ Pʀᴏɢʀᴇss, C=Cᴏɴꜰᴇʀᴇɴᴄᴇ, W=Wᴏʀᴋsʜᴏᴘ/Pᴏsᴛᴇʀ

[S.3] **Global Readiness of Language Technology for Healthcare : What would it Take to Combat the Next Pandemic ?** [%]
Ishani Mondal, Kabir Ahuja, Mohit Jain, Jacki O'Neil, Kalika Bali, Monojit Choudhury
[In Submission] ARR 2022 [ARR 2022]

[S.2] **Natural Instructions : Benchmarking Generalization to New Tasks from Natural Language Instructions**
Swaroop Ranjan Mishra, Daniel Khasabi, Ishani Mondal
[In Submission] ARR 2022 [ARR 2022]

[S.1] **Understanding Sociolinguistic and Pragmatic Aspects of Peer Support in Multilingual Healthcare Conversations**
Ishani Mondal, Mohit Jain, Kalika Bali, Monojit Choudhury, Jacki O'Neil
Working Draft [%] [[%]]

[C.5] **End-to-End Construction of NLP Knowledge Graph** [%][◼]
Ishani Mondal, Yufang Hou, Charles Jochim
Annual Conference of the Association for Computational Linguistics Findings of ACL'21 [Findings of ACL'21]

[C.4] **BBAEG : Towards BERT-based Biomedical Adversarial Example Generation for Text Classification** [%][◼]
Ishani Mondal
North American Chapter of the Association for Computational Linguistics NAACL'21 [NAACL'21]

[C.3] **Multi-Objective Few-shot Learning for Fair Classification** [%][◼]
Ishani Mondal, Procheta Sen, Debasis Ganguly
ACM International Conference on Information and Knowledge Management CIKM'21 [CIKM'21]

[C.2] **ALEX : Active Learning based Enhancement of a Classification Model's EXplainability** [%][◼]
Ishani Mondal, Debasis Ganguly
ACM International Conference on Information and Knowledge Management CIKM'20 [CIKM'20]

[C.1] **Drug-Drug Interactions Prediction Based on Drug Embedding and Graph Auto-Encoder** [%]
Ishani Mondal, Sukannya Purkayastha, Sudeshna Sarkar, Pawan Goyal, Jitesh K Pillai
IEEE 19th International Conference on Bioinformatics and Bioengineering IEEE BIBE'19 [IEEE BIBE'19]

[W.7] **A Linguistic Annotation Framework to Study Interactions in Multilingual Healthcare Conversational Forums** [%]
Ishani Mondal, Kalika Bali, Mohit Jain, Monojit Choudhury, Ashish Sharma, Evans Gitau, Jacki O'Neill, Kagonya Awori, Sarah Gitau
15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop LAW-DMR@EMNLP'21 [LAW-DMR@EMNLP'21]

[W.6] **BERTKG-DDI : Towards Incorporating Entity-specific Knowledge Graph Information in Predicting Drug-Drug Interactions** [%]
Ishani Mondal
Workshop on Scientific Document Understanding (SDU) SDU@AAAI'21 [SDU@AAAI'21]

[W.5] **BERTChem-DDI : Improved Drug-Drug Interaction Prediction from text using Chemical Structure Information** [%]
Ishani Mondal
Knowledgeable NLP : the First Workshop on Integrating Structured Knowledge and Neural Networks for NLP KnowledgeableNLP@AACL-IJCNLP'20 [KnowledgeableNLP@AACL-IJCNLP'20]

[W.4] **Extracting Semantic Aspects for Structured Representation of Clinical Trial Eligibility Criteria** [%]
Ishani Mondal, Tirthankar Dasgupta, Abir Naskar, Sudeshna Jana, Lipika Dey.
3rd Clinical Natural Language Processing Workshop ClinicalNLP@EMNLP'20 [ClinicalNLP@EMNLP'20]

[W.3] **Automatic Segregation and Classification of Inclusion and Exclusion Criteria of Clinical Trials to improve Patient Eligibility Matching** [%]
Tirthankar Dasgupta, Ishani Mondal, Abir Naskar, Sudeshna Jana, Lipika Dey
Explainable AI in Healthcare and Medicine HealthIntelligence@AAAI'20 [HealthIntelligence@AAAI'20]

[W.2] **Approaches to biomedical coreference resolution** [%]
Ishani Mondal
Proceedings of the 7th ACM IKDD CoDS and 25th COMAD CoDS COMAD'20 [CoDS COMAD'20]

[W.1]   Medical Entity Linking using Triplet Network [🔗]
Ishani Mondal, Sukannya Purkayastha, Sudeshna Sarkar, Pawan Goyal, Jitesh Pillai, Amitava Bhattacharyya, Mahanandeeshwar Gattu
2nd Clinical Natural Language Processing Workshop   ClinicalNLP@NAACL'19   [**ClinicalNLP@NAACL'19**]

SELECTED RELEVANT RESEARCH PROJECTS

1. Understanding Socio-Linguistic behavior of Peer Supporters in Code-Switched Conversations
FEB'21 - PRESENT

> *Advisor : Kalika Bali, Monojit Choudhury, Mohit Jain, Maxamed Axmed (Microsoft Research)*
> We conduct an exploratory study on multilingual interaction patterns in two health-focused WhatsApp groups in Kenya serving multilingual HIV positive youths (proficient in English, Swahili and Sheng), focusing on understanding language and sentiment preferences for expressing specific conversational intents, linguistic accommodation, sentiment patterns and functions of switching languages.
> Studied the patterns of engagement using social network analysis.
> We observe that native languages are preferred for informal conversations, linguistic accommodation pattern varies with the group's demography, and structural switching patterns are more dominant than pragmatic switching.

2. Empowering Low-Resource Languages using Language Technologies
APRIL'21 - PRESENT

> *Advisor : Kalika Bali, Monojit Choudhury (Microsoft Research)*
> Developing a pipeline for grapheme to phoneme converter for the low-resource languages like Gondi and Mundari.
> Exploring sentiment analysis on code-mixed conversations using Massively Multilingual Language Models (MMLMs).

3. Exploring the bottlenecks of SOTA models on Downstream tasks
SEPT'21 - PRESENT

> *Advisor : Kalika Bali, Monojit Choudhury, Mohit Jain (Microsoft Research)*
> Performed qualitative and quantitative study on healthcare coversational agents to understand their bottlenecks in terms of their linguistic, pragmatic and mathematical reasoning.
> Explored the linguistic disparity of the commercial conversational frameworks and proposed a global readiness metric for each country to determine their effectiveness to combat a healthcare pandemic (like COVID) using Language Technology.

4. Distilling Relevant Knowledge from Scientific Literature and Performing Temporal Analysis
APRIL'20 - PRESENT

> *Advisor : Yufang Hou, Chales Jochim (IBM Research)*
> Constructed a Knowledge Graph after extracting relevant information about the scientific entities such as NLP Tasks, Datasets, Evaluation Metrics. We have proposed novel methods for coreference and relation extraction among these entities to achieve the task. We also came up with promising strategies to evaluate the constructed graph.
> Performing a diachronic analysis of the scientific entities using citation networks to determine their temporal stability.

5. Effective Drug Repurposing through Literature Mining and External Knowledge Sources
MAY'17 - JULY'20

> *Advisor : Sudeshna Sarkar, Pawan Goyal (IIT Kharagpur)*
> Proposed novel methods for biomedical coreference resolution and entity linking for entities such as genes, drugs.
> Proposed novel methods for predicting Drug-Drug Interaction using both literature and external Knowledge sources such as chemical structure information and Knowledge Graph.
> Explored the adversarial robustness of the state-of-the-art biomedical NLP models on classification tasks.

6. Fairness and Interpretability in NLP Systems
MAY'20 - JULY'21

> *Advisor : Debasis Ganguly (University of Glasgow)*
> Proposed multi-objective debiasing methods for predicting the sentiment. It uses both semi-supervised and unsupervised methods (using clustering heuristic) to alleviate the issue of training data crisis.
> To improve the explainability of classifiers, I have proposed an active learning based method using SHAP values.

7. Semantic Aspect Extraction in the Patient's Eligibility Criteria in Clinical Trials
MAY'19 - JULY'19

> *Advisor : Tirthankar Dasgupta, Lipika Dey (TCS Research)*
> Eligibility criteria in the clinical trials specify the characteristics that a patient must or must not possess in order to be treated according to a standard clinical care guideline. I proposed an attention aware CNN-Bi-LSTM model for automatic segregation and classification of Inclusion and Exclusion Criteria of clinical trials to improve patient eligibility matching.
> In order to deal with paucity of aspect-annotated clinical trial data, I propose a novel weakly-supervised co-training based method which can exploit a large pool of unlabeled criteria sentences to augment the limited supervised training data.

Academic Service

Reviewer	PeerJ Computer Science, BMC BioInformatics 2021, MLH4H@NeuriPS'21
Sub-Reviewer	ACL'21, NAACL'21, WSDM'21, LREC'20, CODS-COMAD'20, Code-Switching Workshop@EMNLP'21, RAI Workshop@KDD'21
Volunteer	GHCI'19, IRISS'19, ISCLS'19

Honours and Awards

Batch Topper by Computer Science Department, Institute of Engineering and Management in 2015

Merit Scholarship awarded in 2011 for being consistent topper in high-school

Grace-Hopper India Scholarship'19 [◉] For attending and presenting poster in 2019

Infosys Women Travel Grant [◉] For attending and presenting paper at NAACL in 2019

Honorable Mention Award in Microsoft Global Hackathon [◉] For development and ideation of 'WhatsInMyReport' a voice-cum-text based multilingual chatbot

Teaching and Leadership Roles

SNLP Reading Group, MSR India *Organizer*   Jul'21 - Present

> Organize a weekly lab-wide Reading Group focused on research taking place in the area of Speech and Natural Language Processing (SNLP). We read recent, classical papers as well as arrange for external talks in related areas.

Technology for Emerging Market Reading Group, MSR India *Organizer*   Dec'21 - Present

> Organize a weekly lab-wide Reading Group focused on research taking place in the area of User-centered Technologies. We read classical papers and arrange for invited talks.

Introduction to Speech and Natural Language Processing (CS60057) *Lead Teaching Assistant*   July'19 - Dec'19

> Course Instructor : Prof. Sudeshna Sarkar
> Responsibilities included evaluating labs, and helping students with the coursework, lab assignments and projects.

Introduction to Machine Learning by NPTEL *Teaching Assistant*   July'19 - Dec'19

> Course Instructor : Prof. Sudeshna Sarkar
> Responsibilities included preparation of question papers and assignments for the students.

Introduction to Deep Learning (CS60010) *Teaching Assistant*   Jan'19 - May'19

> Course Instructor : Prof. Sudeshna Sarkar
> Responsibilities included evaluating labs, and helping students with the coursework and lab assignments.

Introduction to Speech and Natural Language Processing (CS60057) *Teaching Assistant*   July'18 - Dec'18

> Course Instructor : Prof. Sudeshna Sarkar
> Responsibilities included evaluating labs, and helping students with the coursework, lab assignments and projects.

Introduction to Programming and Data Structure (CS10001) *Teaching Assistant*   Jan'18 - May'18

> Course Instructor : Prof. Aritra Hazra
> Responsibilities included evaluating labs, and helping students with the coursework and lab assignments.

Technical Skills

Languages	Python, C, Java , MATLAB, Bash Scripts
Frameworks	PyTorch, Keras, TensorFlow, NumPy, Pandas, Scikit-Learn, Scipy, NLTK, Spacy, Scispacy
DevOps and Version Control	Microsoft Azure, Amazon Web Services, Google Cloud, Git, SVN
Relevant CourseWork	Machine Learning, Deep Learning, Speech and Natural Language Processing, Information Retrieval, Algorithms, Programming and Data Structures, OOPS, Discrete Mathematics

References

> Prof. Pawan Goyal . . . . *Associate Professor, Computer Science Engineering Department, IIT Kharagpur* [◉]
> Prof. Sudeshna Sarkar . . . .*Professor, Computer Science Engineering Department, IIT Kharagpur* [◉]
> Prof. Debasis Ganguly . . . . *Assistant Professor, University of Glasgow, UK* [◉]
> Dr. Yufang Hou . . . . *Research Staff Member, IBM Research, UK* [◉]
> Dr. Kalika Bali . . . . *Principal Researcher, Microsoft Research, India* [◉]
> Dr. Monojit Choudhury . . . . *Principal Researcher, Microsoft Research, India* [◉]