SANDIPAN DANDAPAT

CS 404, Dept. of Computer Science & Engineering
Indian Institute of Technology Hyderabad
Kandi-502284, Sangareddy
Telangana, INDIA
Email: sdandapat@cse.iith.ac.in; sandipandandapat@gmail.com
Cell: +919740103939

INTERESTS

Large Language Models, Artificial Intelligence, Natural Language Processing

EDUCATION

Ph.D. Centre for Next Generation Localization, School of Computing, Dublin City University (Mar. 2009 – May, 2012)
Advisors: Prof. Andy Way and Dr. Sara Morrissey
Thesis Title: Mitigating the Problems of SMT using EBMT
Description: The main motivation of this work is to use example-based MT approaches to overcome some the difficulties of state-of-the-art phrase-based Statistical MT systems. Novel EBMT algorithms (using translation memories) were designed to mitigate the problem of statistical MT for translating minority languages.

M.S. Department of Computer Science & Engineering, IIT Kharagpur, INDIA (Jan. 2005 – Feb. 2008)
Advisors: Prof Sudeshna Sarkar and Prof Anupam Basu, CSE, IIT Kharagpur
Thesis Title: Part-of-Speech Tagging for Bengali
Description: The aim of the work is to develop a Part-of-Speech tagger for Indian Languages especially for resource poor languages. The key techniques used are different machine learning approaches (especially graphical models like HMM, MaxEnt and CRF). The system is being tested on Indian languages (Hindi and Bengali). The system can be used in a variety of applications like search engines, Machine Translation systems, etc.
Courses Taken: Machine Learning, Intelligent Systems, Data Mining, Human Computer Interaction.
CGPA: 9.51/10

Post Graduate Diploma (Computational Linguistics). IIIT Hyderabad, INDIA (2002 – 2004)
CGPA: 7.72/10

B.Tech. Computer Science & Engineering, Haldia Institute of Technology, INDIA (1998 – 2002)
Percentage: 73.3%

Higher Secondary (XII). West Bengal Council of Higher Secondary Education. 1997.
Percentage: 67.2%

Madhyamik (X). West Bengal Board of Secondary Education, 1995.
Percentage: 79.2%

EXPERIENCES

Principal Applied Researcher, Microsoft R&D India Pvt Ltd. (Mar, 2020 – Sep, 2025)
I have worked on four different verticals of making Bing trustworthy which include protecting adults, defensive search and helping different experience to resolve their issues. While we develop deep models to classify, intervene the defensive issues, I helped the team to onboard to GPT-X model which improve in quality, cost and agility. In addition, the work also includes how to measure defensive leakage across experiences.

Senior Applied Researcher, Microsoft R&D India Pvt Ltd. (June, 2016 – Feb, 2020)

Primarily I have worked on Machine Translation systems for Indian Languages (IL). I worked towards the first appearance of Bangla—English, Telugu—English, Tamil—English MT system in Bing, and have improved the Hindi—English MT quality. The challenge was how to develop Indian Language MT system which has less training data and richer morphology using both statistical and deep models.

Research Scientist, Xerox Research Centre India (Jan, 2015 – May, 2016)
I worked on different NLP components to estimate customer satisfaction in real time in contact centre conversation. This include understanding customer behavior, organizational compliance, conversational characteristics etc.

Post-doctoral Researcher, CNGL, Dublin City University (June, 2014 – December, 2014)
I worked on incremental retraining of SMT system using CDEC and Moses using linked data. This also include careful supplementary data selection for incremental retraining and integration of SMT and NER component. This was part of FP7 project FALCON.

Assistant Professor, Dept. of CSE, IIT Guwahati (January, 2013 – May, 2014)
taught both graduate and undergraduate courses, supervised six BTech projects and one MTech student, and engaged in research focused on translation. My teaching included core courses in Programming and Data Structures, as well as an elective in Natural Language Processing.

Post-doctoral Researcher, CNGL, Dublin City University (June, 2012 – December, 2012)
I worked on the project TMTprime – which offers translation recommendation to reduce the overall cost of post-editing of MT output. I was working on the development of fast and robust and scalable TM module. I also worked on quality estimation of parallel data. Uses machine learning technique for automatic classification of noisy parallel data.

Research Intern, Microsoft Research Redmond, WA (June 2011 – August, 2011)
I was working towards the development of an automatic MT system for a new language pair English—Bangla. The particular language pair had almost no data to train a machine translation system. I was assigned to explore strategies to build a reasonably good amount of inventory of parallel data for new language pair, which can further be adopted for any language pair. I hope that the infrastructure and testing I did will be essential for future endeavors.

Research Intern and Research Vendor, Microsoft Research Lab India Private Limited, (February, 2008 – January, 2009)
I have worked on methods for the fast creation of human annotated data (which is crucial for supervised machine learning algorithms) for several syntactic processing tasks. The task mainly involves creation of user-friendly tools for complex annotation tasks and incorporating machine intelligence.

PUBLICATIONS, PATENTS, ACHEVIMENTS

Publications

- Shrawgi, Hari, Prasanjit Rath, Tushar Singhal, and **Sandipan Dandapat**. "Uncovering Stereotypes in Large Language Models: A Task Complexity-based Approach." In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (**EACL**) (Volume 1: Long Papers)*, pp. 1841-1857. 2024. (Core Rank: A)
- Diddee, Harshita, Anurag Shukla, Tanuja Ganu, Vivek Seshadri, **Sandipan Dandapat**, Monojit Choudhury, and Kalika Bali. "INMT-Lite: Accelerating Low-Resource Language Data Collection via Offline Interactive Neural Machine Translation." In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (**LREC-COLING**)*, pp. 9097-9109. 2024.
- Jain, Raghav, Daivik Sojitra, Arkadeep Acharya, Sriparna Saha, Adam Jatowt, and **Sandipan Dandapat**. "Do language models have a common sense regarding time? revisiting temporal commonsense reasoning in

the era of large language models." In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (**EMNLP**)*, pp. 6750-6774. 2023. (Core Rank: A*)

- Sood, Ojasvin, and **Sandipan Dandapat**. "Problematic Webpage Identification: A Trilogy of Hatespeech, Search Engines and GPT." *The 7th Workshop on Online Abuse and Harms (WOAH)*. 2023.
- Kumar, Shanu, Soujanya Abbaraju, **Sandipan Dandapat**, Sunayana Sitaram, and Monojit Choudhury. "DiTTO: A Feature Representation Imitation Approach for Improving Cross-Lingual Transfer." In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* (***EACL***). pp. 385-406. 2023. (Core Rank: A)
- Vashishtha, Aniket, S. Sai Prasad, Payal Bajaj, Vishrav Chaudhary, Kate Cook, **Sandipan Dandapat**, Sunayana Sitaram, and Monojit Choudhury. "Performance and Risk Trade-offs for Multi-word Text Prediction at Scale." In *Findings of the Association for Computational Linguistics* (***EACL***), pp. 2181-2197. 2023. (Core Rank: A)
- Ghosal, Deepanway, Somak Aditya, **Sandipan Dandapat**, and Monojit Choudhury. "Vector Space Interpolation for Query Expansion." In *Findings of the Association for Computational Linguistics* **AACL-IJCNLP** 2022: 405.
- Karthikeyan, K., Shaily Bhatt, Pankaj Singh, Somak Aditya, **Sandipan Dandapat**, Sunayana Sitaram, and Monojit Choudhury. "Multilingual CheckList: Generation and Evaluation." In *Findings of the Association for Computational Linguistics*: **AACL-IJCNLP** 2022, pp. 282-295. 2022.
- Ahuja, Kabir, Antonios Anastasopoulos, Barun Patra, Graham Neubig, Monojit Choudhury, **Sandipan Dandapat**, Sunayana Sitaram, and Vishrav Chaudhary. "The SUMEval 2022 Shared Task on Performance Prediction of Multilingual Pre-trained Language Models." **AACL 2022** (2022): 1
- Diddee, Harshita, **Sandipan Dandapat**, Monojit Choudhury, Tanuja Ganu, and Kalika Bali. "Too Brittle To Touch: Comparing the Stability of Quantization and Distillation Towards Developing Lightweight Low-Resource MT Models." **WMT** 2022 (2022): 870.
- Ahuja, Kabir, Sunayana Sitaram, **Sandipan Dandapat**, and Monojit Choudhury. "On the Calibration of Massively Multilingual Language Models." In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 4310-4323. **EMNLP** 2022. (Core Rank: A*)
- Kumar, Shanu, **Sandipan Dandapat**, and Monojit Choudhury. ""Diversity and Uncertainty in Moderation" are the Key to Data Selection for Multilingual Few-shot Transfer." In *Findings of the Association for Computational Linguistics*: **NAACL** 2022, pp. 1042-1055. 2022. (Core Rank: A)
- Srinivasan, Anirudh, Gauri Kholkar, Rahul Kejriwal, Tanuja Ganu, **Sandipan Dandapat**, Sunayana Sitaram, Balakrishnan Santhanam, Somak Aditya, Kalika Bali, and Monojit Choudhury. "Litmus predictor: An ai assistant for building reliable, high-performing and fair multilingual nlp systems." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11, pp. 13227-13229. **AAAI** 2022. (Core Rank: A*)
- Ahuja, Kabir, Monojit Choudhury, and **Sandipan Dandapat**. "On the Economics of Multilingual Few-shot Learning: Modeling the Cost-Performance Trade-offs of Machine Translated and Manual Data." In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1369-1384. **NAACL** 2022.. (Core Rank: A)
- Ahuja, Kabir, **Sandipan Dandapat**, Sunayana Sitaram, and Monojit Choudhury. "Beyond Static models and test sets: Benchmarking the potential of pre-trained models across tasks and languages." In *Proceedings of NLP Power! The First Workshop on Efficient Benchmarking in NLP*, pp. 64-74. 2022.
- Ahuja, Kabir, Shanu Kumar, **Sandipan Dandapat**, and Monojit Choudhury. "Multi Task Learning For Zero Shot Performance Prediction of Multilingual Models." In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), pp. 5454-5467. **ACL** 2022. (Core Rank: A*)
- Bhatt, Shaily, Poonam Goyal, **Sandipan Dandapat**, Monojit Choudhury, and Sunayana Sitaram. "On the Universality of Deep Contextual Language Models." In *Proceedings of the 18th International Conference on Natural Language Processing* (ICON), pp. 106-119. 2021.
- Santy, Sebastin, Kalika Bali, Monojit Choudhury, **Sandipan Dandapat**, Tanuja Ganu, Anurag Shukla, Jahanvi Shah, and Vivek Seshadri. "Language Translation as a Socio-Technical System: Case-Studies of Mixed-Initiative Interactions." In *ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 156-172. 2021.

- Bhatt, Shaily, Rahul Jain, **Sandipan Dandapat**, and Sunayana Sitaram. "A case study of efficacy and challenges in practical human-in-loop evaluation of nlp systems using checklist." In *Proceedings of the Workshop on Human Evaluation of NLP Systems* (HumEval), pp. 120-130. 2021.
- Mitra, Rajarshee, Manish Gupta, and **Sandipan Dandapat**. "Transformer Models for Recommending Related Questions in Web Search." In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 2153-2156. **CIKM** 2020. (Core Rank: A)
- Srinivasan, Anirudh, **Sandipan Dandapat**, and Monojit Choudhury. "Code-mixed parse trees and how to find them." In Proceedings of the The *4th Workshop on Computational Approaches to Code Switching*, pp. 57-64. 2020.
- Khanuja, Simran, **Sandipan Dandapat**, Anirudh Srinivasan, Sunayana Sitaram, and Monojit Choudhury. "GLUECoS: An Evaluation Benchmark for Code-Switched NLP." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3575-3585. **ACL** 2020. (Core Rank: A*)
- Khanujaa, Simran, **Sandipan Dandapat**, Sunayana Sitarama, and Monojit Choudhurya. "A New Dataset for Natural Language Inference from Code-mixed Conversations." In *LREC 2020 Workshop Language Resources and Evaluation Conference* 11–16 May 2020, p. 9.
- Choudhury, Monojit, Anirudh Srinivasan, and **Sandipan Dandapat**. "Processing and understanding mixed language data." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (**EMNLP-IJCNLP**): Tutorial Abstracts. 2019. (Core Rank: A*)
- Santy, Sebastin, **Sandipan Dandapat**, Monojit Choudhury, and Kalika Bali. "INMT: Interactive neural machine translation prediction." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (**EMNLP-IJCNLP**): System Demonstrations, pp. 103-108. 2019. (Core Rank: A*)
- Pratapa, Adithya, Gayatri Bhat, Monojit Choudhury, Sunayana Sitaram, **Sandipan Dandapat**, and Kalika Bali. "Language modeling for code-mixing: The role of linguistic theory based synthetic data." In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), pp. 1543-1553. **ACL** 2018. (Core Rank: A*)
- Sharma, Raksha, Pushpak Bhattacharyya, **Sandipan Dandapat**, and Himanshu Sharad Bhatt. "Identifying transferable information across domains for cross-domain sentiment classification." In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), pp. 968-978. **ACL** 2018. (Core Rank: A*)
- Kumar, Adarsh, **Sandipan Dandapat**, and Sushil Chordia. "Translating Web Search Queries into Natural Language Questions." In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018). 2018.
- **Dandapat, Sandipan**, and Christian Federmann. "Iterative data augmentation for neural machine translation: a low resource case study for English-Telugu." In Proceedings of the *21st Annual Conference of the European Association for Machine Translation*, pp. 307-312. EAMT 2018.
- **Dandapat, Sandipan**, and William Lewis. "Training deployable general domain mt for a low resource language pair: English–Bangla." In Proceedings of the *21st Annual Conference of the European Association for Machine Translation*. EAMT 2018.
- Shreshtha Mundra, Anirban Sen, Manjira Sinha, Sandya Mannarswamy, **Sandipan Dandapat**, and Shourya Roy. "Fine-Grained Emotion Detection in Contact Center Chat Utterances." In *Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2017*, pp. 337-349. Springer, Cham, 2017.
- Shourya Roy, **Sandipan Dandapat**, and Y. Narahari. "A Fluctuation Smoothing Approach for Unsupervised Automatic Short Answer Grading." *NLPTEA 2016* (2016): 82.
- Shourya Roy, **Sandipan Dandapat**, Ajay Nagesh, and Y. Narahari. "Wisdom of Students: A Consistent Automatic Short Answer Grading Technique." In *13th International Conference on Natural Language Processing, ICON 2016*, p. 178. 2016.
- Mariappan Ragunathan, Balaji Peddamuthu, Preethi R. Raajaratnam, **Sandipan Dandapat**, Neeta Pande, and Shourya Roy. "QART: A Tool for Quality Assurance in Real-Time in Contact Centers." In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, ***CIKM** 2016*, pp. 2493-2496. ACM, 2016. (Core Rank: A)
- **Sandipan Dandapat** and Andy Way. Improved Named Entity Recognition using Machine Translation-based Cross-lingual Information. *Computación y Sistemas* 20.3 (2016): 495-504.

- Himanshu S. Bhatt, **Sandipan Dandapat**, Peddamuthu Balaji, Shourya Roy, Sharmistha and Deepali Semwal. SODA : Service Oriented Domain Adaptation Architecture for Microblog Categorization. In *The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies , **NAACL HLT** 2013,* San Diego, CA. (Core Rank: A)
- **Sandipan Dandapat** and Andy Way. 2016. Improved Named Entity Recognition using Machine Translation-based Cross-lingual Information. In *CICLING 2016: Proceedings of 17th International Conference on Intelligent Text Processing and Computational Linguistics*, Konya, Turkey.
- Shourya Roy, Ragunathan Mariappan, **Sandipan Dandapat**, Saurabh Srivastava, Sainyam Galhotra, Balaji Peddamuthu. QA^{RT}: A System for Real-time Holistic Quality Assurance for Contact Center Dialogues. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, **AAAI** 2016*, Phoenix, AZ. (Core Rank: A*)
- Sharath Reddy Gunamgari, **Sandipan Dandapat** and Monojit Choudhur. Hierarchical Recursive Tagset for Annotating Cooking Recipes. In *Proceedings of the 11th International Conference on Natural Language Processing, ICON-2014***,** Goa, India.
- Santanu Pal, Ankit Srivastava, **Sandipan Dandapat**, Josef van Genabith, Qun Liu and Andy Way. USAAR-DCU Hybrid Machine Translation System for ICON 2014. In *Proceedings of the 11th International Conference on Natural Language Processing, ICON-2014***,** Goa, India.
- **Sandipan Dandapat,** Declan Groves. MTWatch: A Tool for the Analysis of Noisy Parallel Data. In *Proceedings of the 9th Edition of the Language Resource and Evaluation Conference, LREC 2014,* Reykjavik, Iceland. pp. 41-45.
- **Sandipan Dandapat,** Shakuntala Mahanta and Sibansu Mukhopadhyay. Fuzzy Match Score and Translation Memory Match: A Linguistic Insight. In *Proceedings of the 10th International Conference on Natural Language Processing, ICON 2013,* Delhi, India. pp. 1-10.
- Aswarth Abhilash Dara, **Sandipan Dandapat**, Declan Groves, Josef van Genabith. 2013. TMTprime: A Recommender System for MT and TM Integration. In *The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies , **NAACL HLT** 2013,* Atlanta, GA. pp. 10-13. (Core Rank: A)
- Johannes Leveling, Debasis Ganguly, **Sandipan Dandapat**, Gareth F. Jones. 2012. Approximate Sentence Retrieval for Scalable and Efficient Example-based Machine Translation. In *Proceedings of the 24th International Conference on Computation Linguistics, COLING 2012*, Mumbai, India. pp. 1571-1586.
- **Sandipan Dandapat**, Sara Morriessy, Andy Way and Josef van Genabith. 2012. Combining EBMT, SMT, TM and IR Technologies for Quality and Scale. In *Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra), a workshop in EACL 2012*, Avignon, France. pp.48-58.
- **Sandipan Dandapat**. 2011. Nitin Indurkhya and Fred J. Damerau (eds.): Handbook of Natural Language Processing (second edition) – CRC Press, Boca Ranton, 2010, xxxiii + 678 pp, Hardbound, ISBN 978-1-4200-8592-1. *Machine Translation.* 25(4):377-381.
- **Sandipan Dandapat**, Sara Morriessy, Andy Way and Mikel L. Forcada. 2011. Using Example-based MT to Support Statistical MT when Translating Homogeneous Data in Resource-Poor Settings. In *Proceedings of the 15th Annual Meeting of the European Association for Machine Translation, EAMT 2011*, Leuven, Belgium. pp. 201-208.
- ***S*andipan Dandapat**, Sara Morriessy, Sudip Kumar Naskar and Harold Somers. 2010. Statistically Motivated Example-based Machine Translation using Translation Memory. In *Proceedings of the 8th International Conference on Natural Language Processing, ICON 2010*, Kharagpur, India. pp. 168-177.
- **Sandipan Dandapat**, Sara Morriessy, Sudip Kumar Naskar and Harold Somers. 2010. Mitigating Problems in Analogy-based EBMT with SMT and vice versa: a Case Study with Named Entity Transliteration. In *Proceedings of the 24th Pacific Asia Conference on Language Information and Computation, PACLIC 2010*, Sendai, Japan.
- **Sandipan Dandapat**, Mikel L. Forcada, Declan Groves, Sergio Penkale, John Tinsley and Andy Way. 2010. OpenMaTrEx: A Free/Open-Source Marker-Driven Example-Based Machine Translation System. In *Proceedings of the 7th International Conference on Natural Language Processing, IceTAL 2010*, Reykjavik, Iceland. pp. 121-126.
- Sergio Penkale, Rejwanul Haque, **Sandipan Dandapat**, Pratyush Banerjee, Ankit K. Srivastava, Jinhua Du, Pavel Pecina, Sudip Kumar Naskar, Mikel L. Forcada, Andy Way. 2010. MaTrEx: The DCU MT

System for WMT 2010. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and Metrics MATR, ACL 2010*, Uppsala, Sweden. pp. 143-148.

- Sara Morrissey, Harold Somers, Robert Smith, Shane Gilchrist and **Sandipan Dandapat**. Building a Sign Language corpus for use in Machine Translation. In *Proceedings of the 4th Workshop on Representation and Processing of Sign Languages: Corpora for Sign Language Technologies*, 2010. Valetta, Malta. pp. 172-177.
- Harold Somers, **Sandipan Dandapat** and Sudip Kumar Naskar. A review of EBMT using proportional analogy. In *Proceedings of the 3rd Workshop on Example-Based Machine Translation*, 2009. Dublin, Ireland. pp. 53-60.
- **Sandipan Dandapat**, Priyanka Biswas, Monojit Choudhury and Kalika Bali. Co mplex Linguistic Annotation - No Easy Way Out! A Case from Bangla and Hindi POS Labeling Task. In *Proceedings of the 3rd Linguistic Annotation Workshop (LAW - III) 2009, a Workshop at the Joint conference of the 47th Annual Meeting of th e Association for Computational Linguistics and the 4th International Joint Conf erence on Natural Language Processing of the Asian Federation of Natural Languag e Processing*, Singapre. pp. 10-18.
- Rejwanul Haque, **Sandipan Dandapat**, Ankit Kumar Srivastava, Sudip Kumar Naskar and Andy Way. English-Hindi Transliteration Using Context-Informesd PB-SMT: the DCU System for NEWS 2009. In *Proceedings of the Named Entities Workshop (NEWS) 2009, a Workshop at the Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing*, Singapre. pp. 104-107.
- Cohan Sujay Carlos, Monojit Choudhury, **Sandipan Dandapat**. Large-Coverage Root Lexicon Extraction for Hindi. In *Proceedings of the 12th Conference of the European Chapter of the ACL, **EACL** 2009. Athens*, Greece. pp. 121-129. (Core Rank: A)
- Priyanka Biswas, **Sandipan Dandapat**, Kalika Bali, Monojit Choudhury. A Corpus-based Study of করে (kare) in Bangla: Theoretical and Computational Perspectives. In *Proceedings of the 6th International Conference on Natural Language Processing, ICON 2008*. Pune, India. pp.
- Sujan Kumar Saha, Sanjay Chatterjee, **Sandipan Dandapat**, Sudeshna Sarkar and Pabitra Mitra . A Hybrid Approach for Named Entity Recognition in Indian Languages. In *Proceedings of the IJCNLP-08 Workshop on Named Entity Recognition for South and South East Asian Languages, NERSSEAL 08*. Hyderabad, India. http://ltrc.iiit.ac.in/ner-ssea-08/. pp. 17-24.
- Tirthankar Dasgupta, **Sandipan Dandapat**, Anupam Basu. A Prototype Machine Translation System from Text-to-Indian Sign Language. In *Proceedings of the IJCNLP-08 Workshop NLP for Less Privileged Languages, NLPLPL 08*. Hyderabad, India. http://ltrc.iiit.ac.in/nlp-lpl-08/. pp. 19-26.
- Debasis Mandal, **Sandipan Dandapat**, Mayank Gupta, Pratyush Banerjee, Sudeshna Sarkar. Bengali and Hindi to English Cross-language Text Retrieval under Limited Resources. In *Proceedings of 8th Workshop of the Cross-Language Evaluation Forum*, Budapest, Hungary, 19-21 Sept 2007, LNCS, Springer.
- **Sandipan Dandapat,** Sudeshna Sarkar and Anupam Basu**.** Automatic Part-of-Speech Tagging for Bengali: An approach for Morphologically Rich Languages in a Poor Resource Scenario. In *Proceedings of the Association of Computational Linguistics (**ACL 2007**),* Prague, Czech Republic, pp. 221-224. (Core Rank: A*)
- **Sandipan Dandapat.** Part-of-Speech Tagging and Chunking with Maximum Entropy Model. In *Proceedings of the SPSAL Workshop*, IJCAI.2007.
- **Sandipan Dandapat** and Sudeshna Sarkar. Part of Speech Tagging for Bengali with Hidden Markov Model. In *Proceedings of the NLPAI Machine Learning Contest*, http://ltrc.iiit.ac.in/nlpai_contest06. June 2006.
- **Sandipan Dandapat,** Pabitra Mitra and Sudeshna Sarkar. Statistical Investigation of Bengali Noun-Verb (N-V) Collocations as Multi-word Expressions. In *Proceedings of the Symposium on Modeling and Shallow Parsing of Indian Languages (MSPIL 2006)*, IIT Bombay, INDIA, pp. 230-233, April 2006.
- **Sandipan Dandapat,** Sudeshna Sarkar and Anupam Basu. A hybrid part-of-speech tagging and it's application to Bengali. In *Proceedings of the International Conference on Computational Intelligence (ICCI 2004)*, Istanbul, TURKEY, pp. 169-172.

- Vibhab Agrwal, **Sandipan Dandapat**, Dipti Mishra Sharma and Rajeev Sangal. Linking Monolingual Resource with Bilingual Resource. *Proc. of the Symposium on Indian Morphology Phonology and Language Engineering (SIMPLE 2004)*. IIT Kharagpur, INDIA, March 2004.

Patents

- Sharma, Raksha, **Sandipan Dandapat**, and Himanshu Sharad Bhatt. "Method and system for training a target domain classifier to label text segments." U.S. Patent 10,460,257, issued October 29, 2019.
- **Dandapat, Sandipan**, Ragunathan Mariappan, Shourya Roy, and Shreshtha Mundra. "Method and system for data processing for real-time text analysis." U.S. Patent 10,210,157, issued February 19, 2019.
- **Dandapat Sandipan**. "Adult Query Intent Classification through Query Expansion". 2021. Filling in Progress. Approved (Filling in Progress).
- Shrawgi, Hari, Tushar Singhal, Madhur Jindal, Prasenjit Ghosh, **Sandipan Dandapat**, Parag Agarwal, Prasanjit Rath. "Generic eval framework for Responsible AI in LLMs : GRAIL". 2024. Approved (Filling in Progress).

Achievements

- Area Chair NAACL 2025
- Area Chair EMNLP 2024
- Tutorial Co-chair AIML Systems 2024
- Senior Area Chair EMNLP 2023
- Search Technology Centre India, Microsoft Spot Award 2022
- Engineering Excellence Award, Xerox Research Centre India, 2016
- Organizing committee member for International Linguistic Olympiad 2016
- Regional Coordinator for the Panini Linguistic Olympiad for Guwahati Chapter
- INVENT Commercialization Award 2013 for MTWatch Project
- Awarded first prize in NLPAI Machine Learning Contest 2006 for Part of Speech Tagging and Chunking for Indian Languages
- Review committee member of ACL, AAAI, TALIP, COLING, EMNLP