

Anubhav Jangra

Pre-Doctoral Researcher, Google Research

🌐 Homepage @ anubhav0603@gmail.com 🐙 Github 🎓 Google Scholar

Education

Jun 2021 | Indian Institute of Technology, Patna | GPA: 8.82/10, Major GPA: 9.12/10
Jul 2017 | Bachelor of Technology in Computer Science and Engineering

Research Experience

Present | Google Research | Advertising Sciences Team [🌐] | Bangalore, India
Jul 2021 | Pre-Doctoral Researcher | Advisor: [Dr. Aravindan Raghuvveer](#)
Exploring Natural Language Generation techniques for creative advertisement generation. Investigating several research areas like text style transfer, data-to-text generation, semantic representations etc.

Aug 2020 | GREYC Lab, ENSI-CAEN [🌐] | Remote / Caen, France
Jul 2020 | Research Intern | Advisor: [Prof. Gaël Dias](#)
Extended patch-based lexical semantic identification frameworks to a multi-modal setting. Developed the dataset and conducted the pilot studies of the project. [ACM MM'22]

Jun 2019 | Graduate School of Informatics, Kyoto University [🌐] | Kyoto, Japan
May 2019 | Research Intern | Advisor: [Prof. Adam Jatowt](#)
Explored various unsupervised optimization techniques to develop multi-modal summarization systems that generate text-image-audio-video summaries.

Jan 2019 | CFILT Lab, IIT Bombay [🌐] | Mumbai, India
Dec 2018 | Research Intern | Advisor: [Prof. Pushpak Bhattacharyya](#)
Examined Unsupervised NMT for distant language pairs (Indo-Aryan languages) using attention based seq2seq models.

Selected Research Projects

Text Style Transfer | Sept'21 - Jun'22

Advisors: [Dr. Aravindan Raghuvveer](#), [Dr. Preksha Nema](#)

- > Developed an AMR graph based framework to improve content preservation in generation. [EMNLP'22]
- > Proposed method significantly out-performs state-of-the-art techniques by achieving on an average **15.2% higher content preservation** with negligible loss (~3%) in style accuracy.
- > Performed human evaluations to illustrate that T-STAR has **50% lesser hallucinations** compared to SoTA TST models.

Multi-modal summarization

Jan'19 - Jul'21

Advisors: [Dr. Sriparna Saha](#), [Prof. Adam Jatowt](#), [Dr. Mohammed Hasanuzzaman](#)

- > Developed and implemented various systems using optimization techniques like integer linear programming, differential evolution, grey wolf optimizer etc. to solve text, image, and video summary generation. [ECIR'20, SIGIR'20, SIGIR'21]
- > Formally defined the complementary/supplementary enhanced multi-modal summaries, and achieved a new state-of-the-art on unsupervised MMS, surpassing the predecessor by almost **twice as better ROUGE-2 scores**. [SIGIR'21]
- > Wrote the first ever literature survey on multi-modal summarization. [under submission]
- > First work towards topic-aware multi-modal news summarization. [Findings ACL'22]
- > Curated large-scale multi-modal multi-lingual summarization corpus spanning over 20 languages. [ARR submission]

Automatic Text Summarization

Jul'19 - Dec'21

Advisors: [Dr. Sriparna Saha](#), [Prof. Pushpak Bhattacharyya](#)

- > **Extractive Summarization.** Utilized nature-inspired algorithms like Differential Evolution, Grey Wolf Optimizer, Water Cycle Algorithm etc. in a multi-objective optimization framework to generate extractive summaries. [KBS'18]
- > **Abstractive Summarization.** Proposed an RL-based 'extractor-abtractor' framework to outperform its predecessors by a margin of **0.5 ROUGE-1, 0.4 ROUGE-2, 1 METEOR, and 0.9 WMS scores**. A knowledge discovery that seq2seq networks like PGN model implicitly extract and paraphrases sentences was brought to light through this work. [ICON'20]
- > **Evaluation Metrics.** Proposed WIDAR, a ROUGE-based evaluation metric that evaluates generated summary by taking into account both the reference summary and input document. WIDAR correlates better than ROUGE by **26%, 76%, 82%, and 15% in coherence, consistency, fluency, and relevance** on human judgement scores provided in the SummEval dataset. It was able to obtain comparable results with the SOTA while requiring $\sim \frac{1}{64}^{th}$ of computational time. [ECIR'22]

Accepted Works

- [C.7] **T-STAR: Truthful Style Transfer using AMR Graph as Intermediate Representation**
 Anubhav Jangra*, Preksha Nema*, Aravindan Raghuvier
The 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, UAE [EMNLP'22]
- [C.6] **Topic-aware Multimodal Summarization**
 Sourajit Mukherjee, Anubhav Jangra, Sriparna Saha, Adam Jatowt,
2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics [Findings in ACL'22]
- [C.5] **Combining Vision and Language Representations for Patch-based Identification of Lexico-Semantic Relations** [🔗]
 Prince Jha, Gaël Dias, Alexis Lechervy, José G Moreno, Anubhav Jangra, Sebastião Pais, Sriparna Saha
30th ACM International Conference on Multimedia, Lisbon, Portugal [ACM MM'22]
- [C.4] **MAKED: Multi-lingual Automatic Keyword Extraction Dataset** [🔗]
 Yash Verma, Anubhav Jangra, Sriparna Saha, Adam Jatowt, Dwaipayan Roy
13th Conference on Language Resources and Evaluation [LREC'22]
- [C.3] **WIDAR - Weighted Input Document Augmented ROUGE** [🔗]
 Raghav Jain*, Vaibhav Mavi*, Anubhav Jangra*, Sriparna Saha
44th European Conference on Information Retrieval, Stavanger, Norway [ECIR'22]
- [J.1] **Identifying Complaints based on Semi-Supervised Mincuts** [🔗]
 Apoorva Singh, Sriparna Saha, Mohammed Hasanuzzaman, Anubhav Jangra
Elsevier's Expert Systems with Applications, Volume 186, 2021 [ESWA'21]
- [C.2] **Multi-modal Supplementary Complementary Summarization using Multi-Objective Optimization** [🔗]
 Anubhav Jangra, Sriparna Saha, Adam Jatowt, Mohammed Hasanuzzaman
44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual [SIGIR'21]
- [C.1] **Semantic Extractor Paraphraser based Abstractive Summarization** [🔗]
 Anubhav Jangra*, Raghav Jain*, Vaibhav Mavi*, Sriparna Saha, Pushpak Bhattacharyya,
17th International Conference on Natural Language Processing, Patna, India [ICON'20]
- [SP.2] **Multi-Modal Summary Generation using Multi-objective Optimization** [🔗]
 Anubhav Jangra, Sriparna Saha, Adam Jatowt, Mohammed Hasanuzzaman,
43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Xi'an, China [SIGIR'20]
- [SP.1] **Text-Image-Video Summary Generation using Joint Integer Linear Programming** [🔗]
 Anubhav Jangra, Adam Jatowt, Mohammed Hasanuzzaman, Sriparna Saha,
42nd European Conference on Information Retrieval, Lisbon, Portugal [ECIR'20]
- [J.1] **Extractive Single Document Summarization using Multiobjective Optimization: Exploring Self-organized Differential Evolution, Grey Wolf Optimizer and Water Cycle Algorithm** [🔗]
 Naveen Saini, Sriparna Saha, Anubhav Jangra, Pushpak Bhattacharyya,
Elsevier's Knowledge Based Systems, 2018 [KBS'18]



Under Submission

- [S.5] **A Survey on Medical Document Summarization**
 Raghav Jain, Anubhav Jangra, Adam Jatowt, Sriparna Saha
ACM Computing Surveys [ACM CSUR]
- [S.4] **Can Multimodal Pointer Generator Transformers produce topically relevant summaries?**
 Sourajit Mukherjee, Anubhav Jangra, Sriparna Saha, Adam Jatowt,
45th European Conference on Information Retrieval (ECIR), Dublin, Ireland [ECIR'23]
- [S.3] **Large Scale Multi-modal Multi-lingual Summarization Dataset**
 Yash Verma, Anubhav Jangra, Sriparna Saha
ACL Rolling Reviews [ARR]
- [S.2] **A Survey on Multi-hop Question Answering** [🔗]
 Vaibhav Mavi, Anubhav Jangra, Adam Jatowt
ACM Computing Surveys [ACM CSUR]
- [S.1] **A Survey on Multi-modal Summarization** [🔗]
 Anubhav Jangra, Adam Jatowt, Sriparna Saha, Mohammed Hasanuzzaman,
ACM Computing Surveys [ACM CSUR]


Talks


- | | |
|---|----------|
| › Abstract Meaning Representation (AMR) Graphs at work! – AI-NLP-ML Lab, IIT Patna | Oct 2022 |
| › Automatic Text Summarization PyData Patna Conference | Dec 2020 |

Other Experiences

Jul 2021	IBM	Remote / Chennai, India
Jun 2021	Global Research Mentee / Advisor: Ganesan Narayanasamy	
	Developed the project framework for <i>Health Care App</i> , that uses knowledge graphs and named-entity recognition to help users self-diagnose themselves.	
Mar 2021	Huawei Technologies Co., Ltd	Remote
Dec 2020	Project Member / Advisor: Dr. Sriparna Saha	
	Developed a Proof of Concept (POC) for the task of automatic tagline generation and product description using existing neural summarization systems for the upcoming collaborative project of IIT Patna and Huawei.	
Jan 2020	TCS Innovation Lab 	Kolkata, India
Dec 2019	Research Intern / Advisor: Arijit Ukil	
	Investigated generative modeling to tackle the insufficiency of data in time-series signal classification.	
Jul 2021	AI-NLP-ML Lab, IIT Patna 	Patna, India
Jul 2018	Undergraduate Research Scholar / Advisor: Dr. Sriparna Saha	
	Worked extensively in the area of summarization (e.g., multi-modal summarization, extractive and abstractive text summarization), complaint mining and multi-label classification.	

Honours and Awards

Google Research AI summer school, 2020  One of the 50 participants out of 1000+ applicants in the NLU track.

MSU-IITR-IISc course and workshop  Attended a short term course and workshop on “Pragmatic Optimization for Practical Problem Solving” conducted by Michigan State university, IIT Roorkee and IISc Bangalore, limited to 40 students.

IIT JEE Ranked in National Top 0.2% (amongst 1,400,000 candidates) in JEE Mains 2017 and Top 1.5% (amongst 2,00,000 candidates) in IIT-JEE Advanced 2017.

Teaching and Leadership Roles

NLP & QuaD-LR Reading Group, Google Research, Bangalore, India	Participant	Oct 2021-Present
› Active participant in our weekly reading group where I regularly present research papers and engage in discussions.		
Google DSC IIT Patna, Patna, India	ML Department Lead	2019-2020
› Supervised three projects and gave lectures on Machine Learning theory and its applications.		
Univeristy of Innsbruck, Austria	Teaching Assistant	Jun 2020
› Part-time Teaching Assistant in the course 2021S703836 VU (Natural Language Processing). Prepared lectures on automatic summarization.		

Academic Service

Reviewer (Journals)	ACM Computing Surveys (since Jan 2021), ACM TALLIP (since May 2020), Applied Artificial Intelligence (since Oct 2021), and IEEE Transactions on Computational Social Systems (since Jan 2022), Expert Systems with Applications (since Sept 2022), Engineering Applications of Artificial Intelligence (since Feb 2022), IEEE Internet Computing (since Feb 2022)
Secondary Reviewer	AAAI 2020, EACL 2021, ACL 2021, EMNLP 2021, CIKM 2021, and WebConf2022
Mentor	Mentored three undergraduate interns, two masters student researchers, and one undergraduate student researcher as part of the AI-NLP-ML lab, IIT Patna.
Volunteer	Volunteered as a reviewer in CS Research Mentorship program to help review applicants from historically marginalized communities for the mentorship program.

Miscellaneous

- › **AnthroKrishi project at Google:** Conducted semi-structured interviews with farmers on understanding motivations and barriers for changing farming practices for carbon sequestration.
- › Outside of work, I love to create origami and write in calligraphy.
- › I have gracefully failed at learning violin in the past.