Key Takeaways from the Second Shared Task on Indian Language Summarization (ILSUM 2023)

Shrey Satapara[1], Parth Mehta[2], Sandip Modha[3] and Debasis Ganguly[4]

[1]*Indian Institute of Technology Hyderabad, India*
[2]*Parmonic, USA*
[3]*LDRP-ITR, Gandhinagar, India*
[4]*University of Glasgow, Scotland, UK*

Abstract

This paper provides an overview of the second edition of the shared task on Indian Language Summarization (ILSUM) organized at the 15th Forum for Information Retrieval Evaluation (FIRE 2023). This edition builds upon ILSUM 2022 by creating additional benchmark data for text summarization in Indian languages. Apart from expanding the datasets of the three languages from the previous edition, namely Hindi, Gujarati and Indian English, a new Bengali dataset was introduced this year. In addition to this, a new misinformation detection subtask was introduced. ILSUM 2023 saw an enthusiastic response, with registrations from over 35 teams. A total of 6 teams submitted runs across both subtasks and 4 teams submitted working notes. Standard ROUGE metrics as well as Bert-score were used as the evaluation metric for the summarization subtask, while macro-F1 score was used for the misinformation detection subtask.

Keywords

Automatic Text Summarization, Indian Languages, Headline Generation, Misinformation Detection

1. Introduction

The second shared task on Indian Language Summarization was continuation of the efforts for bridging the gap in progress of NLP research between resource-rich languages like English, Spanish, Chinese, etc as opposed to more resource-constrained Indian languages. Platforms like the Forum for Information Retrieval Evaluation (FIRE)[1] has been consistently trying to bridge this gap by building reusable and open source test collections. The progress has been noteworthy in several language-dependent tasks like hate speech detection[2, 3, 4, 5, 6], Sentiment analysis[7, 8], mixed script IR[9, 10], Fake news detection[11, 12], authorship attribution[13, 14] as well as language independent tasks like Indian legal document retrieval and summarization[15, 16, 17, 18, 19, 20], IR from microblogs[21], IR for software engineering[22], etc. Several large-scale datasets and pre-trained models have become publicly available. AI-

Forum for Information Retrieval Evaluation, December 15-18, 2023, India
shreysatapara@gmail.com (S. Satapara); parth.mehta126@gmail.com (P. Mehta); sjmodha@gmail.com (S. Modha); debforit@gmail.com (D. Ganguly)
0000-0001-6222-1288 (S. Satapara); 0000-0002-4509-1298 (P. Mehta); 0000-0003-2427-2433 (S. Modha); 0000-0003-0050-7138 (D. Ganguly)

4BHARAT[1] is another initiative that is playing a pivotal role in bridging this gap, especially in machine translation and Indian language LLMs.

With the series of ILSUM tasks[23, 24, 25] we aim to replicate this for Automatic text summarization where research is skewed towards English[26, 27, 28] and other resource-rich languages, while the focus on other resource-poor languages is almost negligible[29]. Previous attempts at building test collections for Indian language summarization were limited in scope with at most a few dozen documents[30, 31, 32, 33, 34, 35]. Moreover, most of these datasets are either not public or are too small to be useful. In contrast, ILSUM 2023 dataset consists of 20,000 article-summary pairs for Hindi, Gujarati, Bengali and Indian Languages. Table 1 presents the details of the ILSUM dataset. The task is to generate a meaningful summary, either extractive or abstractive, for each article.

We also introduce a new subtask on misinformation detection in LLM generated summaries. This subtask was limited to Indian English in the current edition. The recent success in language generation capabilities of large language models (LLMs) [36], such as GPT [37], Llama [38] etc., has raised concerns about their possible misuse for generating fake news and spreading misinformation. This problem can easily extend to summaries where instead of fabricating an entire story, miscreants can use a real new article and generate a summary tailored to suit their purpose. In this subtask participants are given a machine generated summary and the task is to identify if the content in the summary are correct, or if they fall into one of four categories of misinformation namely incorrect numerical quantities, fabrication, false attribution or misrepresentation. Both subtasks are explained in detail in the next section, followed by a description of the approaches used by the participating teams.

2. Task Definition

The second shared task on Indian Language Summarization continued the effort of creating benchmark datasets for text summarization in Indian languages. The current edition saw the inclusion of Bengali alongside Hindi, Gujarati and Indian English. Bengali is one of the most widely spoken languages in the world with over 250 million speakers, the majority of them from India and Bangladesh. Datasets for all languages in ILSUM 2022[cite ilsum] were extended to include more articles and summaries. Apart from this we also introduced a new subtask on misinformation detection in machine-generated summaries. In the following subsections, we discuss in detail both tasks and the corresponding datasets.

2.1. Task 1 Text Summarization For Indian Languages

The objective of this task is the same as the first edition of ILSUM, which follows the standard definition of text summarization task (Given an article, participants are asked to generate a fixed-length summary in either an abstractive or extractive way). This year, we extended by

[1] https://ai4bharat.iitm.ac.in

adding approximately 15000 more articles on top of the previous edition's dataset and introduced one more language. As introduced in the previous edition, the current dataset poses a unique challenge of code-mixing and script mixing. It is very common for news articles to borrow phrases from English, even if the article itself is written in an Indian Language.

Examples like these are a common occurrence both in the headlines as well as in the articles.

- Gujarati: ”IND vs SA, 5મી T20 તસવીરોમાં: વરસાદે વિલન બની મજા બગાડી” (India vs SA, 5th T20 in pictures: rain spoils the match)

- Hindi: ”LIC के IPO में पैसा लगाने वालों का टूटा दिल, आई एक और नुकसानदेह खबर” (Investors of LIC IPO left broken hearted, yet another bad news)

Language	Training Set	Test Set	Total
Hindi	21225	3000	24225
Gujarati	33630	2999	36629
Bengali	12356	2951	15307
English	28342	2895	31237

Table 1
Training and Test Data Distribution for Different Languages in Task 1

2.2. Task 2 Detecting Factual Incorrectness in Machine-Generated Summaries

This task aims to identify incorrectness in machine-generated summaries, which is an important step in ensuring the reliability and accuracy of information. While evaluating these summaries against the original article, the key focus is to detect and classify different types of incorrectness. In this task, we provided the dataset with four different types of inaccuracies along with a fifth class containing correct summaries. We use the GPT-4 model to generate incorrect summaries of each class, and the GPT-3.5 model to produce the correct summaries using carefully crafted prompts to generate automatic summaries for each type of incorrectness without any manual intervention. Following are the types of incorrectness present in the dataset. Detailed description of how the dataset was created is available in [39].

- **Misrepresentation**: This involves presenting information in a way that is misleading or that gives a false impression. This could be done by exaggerating certain aspects, understating others, or twisting facts to fit a particular narrative.
- **Inaccurate Quantities or Measurements**: Factual incorrectness can occur when precise quantities, measurements, or statistics are misrepresented, whether through error or intent.
- **False Attribution**: Incorrectly attributing a statement, idea, or action to a person or group is another form of factual incorrectness.
- **Fabrication**: Making up data, sources, or events is a severe form of factual incorrectness. This involves creating ”facts” that have no basis in reality.

For this task, text articles and generated summaries are provided with one associated label of the type of incorrectness in training data. Still, participants are asked to predict all possible labels associated with text summaries in test data, as one summary can have multiple types of incorrectness. Example Article with all types of incorrectness is available at https://ilsum.github.io/ilsum/2023/index.html. Table 2 contains dataset statistics for Task 2 dataset. The class predictions on test data are evaluated using Macro F1 score.

Class	Training Set	Test Set	Total
Misrepresentation	294	25	319
Inaccurate Quantities	195	10	205
False Attribution	250	13	263
Fabrication	250	32	282
Correct	5000	143	5143

Table 2
Task 2 Dataset Statistics

3. Results and Disussion

In this section, we discuss results of the participating teams. Compared to the last edition, where we only used the ROUGE score for evaluation, we added another ranking based on the BERT Score for a fair evaluation of abstractive summaries. However, we observe very high co-relation between the BERT score and ROUGE. Especially the rankings of the system are exactly same irrespective of the choice of metric. Below we report the results and approaches used for each of the task and language.

3.1. Task 1 Hindi

For text summarization in Hindi two teams submitted a total of 6 runs. Team Irlab-IITBHU utilized name entity-aware text summarization, NER emerges as important factor to extract in-depth information and prioritising key entities for the summary by utilizing a pre-trained Muril-based HindiNER model and fine-tuning MBART-50(rank 1), mT5 with name entities(rank 2), IndicBART(rank 3), IndicBARTSS(rank 4) and indicBART with name entities(rank 6). Table 3 contains results of all submissions for text summarization in Hindi.

rank	Team Name	BERT SCORE			ROUGE (F1 Scores)			
		Precision	Recall	F1 Score	Rouge-1	Rouge-2	Rouge-4	Rouge-L
1	Irlab-IITBHU	0.8226	0.8048	0.813	0.5625	0.4715	0.4032	0.5373
2	Irlab-IITBHU	0.797	0.8073	0.8017	0.5409	0.4592	0.4007	0.5153
3	Irlab-IITBHU	0.8085	0.7948	0.8008	0.5359	0.4551	0.3973	0.5128
4	Irlab-IITBHU	0.8005	0.8003	0.7998	0.5328	0.4496	0.3912	0.5084
5	BITS Pilani	0.7609	0.682	0.7186	0.2988	0.1707	0.1196	0.2476
6	Irlab-IITBHU	0.7153	0.7037	0.7089	0.2801	0.1568	0.0836	0.2423

Table 3
Performance of teams on Language summarization in Hindi

3.2. Task 1 Gujarati and Bengali

For Gujarati and Bengali text summarisation, only one team submitted only one submission. Team *BITS Pilani* fine-tuned mT5(mT5-multilingual-XLSum) model on the ILSUM dataset for all four languages. Results for text summarization in Gujarati and Bengali are available in Table 4 and Table 5

rank	Team Name	BERT SCORE			ROUGE (F1 Scores)			
		Precision	Recall	F1 Score	Rouge-1	Rouge-2	Rouge-4	Rouge-L
1	BITS Pilani	0.7423	0.688	0.7135	0.174	0.0747	0.0333	0.1655

Table 4
Performance of teams on Language summarization in Gujarati

rank	Team Name	BERT SCORE			ROUGE (F1 Scores)			
		Precision	Recall	F1 Score	Rouge-1	Rouge-2	Rouge-4	Rouge-L
1	BITS Pilani	0.7058	0.6554	0.679	0.12	0.0567	0.0254	0.1087

Table 5
Performance of teams on Language summarization in Bengali

3.3. Task 1 English

For English, four teams submitted one run each. Team *NITK - AI* outperformed other teams where they fine-tuned T5-base on ILSUM English dataset. Team *Eclipse* also fine-tuned T5-base model standing second on the leaderboard. Results of all four submissions by all teams are available in Table 6.

rank	Team Name	BERT SCORE			ROUGE (F1 Scores)			
		Precision	Recall	F1 Score	Rouge-1	Rouge-2	Rouge-4	Rouge-L
1	NITK - AI	0.8752	0.8684	0.8716	0.3321	0.1731	0.121	0.282
2	Eclipse	0.8505	0.8733	0.8616	0.3022	0.1111	0.042	0.2504
3	BITS Pilani	0.8724	0.8462	0.8589	0.2354	0.0604	0.0147	0.182
4	ASH	0.8277	0.8036	0.8153	0.137	0.017	0.0004	0.1181

Table 6
Performance of teams on Language summarization in English

3.4. Task 2 Detecting Factual Incorrectness in Machine-Generated Summaries

In this subtask, only one team submitted five runs, exploring zero-shot prompting using GPT-3.5 Turbo. Where they explored zero-shot prompting to identify if an article and summary pair belong to a particular class or not with different order of classes. The best result they obtained was by using an ensemble of predictions from all four different class orders they explored. The results obtained on this task are available in Table 7

Class	F1 Score
Fabrication	0.152
False Attribution	0.093
Incorrect Quantities	0.291
Misrepresentation	0.335
MACRO F1	0.527

Table 7
Performance of the participation team on Misinformation detection task

4. Conclusion and Future Work

The Indian Language Summarization (ILSUM) track at FIRE 2023 continued the efforts to create benchmark corpora for text summarization in Indian languages. Two major updates from last year were inclusion of Bengali in the summarization task, and inclusion of a new subtask on misinformation detection in machine generated summaries. Like previous edition majority of the summarization systems for task 1 were based on pre-trained large language models like MT5, MBart, and IndicBART. A notable exception was the approach proposed by IIT-BHU who used a combination of NER and pretrained language models. It was also the best performing approach for Hindi, highlighting scope for improvements over pre-trained LLMs. In the next edition of the ILSUM we plan to extend the summarization subtask to new languages, especially Dravidian languages. For the misinformation detection subtask we aim at providing fine-grain annotations about the part of summaries which are factually incorrect instead of simply labelling the entire summary as incorrect.

References

[1] P. Mehta, T. Mandl, P. Majumder, S. Gangopadhyay, Report on the FIRE 2020 evaluation initiative, SIGIR Forum 55 (2021) 3:1–3:11. URL: https://doi.org/10.1145/3476415.3476418. doi:10.1145/3476415.3476418.

[2] T. Mandl, S. Modha, G. K. Shahi, H. Madhu, S. Satapara, P. Majumder, J. Schäfer, T. Ranasinghe, M. Zampieri, D. Nandini, A. K. Jaiswal, Overview of the HASOC subtrack at FIRE 2021: Hatespeech and offensive content identification in english and indo-aryan languages, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation, Gandhinagar, India, December 13-17, 2021, volume 3159 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 1–19. URL: http://ceur-ws.org/Vol-3159/T1-1.pdf.

[3] T. Mandl, S. Modha, G. K. Shahi, A. K. Jaiswal, D. Nandini, D. Patel, P. Majumder, J. Schäfer, Overview of the HASOC track at FIRE 2020: Hate speech and offensive content identification in indo-european languages, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2020 - Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, volume 2826 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 87–111. URL: http://ceur-ws.org/Vol-2826/T2-1.pdf.

[4] S. Modha, T. Mandl, P. Majumder, D. Patel, Overview of the HASOC track at FIRE 2019: Hate speech and offensive content identification in indo-european languages, in: P. Mehta, P. Rosso, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019, volume 2517 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 167–190. URL: http://ceur-ws.org/Vol-2517/T3-1.pdf.

[5] H. Madhu, S. Satapara, S. Modha, T. Mandl, P. Majumder, Detecting offensive speech in conversational code-mixed dialogue on social media: A contextual dataset and benchmark experiments, Expert Systems with Applications (2022) 119342.

[6] S. Modha, P. Majumder, T. Mandl, C. Mandalia, Detecting and visualizing hate speech in social media: A cyber watchdog for surveillance, Expert Syst. Appl. 161 (2020) 113725. URL: https://doi.org/10.1016/j.eswa.2020.113725. doi:10.1016/j.eswa.2020.113725.

[7] M. Subramanian, R. Ponnusamy, S. Benhur, K. Shanmugavadivel, A. Ganesan, D. Ravi, G. K. Shanmugasundaram, R. Priyadharshini, B. R. Chakravarthi, Offensive language detection in tamil youtube comments by adapters and cross-domain knowledge transfer, Comput. Speech Lang. 76 (2022) 101404. URL: https://doi.org/10.1016/j.csl.2022.101404. doi:10.1016/j.csl.2022.101404.

[8] B. R. Chakravarthi, P. K. Kumaresan, R. Sakuntharaj, A. K. Madasamy, S. Thavareesan, B. Premjith, S. K, S. C. Navaneethakrishnan, J. P. McCrae, T. Mandl, Overview of the hasoc-dravidiancodemix shared task on offensive language detection in tamil and malayalam, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation, Gandhinagar, India, December 13-17, 2021, volume 3159 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 589–602. URL: http://ceur-ws.org/Vol-3159/T3-1.pdf.

[9] S. Banerjee, K. Chakma, S. K. Naskar, A. Das, P. Rosso, S. Bandyopadhyay, M. Choudhury, Overview of the mixed script information retrieval (MSIR) at FIRE-2016, in: P. Majumder, M. Mitra, P. Mehta, J. Sankhavara, K. Ghosh (Eds.), Working notes of FIRE 2016 - Forum for Information Retrieval Evaluation, Kolkata, India, December 7-10, 2016, volume 1737 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2016, pp. 94–99. URL: http://ceur-ws.org/Vol-1737/T3-1.pdf.

[10] P. Gupta, K. Bali, R. E. Banchs, M. Choudhury, P. Rosso, Query expansion for mixed-script information retrieval, in: S. Geva, A. Trotman, P. Bruza, C. L. A. Clarke, K. Järvelin (Eds.), The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14, Gold Coast , QLD, Australia - July 06 - 11, 2014, ACM, 2014, pp. 677–686. URL: https://doi.org/10.1145/2600428.2609622. doi:10.1145/2600428.2609622.

[11] M. Amjad, G. Sidorov, A. Zhila, Data augmentation using machine translation for fake news detection in the urdu language, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020, European Language Resources Association, 2020, pp. 2537–2542. URL: https://aclanthology.org/2020.lrec-1.309/.

[12] M. Amjad, N. Ashraf, A. Zhila, G. Sidorov, A. Zubiaga, A. F. Gelbukh, Threatening language detection and target identification in urdu tweets, IEEE Access

9 (2021) 128302–128313. URL: https://doi.org/10.1109/ACCESS.2021.3112500. doi:10.1109/ACCESS.2021.3112500.
[13] P. Mehta, P. Majumder, Optimum parameter selection for K.L.D. based authorship attribution in gujarati, in: Sixth International Joint Conference on Natural Language Processing, IJCNLP 2013, Nagoya, Japan, October 14-18, 2013, Asian Federation of Natural Language Processing / ACL, 2013, pp. 1102–1106. URL: https://aclanthology.org/I13-1155/.
[14] P. Mehta, P. Majumder, Large scale quantitative analysis of three indo-aryan languages, J. Quant. Linguistics 23 (2016) 109–132. URL: https://doi.org/10.1080/09296174.2015.1071151. doi:10.1080/09296174.2015.1071151.
[15] P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, P. Mehta, A. Bhattacharya, P. Majumder, Overview of the FIRE 2019 AILA track: Artificial intelligence for legal assistance, in: P. Mehta, P. Rosso, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019, volume 2517 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 1–12. URL: http://ceur-ws.org/Vol-2517/T1-1.pdf.
[16] P. Bhattacharya, P. Mehta, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya, P. Majumder, FIRE 2020 AILA track: Artificial intelligence for legal assistance, in: P. Majumder, M. Mitra, S. Gangopadhyay, P. Mehta (Eds.), FIRE 2020: Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, ACM, 2020, pp. 1–3. URL: https://doi.org/10.1145/3441501.3441510. doi:10.1145/3441501.3441510.
[17] V. Parikh, U. Bhattacharya, P. Mehta, A. Bandyopadhyay, P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya, P. Majumder, AILA 2021: Shared task on artificial intelligence for legal assistance, in: D. Ganguly, S. Gangopadhyay, M. Mitra, P. Majumder (Eds.), FIRE 2021: Forum for Information Retrieval Evaluation, Virtual Event, India, December 13 - 17, 2021, ACM, 2021, pp. 12–15. URL: https://doi.org/10.1145/3503162.3506571. doi:10.1145/3503162.3506571.
[18] V. Parikh, V. Mathur, P. Mehta, N. Mittal, P. Majumder, Lawsum: A weakly supervised approach for indian legal document summarization, CoRR abs/2110.01188 (2021). URL: https://arxiv.org/abs/2110.01188. arXiv:2110.01188.
[19] S. Ghosh, A. Wyner, Identification of rhetorical roles of sentences in indian legal judgments, in: Legal Knowledge and Information Systems: JURIX 2019: The Thirty-second Annual Conference, volume 322, IOS Press, 2019, p. 3.
[20] S. Parashar, N. Mittal, P. Mehta, Casrank: A ranking algorithm for legal statute retrieval, Multimedia Tools and Applications (2023) 1–18.
[21] M. Basu, S. Ghosh, K. Ghosh, Overview of the fire 2018 track: Information retrieval from microblogs during disasters (irmidis), in: Proceedings of the 10th annual meeting of the Forum for Information Retrieval Evaluation, 2018, pp. 1–5.
[22] S. Majumdar, A. Bandyopadhyay, S. Chattopadhyay, P. P. Das, P. D. Clough, P. Majumder, Overview of the irse track at fire 2022: Information retrieval in software engineering, in: Forum for Information Retrieval Evaluation, ACM, 2022.
[23] S. Satapara, B. Modha, S. Modha, P. Mehta, Fire 2022 ilsum track: Indian language summarization, in: Proceedings of the 14th Forum for Information Retrieval Evaluation, ACM, 2022.
[24] S. Satapara, P. Mehta, S. Modha, D. Ganguly, Indian language summarization at fire 2023,

in: Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE 2023, Goa, India. December 15-18, 2023, ACM, 2023.
[25] S. Satapara, B. Modha, S. Modha, P. Mehta, Findings of the first shared task on indian language summarization (ILSUM): approaches challenges and the path ahead, in: K. Ghosh, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation, Kolkata, India, December 9-13, 2022, volume 3395 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 369–382. URL: https://ceur-ws.org/Vol-3395/T6-1.pdf.
[26] P. Mehta, From extractive to abstractive summarization: A journey, in: H. He, T. Lei, W. Roberts (Eds.), Proceedings of the ACL 2016 Student Research Workshop, Berlin, Germany, August 7-12, 2016, Association for Computational Linguistics, 2016, pp. 100–106. URL: https://doi.org/10.18653/v1/P16-3015. doi:10.18653/v1/P16-3015.
[27] P. Mehta, P. Majumder, Effective aggregation of various summarization techniques, Inf. Process. Manag. 54 (2018) 145–158. URL: https://doi.org/10.1016/j.ipm.2017.11.002. doi:10.1016/j.ipm.2017.11.002.
[28] S. Modha, P. Majumder, T. Mandl, R. Singla, Design and analysis of microblog-based summarization system, Social Network Analysis and Mining 11 (2021) 1–16. URL: https://doi.org/10.1007/s13278-021-00830-3.
[29] S. Sinha, G. N. Jha, An overview of indian language datasets used for text summarization, CoRR abs/2203.16127 (2022). URL: https://doi.org/10.48550/arXiv.2203.16127. doi:10.48550/arXiv.2203.16127. arXiv:2203.16127.
[30] S. Barve, S. Desai, R. Sardinha, Query-based extractive text summarization for sanskrit, in: S. Das, T. Pal, S. Kar, S. C. Satapathy, J. K. Mandal (Eds.), Proceedings of the 4th International Conference on Frontiers in Intelligent Computing: Theory and Applications, FICTA 2015, Durgapur, India, 16-18 November 2015, volume 404 of *Advances in Intelligent Systems and Computing*, Springer, 2015, pp. 559–568. URL: https://doi.org/10.1007/978-81-322-2695-6_47. doi:10.1007/978-81-322-2695-6_47.
[31] R. R. Chowdhury, M. T. Nayeem, T. T. Mim, M. S. R. Chowdhury, T. Jannat, Unsupervised abstractive summarization of bengali text documents, in: P. Merlo, J. Tiedemann, R. Tsarfaty (Eds.), Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021, Association for Computational Linguistics, 2021, pp. 2612–2619. URL: https://doi.org/10.18653/v1/2021.eacl-main.224. doi:10.18653/v1/2021.eacl-main.224.
[32] J. D'Silva, U. Sharma, Development of a konkani language dataset for automatic text summarization and its challenges, International Journal of Engineering Research and Technology. International Research Publication House. ISSN (2019) 0974–3154.
[33] V. R. Embar, S. R. Deshpande, A. Vaishnavi, V. Jain, J. S. Kallimani, saramsha-a kannada abstractive summarizer, in: 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2013, pp. 540–544.
[34] S. Gandotra, B. Arora, Feature selection and extraction for dogri text summarization, in: Rising Threats in Expert Applications and Solutions, Springer, 2021, pp. 549–556.
[35] R. Kabeer, S. M. Idicula, Text summarization for malayalam documents - an experience, in: International Conference on Data Science & Engineering, ICDSE 2014, Kochi, India, August 26-28, 2014, IEEE, 2014, pp. 145–150. URL: https://doi.org/10.1109/ICDSE.2014.

6974627. doi:10.1109/ICDSE.2014.6974627.

[36] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving language understanding with unsupervised learning, 2018. URL: https://openai.com/research/language-unsupervised.

[37] OpenAI, Gpt-4 technical report, 2023. arXiv:2303.08774.

[38] R. Taori, I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, P. Liang, T. B. Hashimoto, Stanford Alpaca: An Instruction-following LLaMA model, 2023. URL: https://github.com/tatsu-lab/stanford_alpaca, publication Title: GitHub repository.

[39] S. Satapara, P. Mehta, D. Ganguly, S. Modha, Fighting fire with fire: Adversarial prompting to generate a misinformation detection dataset, 2024. arXiv:arXiv:2401.04481.