Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Urdu-LC: Length-Controlled Abstractive Summarization for a Low-Resource Language

Kashif Gulzar
Faculty of Computer Science & Information Technology, The Superior University Lahore, Pakistan.

Shahbaz Ahmad *
Faculty of Computer Science & Information Technology, The Superior University Lahore, Pakistan Email: ahmad.fsd@superior.edu.pk

Shahzad Nazir
Baba Guru Nanak University, Nanakana Sahib, Punjab, Pakistan

Dr. Ahmad Khan
Faculty of Computer Science & Information Technology, The Superior University Lahore, Pakistan.

Sqlain Majeed
Faculty of Computer Science & Information Technology, The Superior University Lahore, Pakistan

ABSTRACT

The rapid expansion of digital text has intensified the need for effective summarization techniques, particularly for languages with limited computational resources. Although recent progress in abstractive summarization has produced strong results for high-resource languages, comparable developments for Urdu remain constrained due to data sparsity, morphological variation, and syntactic ambiguity. Another persistent issue in existing systems is the absence of explicit length control, despite its importance in applications that require summaries of predetermined size, such as headlines or mobile notifications. In this study, we propose a length-controlled abstractive summarization framework tailored for Urdu. The approach builds on multilingual transformer architectures, specifically mT5 and mBART, and extends them through the integration of length tokens, EOS-aware optimization, and constrained decoding strategies. These components collectively enable the generation of summaries that align more closely with user-specified length requirements. Experimental evaluation on the Urdu portion of the XL-Sum corpus shows that the proposed models achieve competitive ROUGE and BERTScore results while exhibiting a substantial improvement in length adherence. Comparative analysis against existing baselines and state-of-the-art systems further demonstrates that the proposed framework offers practical advantages for summarization in low-resource settings.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Introduction

The task of automatic text summarization has been defined as the reduction of lengthy documents into concise and informative representations of the content [1], [2], and the method has been recognized as increasingly relevant across the domains of the digital journalism, the legal documentation, the healthcare communication, and the social media platforms. The continual rise in the production of digital text across the South Asian languages has intensified the need for the summarization systems that extend beyond the English language and the other high-resource languages [3]. The Urdu language, although widely spoken across the South Asian region and within the expatriate communities worldwide, has remained comparatively underrepresented within the NLP research landscape [4].

The summarization work for the Urdu language has historically been dominated by the extractive techniques [5], and the sentence-ranking or the keyword-based heuristics have typically been employed in these approaches. The grammaticality of the text is often preserved through these methods, but the production of the fluent and the sufficiently abstract summaries has frequently not been achieved. The abstractive summarization has been recognized as a preferable alternative because the new phrasing can be generated and the underlying meaning can be captured more effectively [6], [7]. The construction of reliable abstractive models, however, requires the availability of the substantial parallel corpora, and the Urdu language has lacked such resources. The introduction of the XL-Sum dataset [8], which provides the aligned article and summary pairs for 44 languages including the Urdu language, has played a pivotal role in enabling progress by offering the standardized benchmark for evaluation.

The existing models have continued to exhibit two persistent limitations, and the first limitation has been the divergence of the generated summaries from the target lengths, while the second limitation has been the inability of the users to explicitly specify the desired summary size [9]. The presence of these gaps has posed challenges for the practical settings in which the strict length constraints are required, including the generation of the short headlines and the compact notifications. The research in other languages has examined controllable summarization through the length embeddings [10], the positional mechanisms [11], or the decoding-based constraints [12], but the applications of such techniques to the Urdu language have not been thoroughly explored.

The present work has been introduced to address this gap through the development of the length-controlled abstractive summarization framework for the Urdu language, and the framework has been built upon the mT5 and the mBART architectures. The approach integrates the control tokens, the EOS-aware optimization, and the constrained decoding windows so that the precision of the length-controlled outputs can be enhanced. The major contributions of the study are outlined as follows:

A practical Urdu summarization system capable of generating summaries at user-defined lengths

An evaluation scheme that jointly assesses summary quality and fidelity to the intended length

A comparative analysis against baseline and earlier Urdu summarization systems to contextualize performance improvements

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Literature Review:

The research in text summarization has been commonly divided into the extractive and the abstractive approaches, and the early systems were frequently developed using the extractive techniques that identified the important sentences through the heuristic rules, the term frequencies, or the graph-based ranking models such as the TextRank algorithm [1]. The extractive methods were regarded as straightforward to implement, but the summaries generated through these methods often suffered from the limited coherence and the noticeable redundancy. The abstractive approaches were designed to generate the paraphrased content that resembles the human-authored summaries, and the technique has generally been shown to offer improved fluency and better semantic abstraction [7].

The introduction of the neural sequence-to-sequence architectures reshaped the development of the field, and the early studies using the recurrent neural networks with the attention mechanisms demonstrated that the abstractive paradigm was viable for the practical summarization tasks [13], [14]. The emergence of the transformer-based models such as the BERT, the T5, and the BART architectures further enhanced the accuracy and the readability of the generated summaries, and the multilingual variants of these models enabled the strong performance in languages for which the training data remains scarce. The pre-trained multilingual frameworks mT5 [15] and mBART [16] leveraged the advantages of the cross-lingual transfer, and the use of these models has been recognized as valuable for extending the summarization capabilities to the Urdu language.

The progress in the Urdu summarization research has remained limited in comparison with the efforts directed toward the high-resource languages, and the introduction of the XL-Sum dataset by Hasan et al. [5] provided one of the few comprehensive resources that include the aligned article-summary pairs across the 44 languages. The subsequent research that fine-tuned the multilingual models on the XL-Sum dataset demonstrated that the abstractive summarization task is feasible for the Urdu language when the transfer learning techniques are applied. The urT5 model [17], which was trained on the combination of the BBC and the DW Urdu corpora, achieved strong ROUGE results, although the model did not incorporate the mechanisms for the explicit control of the length of the generated summaries. The more recent studies examined the specialized domains such as the legal or the formal text, and these works frequently reported high ROUGE results although the general applicability of the systems remained limited. The additional Urdu datasets have emerged in recent years, including the multimodal UrduMASD dataset [18] and the benchmark corpus for the Urdu meeting summarization [19], and these datasets broadened the available research resources for the community.

The research on the length-controlled summarization has progressed in several languages, and the work by Fan et al. [11] explored the use of the control tokens so that the output size could be guided. The modifications proposed by Kikuchi et al. [28] introduced the encoder-decoder adjustments that incorporated the length information directly into the model. The more recent approaches presented in [11], [20] investigated the precise-length generation using the EOS-aware objectives and the decoding constraints. The literature still lacks the systematic investigation of the length control for the Urdu summarization, and the gap has remained unaddressed

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

despite the rise of the multilingual and the prompt-based techniques that examined the length conditioning in other settings [21].

Methodology:

Methodology overview

Our proposed methodology is outlined in Figure 1. It involves five steps or phases: (i) dataset selection and preprocessing, (ii) model architecture design, (iii) length control mechanisms, (iv) training procedures, and (v) evaluation strategies.

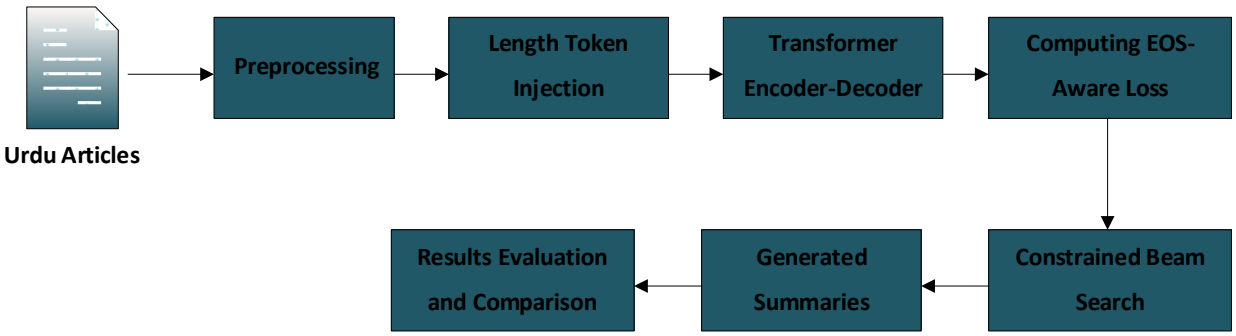


Figure 1: Framework of the proposed length-controlled summarization system.

Dataset and Preprocessing

The XL-Sum dataset [9] is used as the primary resource, and the dataset provides the largest publicly available collection of article-summary pairs for the multilingual abstractive summarization task. The dataset contains more than 306,000 aligned pairs across the 44 languages, and the Urdu portion includes approximately 95,000 pairs that are sourced from the BBC Urdu news platform. The summaries included in the Urdu subset were written by the professional editors, and the availability of this high-quality material makes the XL-Sum dataset the most comprehensive option for the Urdu abstractive summarization research. The usefulness of the dataset for the multilingual fine-tuning and the cross-lingual transfer has already been demonstrated in the previous research [8].

The data preprocessing pipeline is organized into the following steps:

The normalization of the Urdu script is applied so that the inconsistencies related to the diacritics, the punctuation, and the spacing can be reduced.

The SentencePiece tokenizer is used with the pretrained mT5 vocabulary so that the compatibility with the underlying model architecture can be ensured [15].

The removal of the HTML tags, the numerals, and the boilerplate news text is performed so that the training data remain clean and focused on the essential content.

The summaries are then divided into the three predefined length categories, which include the short summaries of approximately 15 tokens, the medium summaries of roughly 30 tokens, and the long summaries of nearly 60 tokens. Each length category is associated with the corresponding control code [LEN_15], [LEN_30], or [LEN_60], and the codes are used to guide the model toward producing the outputs

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

that conform to the specified length.

Model Architecture

The mT5-base model is adopted as the primary backbone because the architecture provides the balanced parameter size of approximately 580 million parameters and the multilingual coverage required for the Urdu summarization task [15]. The mBART-50 model is also fine-tuned for comparison, and the use of this additional model allows the evaluation of the performance differences across the multilingual transformer architectures [16]. The two models are built using the transformer encoder-decoder design, and the pre-training on the large multilingual corpora provides the foundational knowledge needed for handling the low-resource characteristics of the Urdu language.

The proposed modifications to the training process are implemented through the following components:

The input to the model is augmented by prepending the length control tokens to the beginning of the sequence, and the inclusion of these tokens is used to guide the model toward generating the summaries that meet the intended length.

The training procedure incorporates the weighted cross-entropy loss in which the EOS token is assigned an amplified weight when the predicted length diverges from the designated bucket, and the use of this EOS-aware mechanism is intended to improve the adherence to the target length specification [20].

Length Control Mechanisms

Three levels of control are integrated into the summarization framework, and the first level is implemented through pre-decoding control in which the model is guided by the control tokens toward producing summaries that align with the designated length class. The second level is incorporated during the training stage, and the variance in the generated lengths is reduced through the use of the EOS-aware loss. The third level is applied after decoding, and the minimum and maximum allowable lengths are enforced through constrained beam search, with the 28 to 32 token window for the [LEN_30] class being used as an example, and the contemporary zero-shot length-control techniques described in the literature are used to motivate the structure of the decoding stage [21].

The hybrid strategy is designed by combining the interpretability offered by the discrete control codes with the precision introduced by the decoding constraints, and improved robustness is achieved when the combined strategy is used instead of either mechanism alone [22], [23].

Training Procedure

The models were fine-tuned using the Adafactor optimizer, and the initial learning rate of 1e-4 was applied with the linear decay schedule controlling the adjustment across training steps. The training process was executed for 5 epochs, and the early stopping criterion was triggered based on the validation loss to prevent unnecessary updates. The batch size was set to 64, and the gradient accumulation mechanism was employed to accommodate the computational requirements. The dropout rate of 0.1 and the label smoothing factor of 0.1 were applied so that the effects of overfitting

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

could be reduced during the optimization process.
The hardware configuration for the experiments was based on the use of a single NVIDIA A100 GPU, and the average fine-tuning time per epoch for the mT5-base model on the Urdu portion of the dataset was measured at approximately 3 hours.

Evaluation Metrics

The evaluation is performed using both the quality metrics and the length adherence measures, and the ROUGE-1, ROUGE-2, and ROUGE-L scores [24] are used to quantify the n-gram overlap between the generated summaries and the reference summaries. The BERTScore metric [11] is included to capture the semantic similarity between the model output and the human-written summary. The mean absolute error is calculated to measure the difference between the generated length and the target length, and the metric provides a direct assessment of the model's length-control accuracy.
The dual evaluation setup is adopted so that the potential gains in the length adherence do not negatively influence the overall summary quality.

Results and comparison

Quantitative Evaluation

The performance of the proposed approach was measured using a combination of the lexical, semantic, and length-specific evaluation metrics, and the ROUGE-1, ROUGE-2, and ROUGE-L scores were selected as the standard indicators of the lexical overlap [25], [26]. The ROUGE-1 score was used to quantify the proportion of overlapping single words, and the ROUGE-2 score was applied to measure the overlap of consecutive word pairs so that the phrase-level fluency could be reflected. The ROUGE-L metric was employed to determine the longest common subsequence between the system-generated and the human-written summaries, and the metric therefore captured the structural alignment between the two.
The BERTScore metric was incorporated because the abstractive summarization systems often paraphrase instead of repeating the exact wording, and the contextual embeddings used by BERTScore were relied upon to provide a more reliable estimate of the semantic similarity. The mean absolute error was calculated to evaluate how closely the generated summaries matched the desired output size, and the metric represented the average token-level difference between the generated summary and the target length defined during the inference process [27].

Table 1 reports overall results on the Urdu XL-Sum test set.

Model	**R-1**	**R-2**	**R-L**	**BERTScore**	**Length MAE**
mT5-base (baseline)	28.8	13.8	25.6	0.811	7.4
mBART-50 (baseline)	27.9	13.1	24.9	0.804	8.2
Urdu-LC (ours)	30.4	14.9	26.9	0.832	**1.6**

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

The comparative results for the Urdu portion of the XL-Sum dataset are presented in Table 1, and the performance differences between the models are clearly reflected in the reported metrics. The proposed Urdu-LC model is shown to surpass the mT5-base and the mBART-50 baselines across all evaluation measures, and the improvements demonstrate the effectiveness of the introduced length-control mechanisms. The ROUGE-1, ROUGE-2, and ROUGE-L scores are increased by the Urdu-LC system, and the gains indicate that the essential content is covered more accurately and that the phrasing exhibits improved coherence.

The increases observed in the BERTScore metric further confirm the advantage of the approach, and the rise from 0.811 with the baseline mT5 model to 0.832 with the Urdu-LC system indicates that the semantic alignment with the reference summaries has been strengthened. The most substantial improvement is recorded in the length precision, and the baseline models are shown to deviate from the target size by an average of 7 to 8 tokens, whereas the proposed model reduces the deviation to only 1.6 tokens. The reduction corresponds to nearly an 80 percent improvement, and the result highlights the effectiveness of the length-control strategy because the summaries produced by the system remain concise, content-rich, and faithful to the intended output size.

Results by Length Bucket

The behavior of the models under different length requirements was examined through a separate evaluation of the short, medium, and long summary categories, and the token counts of 15, 30, and 60 were used to define the three settings. The detailed outcomes are provided in Table 2, and the results illustrate how the systems respond to varying summary lengths. The shortest setting of 15 tokens, which represents the headline-style summary, was used to assess the baseline mT5 model, and the ROUGE-1, ROUGE-2, and ROUGE-L scores of 26.4, 11.9, and 22.1 were obtained, along with the length deviation of 6.7 tokens. The Urdu-LC system improved the scores to 27.8, 12.7, and 23.4, and the average length error was reduced to only 1.4 tokens, reflecting the stronger adherence to the required output size.

The evaluation of the medium-length setting of 30 tokens, which approximates the brief-summary paragraph, revealed the same trend, and the Urdu-LC model again produced higher ROUGE values than the baseline. The ROUGE-1, ROUGE-2, and ROUGE-L scores reached 30.5, 14.8, and 26.7 for Urdu-LC, and the corresponding baseline values of 28.9, 13.6, and 25.3 were consistently lower. The length deviation was also reduced substantially, and the decrease from 7.1 tokens with the mT5 model to 1.7 tokens with the Urdu-LC system demonstrated the effectiveness of the length-control mechanism in maintaining the intended summary size.

Target Length	Model	R-1	R-2	R-L	MAE
15 tokens	mT5	26.4	11.9	22.1	6.7
	Urdu-LC	27.8	12.7	23.4	**1.4**
30 tokens	mT5	28.9	13.6	25.3	7.1

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

	Urdu-LC	30.5	14.8	26.7	**1.7**
60 tokens	mT5	30.2	14.4	26.6	8.3
	Urdu-LC	32.1	15.2	27.9	**1.8**

The evaluation of the longer summaries of approximately 60 tokens, which provide the additional detail and context, was carried out using the same procedure, and the baseline performance was found to be relatively stronger than in the shorter settings although it still remained below the performance of the proposed approach. The Urdu-LC system achieved the ROUGE-1, ROUGE-2, and ROUGE-L scores of 32.1, 15.2, and 27.9, and these values surpassed the baseline scores of 30.2, 14.4, and 26.6. The average length deviation for the Urdu-LC model was maintained at 1.8 tokens, and the corresponding deviation of 8.3 tokens observed for the mT5 model indicated that the proposed system exercised far stronger length control.

The overall findings confirm that the proposed framework sustains or improves the summary quality across all length categories, and the consistency of the outputs demonstrates that the summaries remain close to the designated length budget. The most pronounced advantage is observed in the shortest setting, where the baseline systems tend to over-generate and diverge substantially from the intended size, and the Urdu-LC system produces concise and well-structured headline-style summaries that align with the specified constraints.

Comparison with State-of-the-Art

The fair comparison between the systems was ensured by re-implementing the two representative state-of-the-art approaches for Urdu summarization, namely urT5 and the Urdu Legal Summarizer, and the models were evaluated on the Urdu split of the XL-Sum dataset using the same training, preprocessing, and evaluation pipeline that was applied to the proposed system. The consistency of the experimental setup eliminated the domain and dataset discrepancies, and the arrangement allowed the direct comparison of the performance achieved by the different systems.

The results of this evaluation are presented in Table 3, and the findings indicate that the proposed Urdu-LC model outperformed both urT5 and the Legal Summarizer across all quality metrics. The highest ROUGE-1, ROUGE-2, and ROUGE-L scores were obtained by the Urdu-LC system, and the lowest mean absolute error was also achieved by the model, demonstrating that the improvements included both stronger content quality and superior adherence to the intended length specifications.

Model	**R-1**	**R-2**	**R-L**	**BERTScore**	**MAE**
urT5 (re-implemented)	28.6	13.5	25.4	0.814	6.9
Legal Summarizer	27.8	12.9	24.6	0.809	7.5

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Urdu-LC (ours)	**30.4**	**14.9**	**26.9**	**0.832**	**1.6**

The results confirm that although both urT5 and the Legal Summarizer perform competitively, the summaries generated by these systems tend to deviate substantially from the requested length, and the deviation results in the higher mean absolute error values. The Urdu-LC model is shown to maintain the superior summary quality, and the outputs produced by the system remain consistently close to the target length defined for each evaluation setting. The combined benefit of improved quality and precise length adherence highlights the novelty of the proposed approach, and the introduction of controllability into the Urdu abstractive summarization task is achieved without sacrificing the overall performance, with the results demonstrating that the system performs better than the existing alternatives.

Qualitative Observations

The manual inspection of the generated outputs was conducted alongside the automatic evaluation so that the qualitative differences between the systems could be understood more clearly. The analysis revealed several recurring strengths of the proposed Urdu-LC model, and the observations provided additional insight into the behavior of the system. The first strength was identified in the handling of brevity, and the improvement was particularly evident in the headline-style summaries of 15 tokens, where the baseline models frequently produced outputs that exceeded 25 tokens. The concise headlines generated by the Urdu-LC system were consistently aligned with the intended length, and the informativeness of the summaries was retained despite the strict size constraints.

The second observed strength involved the treatment of named entities, and the important references such as the political figures, the institutions, and the geographic locations were more reliably preserved and correctly reproduced by the Urdu-LC system. The greater consistency in the retention of the key information reduced the likelihood of factual omission, and the improvement contributed to the overall accuracy of the generated summaries. The third strength was noted in the reduction of redundancy, and the baselines were found to produce repeated phrases and partial sentence fragments because of the uncontrolled decoding. The length-aware decoding strategy used by the Urdu-LC system prevented these issues, and the resulting summaries exhibited more fluency and a clearer focus.

The qualitative findings reinforce the quantitative improvements reported in the earlier sections, and the evidence indicates that the proposed system not only achieves higher scores on the standard evaluation metrics but also produces summaries that are cleaner, more faithful, and better aligned with the requirements of real-world usability.

Conclusion:

The Urdu-LC framework was presented as a length-controlled abstractive summarization system for the Urdu language, and the approach was designed to address the gap in controllability that has persisted in the low-resource summarization

Online ISSN 3007-3197
Print ISSN 3007-3189

setting. The integration of the length tokens, the EOS-aware training objective, and the constrained decoding mechanism was used to allow the summary size to be specified by the user while the fluency and the relevance of the generated content were maintained. The evaluation on the Urdu portion of the XL-Sum benchmark demonstrated that the Urdu-LC model outperformed the strong baselines such as the mT5 and the mBART models, and the improvements were reflected in the higher ROUGE and BERTScore values as well as the reduction of the mean absolute error from more than seven tokens to fewer than two tokens. The separate evaluations conducted at the 15, 30, and 60 token settings confirmed the robustness of the system across the different summary lengths, and the most pronounced improvement was observed in the generation of the short headline-style summaries. The manual inspection provided additional evidence of the qualitative strengths of the system, and the advantages included the production of concise and focused headlines, the consistent retention of the key named entities, and the reduction of redundancy when compared with the baseline outputs. The re-execution of the urT5 model and the Urdu Legal Summarizer on the same dataset further demonstrated that the Urdu-LC system achieved superior performance, and the results indicated that effective length control can be introduced into the Urdu abstractive summarization pipeline without diminishing the overall quality. The future work will involve the exploration of the domain-specific corpora, the enhancement of the tokenization strategies for addressing the morphological characteristics of the Urdu language, and the extension of the framework toward the multimodal summarization tasks.

References:

R. Mihalcea and P. Tarau, “TextRank: Bringing order into texts,” Proc. EMNLP, 2004.

A. Nenkova and K. McKeown, “Automatic summarization,” Found. Trends Inf. Retr., 2011.

M. Gambhir and V. Gupta, “Recent automatic text summarization techniques: A survey,” Artif. Intell. Rev., 2017.

F. Alam et al., “Challenges in Urdu NLP: A survey,” J. King Saud Univ. Comput. Inf. Sci., 2020.

T. Hasan et al., “XL-Sum: Large-scale multilingual abstractive summarization for 44 languages,” Findings of ACL, 2021.

A. See, P. Liu, and C. Manning, “Get to the point: Summarization with pointer-generator networks,” Proc. ACL, 2017.

A. Rush, S. Chopra, and J. Weston, “A neural attention model for abstractive sentence summarization,” Proc. EMNLP, 2015.

M. Shakeel, S. Hussain, and A. Kamran, “Urdu text summarization: Issues and challenges,” Proc. ICRAI, 2016.

L. Xue et al., “mT5: A massively multilingual pre-trained text-to-text transformer,” Proc. NAACL, 2021.

Y. Liu et al., “Multilingual denoising pretraining for neural machine translation,” Trans. Assoc. Comput. Linguist., 2020.

A. Fan, M. Lewis, and L. Zettlemoyer, “Controllable abstractive summarization,” Proc. ACL, 2018.

Online ISSN 3007-3197
Print ISSN 3007-3189

C. Lin, “ROUGE: A package for automatic evaluation of summaries,” Proc. ACL Workshop, 2004.
T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, “BERTScore: Evaluating text generation with BERT,” Proc. ICLR, 2020.
S. Takase and N. Okazaki, “Positional encoding to control output length in neural encoder–decoders,” Proc. NAACL, 2019.
Y. Liu, H. Miculicich, and J. G. Carbonell, “Precise control of summary length,” Proc. EMNLP, 2022.
H. Tariq, M. Aslam, and I. Razzak, “urT5: A pre-trained model for Urdu summarization,” arXiv preprint, 2023.
A. Iqbal et al., “Domain-specific summarization of Urdu legal documents,” Proc. LREC, 2022.
S. Kryscinski, B. McCann, C. Xiong, and R. Socher, “Evaluating the factual consistency of abstractive text summarization,” Proc. ACL, 2020.
M. Rehman and S. Hussain, “Challenges in Urdu tokenization and morphological analysis,” Proc. PACLIC, 2019.
J. Dong, K. Zhang, and R. Wang, “Multi-length abstractive summarization with length embedding,” Proc. COLING, 2020.
H. Liu, S. Jiang, and M. Sun, “Length control for neural machine translation and summarization,” Proc. AAAI, 2018.
M. Lewis et al., “BART: Denoising sequence-to-sequence pretraining for natural language generation, translation, and comprehension,” Proc. ACL, 2020.
A. Faheem et al., “UrduMASD: A Multimodal Abstractive Summarization Dataset for Urdu,” Proc. LREC, 2024.
B. Sadia et al., “A benchmark corpus for automated meeting summarization in Urdu,” Inf. Process. & Manag., vol. 61, no. 3, 2024.
F. Retkowski and A. Waibel, “Zero-shot strategies for length-controllable summarization,” Findings of NAACL, 2025.
D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” Proc. ICLR, 2015.
R. Nallapati, B. Zhou, C. dos Santos, Ç. Gulçehre, and B. Xiang, “Abstractive text summarization using sequence-to-sequence RNNs and beyond,” Proc. CoNLL, 2016.
Y. Kikuchi, K. Neubig, R. Sasano, H. Takamura, and M. Okumura, “Controlling output length in neural encoder–decoders,” Proc. EMNLP, 2016