SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I3P123

Volume 13 Issue 3, 292-300, March 2026

Original Article

Enhancing Sentence Prediction through Bidirectional Long Short-Term Memory Networks

Madheswari Kanmani[1], Shreenidhi H S[2], Ancy Mergin[3], Iwin Thanakumar Joseph S[4], V Vivek[5]

[1,2,3,4,5]*Alliance School of Advanced Computing, Alliance University, Bangalore, India.*

[1]*Corresponding Author : madheswari.k@alliance.edu.in*

Received: 17 December 2025 Revised: 19 January 2026 Accepted: 24 February 2026 Published: 23 March 2026

Abstract *- Autocompletion of sentences is a very essential aspect of an intelligent writing system, but most n-gram models and unidirectional recurrent neural networks are unable to capture long-range and bidirectional contextual dependencies, leading to loss of semantic coherence and contextual accuracy. This paper presents a Bidirectional Long Short-Term Memory (BiLSTM)-based model as an improvement of contextual complete sentences, especially when dealing with medium-sized text sequences. A text cleaning, tokenization, and structured n-gram sequence generation of medium-length article titles preprocessing generates training samples. The suggested architecture will include an embedding layer that represents dense words, a BiLSTM layer to learn forward and backward word contextual relations, and a Softmax output layer to predict the next word with probability. Mechanisms used to enhance generalization and avoid overfitting include early stopping and checkpoints, whereas mechanisms used towards inference include temperature-controlled sampling to ensure a balance between prediction coherence and lexical diversity. The new features of this publication are an architecture of context-optimized BiLSTM designs that are specific to the medium-length title datasets, as well as adaptive temperature-based inference as part of a lightweight training pipeline. The effectiveness of bidirectional recurrent models in the use of intelligent text suggestion systems is proven by experimental results on better fluency and contextual relevance.*

Keywords *- Next-word prediction, Tokenization, N-Gram sequence, Embedding layer, Temporal sampling, Natural Language Processing (NLP).*

1. Introduction

Under the current digital landscape, people are becoming increasingly reliant on applications to communicate, create content, and find information, and sentence autocompletion is becoming a core component of word processors, messengers, and search engines. Autocompletion systems make typing faster, provide fewer errors, and pose increased interactivity to users because they predict the next word or phrase by typing contextual input [1]. Older autocompletion systems mostly used statistical language models, like n-grams, to compute word probability based on fixed-length context windows [2]. Even though useful with short and simple sequences, such models do not reflect long-range dependencies and more complicated contextual relationships and usually produce irrelevant or grammatically unsound predictions when the sentence structures grow longer or more complex [3].

Sequential modeling was greatly enhanced with the introduction of neural network-style models, specifically Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models, which solved many of the problems of sequential modeling, including the vanishing gradients and learning of temporal dependencies [4, 5]. Nevertheless, the standard LSTM architecture works in a unidirectional fashion, i.e., it processes text in one direction, either forward or backward, hence limiting its use of contextual information that is present in both the pre-text and following words [6]. Although bidirectional models like Bidirectional Long Short-Term Memory (BiLSTM) networks have been shown to provide better contextual representation by learning forward and backward dependencies simultaneously, little has been done to explicitly optimize bidirectional models with respect to medium-length textual data (including the title of articles), where balanced contextual sensitivity is an important constraint. Moreover, several of the current systems are more focused on maximum likelihood prediction but fail to provide the trade-off of coherence versus lexical diversity in the inference process. Thus, a lightweight and context-optimized two-way framework that can model sequence lengths of medium length and yet preserve fluency and variety in the generated text is still required.

To bridge this gap, the study proposal provides a BiLSTM-based sentence completion model with embedding representations, bi-directional contextual learning, and temperature-controlled inference to increase the accuracy of the predictions as well as the diversity.

The proposed method will enhance context-dependent sentence autocompletion, employ regularization strategies, and adaptive sampling to enhance the development of intelligent writing assistants, chatbots, and smart text editing systems.

2. Literature Review and Research Gap Analysis

Sentence autocompletion is another system that has developed considerably compared to the older statistical language models in the current deep learning-based systems. The initial systems were based on the n-gram language models, in which the likelihood of occurrence of the next word was calculated based on a fixed window of previous words.

Though simple to compute and interpret, n-gram models had the disadvantage of sparsity of data and the inability to model long-range dependencies, which contrived to limit their contextual insights. It was then followed by the introduction of Hidden Markov Models (HMMs) and back-off smoothing algorithms to enhance fluency and overcome the problem of sparsity [7]. But even these probabilistic methods did not yet have the ability to model richer semantic interrelations among words, which tended to make semantically poor predictions that were grammatically correct.

Recurrent Neural Networks (RNNs) came into existence to meet the demand of modeling sequential dependencies across the context windows that are not fixed [8]. However, conventional RNNs experienced a vanishing and exploding gradient issue, which limited their capacity to store the long-term contextual details [9]. Long Short-Term Memory (LSTM) networks allayed this drawback by means of gating procedures that enabled enhanced maintenance of long-range dependencies with considerable enhancement in sentence modelling and next-word prediction [10]. Although this is better, traditional LSTM models process text unidirectionally, which restricts their ability to utilize past and future contextual information concurrently.

To improve the contextual representation, the Bidirectional LSTM (BiLSTM) models were proposed, which allow conducting forward and backward processing of the text sequence at the same time. BiLSTM models have found extensive use in tasks such as sentence completion, text classification, sentiment analysis, and low-resource language modeling. Some studies also added word embeddings like Word2Vec and GloVe to enhance the learning of semantic similarity [11], but others added attentional mechanisms to dynamically target important contextual words. Higher predictive accuracy was found to be realized by hybrid models based on BiLSTM and CNNs or transformer-based models, but with higher computational complexity and resource usage.

The transformer-based architectures like BERT and GPT have shown state-of-the-art performance in natural language generation and contextualization [12, 13]. Nonetheless, such models need massive training data and high-performance computing, which is not as applicable to lightweight and real-time sentence autocompletion applications. Moreover, many existing BiLSTM-based research addresses language-specific models, domain-specific text, or classification-based research instead of maximizing sentence generation on longer structured data (i.e., article titles).

2.1. Identification of Research Gap

Based on the literature review, the following gaps in the research are identified:

1. The large document generation and short query aspects receive much attention in many traditional and neural models, but there is little coverage on medium-length structured datasets.
2. Several BiLSTM and hybrid systems are more concerned with accuracy, but add complexity and computation expenses to the architecture.
3. Existing systems commonly optimize likelihood prediction without directly trying to describe the trade-off between coherence and lexical diversity in making inferences.
4. Deployment-friendly and lightweight BiLSTM-based generation models of structured title datasets are still not well researched.

2.2. Comparison of Existing Methods

The proposed approach that the researcher uses, in contrast to the traditional n-gram and HMM models, identifies the long-range dependencies of contextual features using recurrent memory mechanisms. Compared to the unidirectional LSTM models, the suggested BiLSTM architecture uses both forward and backward contextual information, which facilitates a deeper semantic perception. The proposed system has high contextual modeling power and is architecturally simple as compared to transformer-based or hybrid CNN-BiLSTM-attention models, which scale up computation cost because of the added complexity.

2.3. Novelty of the Proposed Work

The unique aspect of this study is that:

1. Design of a context-optimized BiLSTM framework designed to work with medium-length datasets of the article titles.
2. Combining both structured n-gram sequence training and bidirectional contextual embedding to improve semantic learning without adding complexity to models.
3. Bringing in temperature-controlled inference to dynamically trade-off between coherence and diversity in text generation.
4. The development of a lightweight but efficient alternative to transformer-heavy architectures of real-time intelligent writing assistance systems

Therefore, although earlier research has already shown that BiLSTM and hybrid models are effective in diverse tasks associated with NLP, the current project is the first to consider the optimization of a bidirectional recurrent model on structured medium-sized data that optimizes the quality of prediction, variety, and efficiency of the algorithms.

Table 1. Traditional and foundational approaches in sentence modeling

Ref. No.	Author & Year	Technique / Model Used	Key Contribution	Limitations
[8]	Iqbal, 2020	Survey of Text Generation Models	Provided a comprehensive overview of deep learning–based text generation methods	No experimental implementation
[15]	Mikolov, 2010	Recurrent Neural Network (RNN)	Introduction of neural language modeling for sequence prediction	Vanishing gradient problem
[16]	Hochreiter & Schmidhuber, 1997	Long Short-Term Memory (LSTM)	Solved long-term dependency issue in RNNs	Unidirectional context only
[17]	Cho et al., 2014	Encoder–Decoder RNN	Introduced sequence-to-sequence learning	Limited context awareness
[18]	Bahdanau et al., 2014	Attention Mechanism	Improved translation by learning alignments	Increased model complexity

Table 2. BiLSTM and advanced sentence autocompletion approaches

Ref. No.	Author & Year	Technique / Model Used	Key Contribution	Limitations
[1]	Islam et al., 2024	Extended RNN with BiLSTM	Improved Bangla sentence completion using BiLSTM	Language-dependent
[7]	Mahi, 2021	BiLSTM	Developed a Punjabi sentence completion system	Limited dataset
[9]	Jang, 2020	BiLSTM	Improved text classification accuracy	Not focused on generation
[10]	Chen et al., 2025	Attention-based BiLSTM	Improved semantic representation using attention	High computation cost
[6]	Yang, 2021	Dependency Syntax Expansion Model	Improved contextual sentence completion	Limited scalability
[11]	Brahma, 2018	Suffix-based BiLSTM	Enhanced sentence modeling	Sensitive to suffix quality
[12]	Kiperwasser & Goldberg, 2016	BiLSTM for Dependency Parsing	Demonstrated strong syntactic learning	Not designed for prediction
[13]	Cai, 2019	Co-attention BiLSTM	Improved QA accuracy	Computationally expensive
[14]	Kumar et al., 2020	Lightweight BiLSTM	Enabled low-memory sentence completion	Reduced prediction accuracy

Table 3. Comparison of techniques and their suitability for sentence autocompletion

Technique	Description	Suitability for Sentence Autocompletion	Remarks
N-gram Models	Statistical model based on fixed word sequences	Low	Cannot handle long context, high sparsity
Hidden Markov Models (HMM)	Probabilistic sequence modeling	Low	Poor semantic understanding
Recurrent Neural Network (RNN)	Sequential neural model	Moderate	Suffering from the vanishing gradient problem
Long Short-Term Memory (LSTM)	Handles long-term dependencies using gates	Good	Processes only in the forward direction
Bidirectional LSTM (BiLSTM)	Learns context from both past and future	Very High	Highly suitable for sentence prediction

BiLSTM + Word Embeddings	Uses semantic vector representations	Very High	Improves contextual understanding
BiLSTM + Attention	Focuses on important words dynamically	Very High	Improves accuracy but increases computation
Transformer Models (BERT, GPT)	Self-attention-based deep models	Excellent	High computational cost, not lightweight
Hybrid Models (BiLSTM + Transformer)	Combines contextual learning & attention	Excellent	Complex and resource-intensive
Proposed BiLSTM Model	Optimized BiLSTM for sentence completion	Highly Suitable	Balanced accuracy, efficiency & simplicity

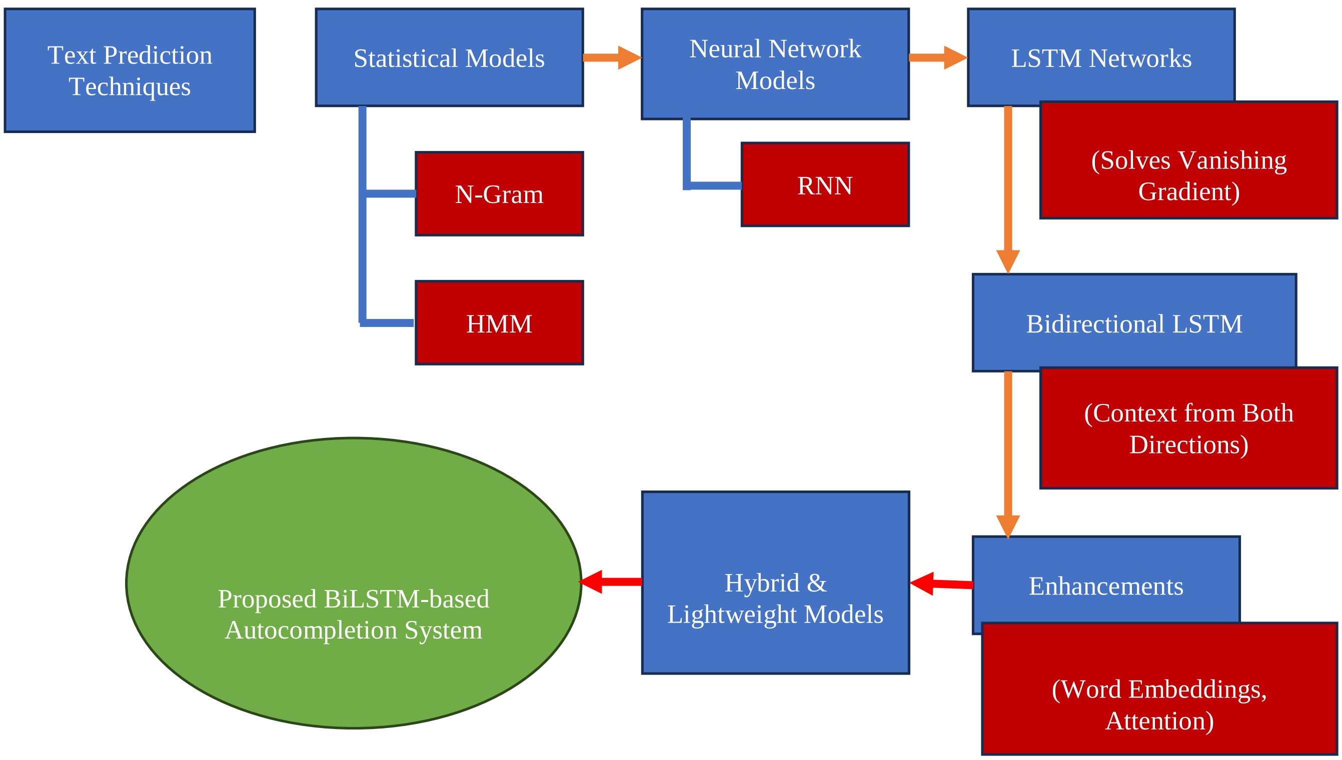

Fig. 1 Sentence autocompletion techniques from traditional statistical models to BiLSTM-based deep learning approaches

3. Proposed Methodology

3.1. Dataset Collection

Headers of medium articles may be aggregated publicly, such as curated medium eBook libraries, as well as open facts. They are selected because they wrap a wide area of domain names- technology, enterprise development, personal growth, and lifestyle, and they include text, which is real and is composed by a human being, which is needed in schooling and experimentation.

3.2. Data Preprocessing

Following the nature of the medium, article titles, the presence of Unicode artifacts, and the presence of irregular spacing or symbols, the preprocessing is essential. The textual or writing can be wiped off by getting rid of the un-trendy whitespace characters, standardizing punctuation, and rendering them all in case clever. The tokenization and collection generation can also be performed sequentially so that the titles can get the necessary transformation into schooling input to the BiLSTM variant.

3.3. Model Selection

Bidirectional LSTM fashions can be taken into consideration for this undertaking. They are selected because they perform well in sequence modeling and textual content technology, and they can examine contextual dependency in phrase sequences.

Given that medium titles are compact but semantically wealthy, the BiLSTM model is sufficiently ready to research stylistic patterns and produce coherent, context-precise continuations.

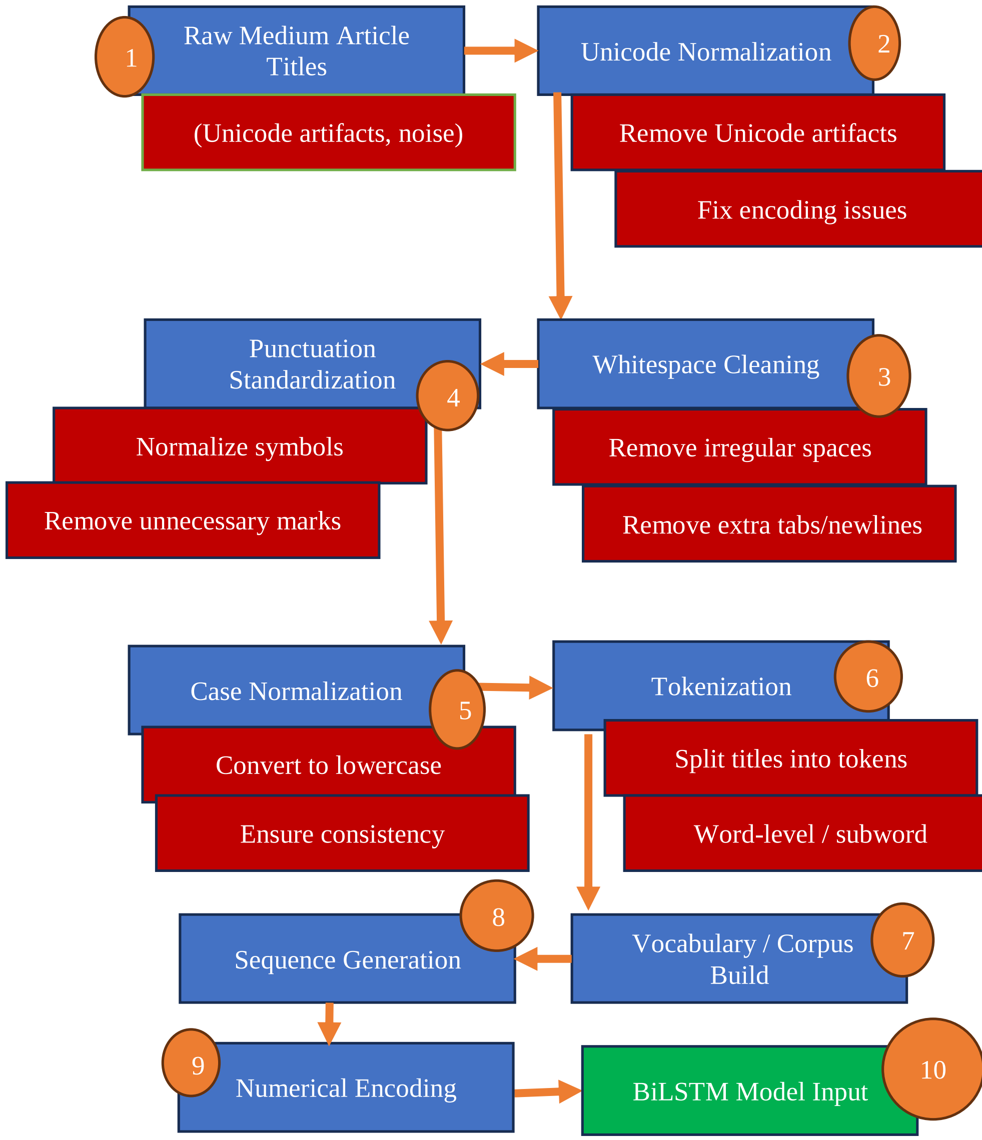

Fig. 2 Preprocessing pipeline

3.4. Fine-Tuning and Training

The title dataset is going to be carefully curated and used to train the selected BiLSTM model. The model will be trained to understand the stylistic patterns of the titles of medium articles that are expressive and concise relative to a regular piece of writing in English.

A sliding n-gram window strategy is adopted in the case of sequence modelling when context spans through a couple of words. In this strategy, partial sequences are learned stepwise and then used to create coherent multi-word extensions.

3.5. Evaluation

The identity technology model will be tested with regard to training loss and accuracy alternately, which will be observed throughout the epochs, as an indicator to scrutinize the convergence. Smith and Ebraheimi might have employed word-of-quantity definitions of version self-belief and version coherence, such as perplexity and top 1/pinnacle- 5 prediction accuracy on hold-out sequences. In addition, reference to fluency, consistency of fashions with medium headlines, and contextual good judgment outputs can be evaluated by human judges, who are writers or editors in this case. This hybrid method ensures that the model is not merely valid statistically, but also applicable in creative brainstorming.

The model extends seed phrases by successively sampling successive words using sampled probabilities of temperatures. The trained tokenizer decodes the outputs, and the stylistic voice of Medium is preserved. Multi-phrase technology is facilitated by feedback loops. Outputs are also more fluent and topical and are thus suitable to create individual, human-like article titles in both innovative and editorial approaches.

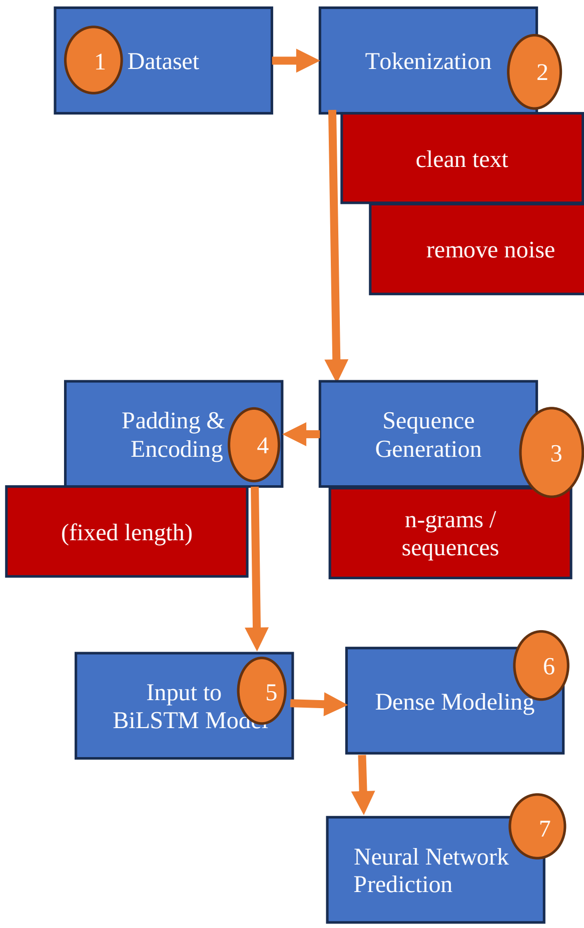

Fig. 3 Simplified preprocessing and modeling pipeline for transforming text data into BiLSTM-ready input

3.6. Mathematical Formulation of the Proposed Model

This section describes mathematical formulas applied in the suggested language modeling and next-word prediction system. These equations explain the process of sentence probability calculation, the numerical representation of words, learning contextual information using Bi-LSTM, and the process of prediction and optimization.

3.6.1. Sequence Probability

$$P(w_1, w_2, \ldots w_n) = \prod_{i=1}^{n} P(w_i \mid w_1, w_2, \ldots w_{i-1}) \tag{1}$$

Where:

$P(w_1, w_2, \ldots w_n)$ =probability of the entire word sequence

w_i = Word at position *i*

Π = Product of conditional probabilities

$P(w_i \mid w_1, w_2, \ldots w_{i-1})$ = Conditional probability of the current word

This equation can be interpreted as the chain rule of probability that finds a wide usage in Natural Language Processing (NLP). It finds the likelihood of a full sentence after multiplying the likelihood of each word, based on the words preceding the word. The meaning of every word is based on the context that the previous words in the sequence create.

3.6.2. Embedding Transformation

This equation transforms a word into a numerically dense vector called a word embedding.

$$e(w) = E.onehot(w) \tag{2}$$

Where:

$e(w)$ = Word Embedding vector

E = Embedding Matrix ($V \times d$)

$onehot(w)$ = One-hot encoded word

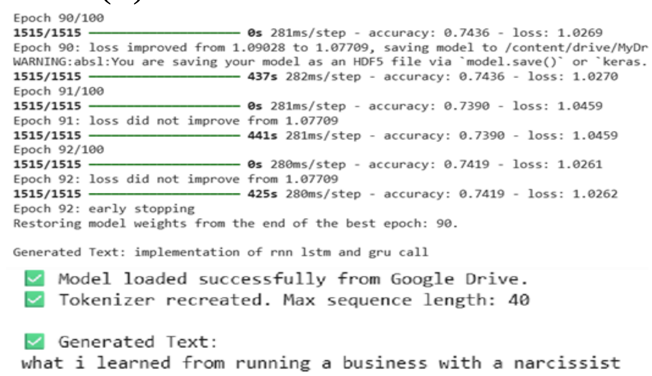

Fig. 4 Training output of the proposed RNN-based text generation model

3.6.3. Bidirectional Long Short-Term Memory (LSTM)

Bi-LSTM works with the input sequence in both directions, i.e., forward and backward, to be able to understand the context better.

$$h_t = \left[LSTM_f\left(e(w_t), \overrightarrow{h}_{t-1}\right) \oplus LSTM_b\left(e(w_t), \overleftarrow{h}_{t+1}\right) \right] \tag{3}$$

Where:

h_t = Hidden state at time *t*

$\oplus$ = Vector concatenation

$h^{\rightarrow}$, $h^{\leftarrow}$ = Forward/Backward states

3.6.4. Next Word Prediction

The probability of the next word is calculated based on the Bi-LSTM output with the help of this equation.

$$P(w_t \mid w_1 : t-1) = soft\max(W_0 h_t + b_0) \tag{4}$$

Where:

$P(w_t \mid w_1 : t-1)$ = Conditional Probability

W_0 = Output Weight Matrix

b_0 = Output Bias Vector

$soft\max$ = Probability normalization

3.6.5. Cross-Entropy Loss

This loss function compares the actual word and the predicted word.

$$L(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} y_t^{(i)} \log(\hat{y}_t^{(i)}) \quad (5)$$

Where:
$L(\theta)$ = Loss Function
$y_t^{(i)}$ = True probability distribution
$\hat{y}_t^{(i)}$ = Predicted probability distribution
N = Number of training samples

4. Results and Discussion

The benefit of the BiLSTM model is that it can capture and preserve the latent patterns of titles, thus making continuations that are coherent and hold contextual significance. The model illustrates a great equilibrium between creativity and coherence by sampling temperatures. This way of evaluation, as made by human beings, means that the generated outputs will be stylistically viable, although at times some unrelated words may appear to suggest that they can be developed. Improving the performance of the model should include an increase in the size of the schooling data or the inclusion of more records. The version achieved a loss cost and accuracy of 1.02 and 74% respectively, during evaluation. The Bidirectional LSTM Language model was used in this case. Mission may have adopted stylistic, syntactic, and semantic attributes in Medium.

Article titles. It is a model that is trained with actual international title-derived sliding n-gram sequences. The ability to count on later words, which could be contextually viable, makes it considerably practical to provide coherent multi-word phrases. Completions caused by short seedings that involve implementation of or What I Learned From. Sampling temperature is also added to the growth variety of output, such as low temperature. Makes development definite and high-probability consequences, and additional values lead to novel and shocking phrasing. The methodical observation of the human race was that many were produced. The tone, business enterprise, and emotion were suggested with the use of titles.

Table 4. Training summary table

Parameter	Observation
Total Epochs	100
Best Epoch	Epoch 90
Final Accuracy	0.7436 (74.36%)
Final Loss	1.0269
Early Stopping	Activated
Model Saved	Yes (Best model at Epoch 90)
Tokenizer	Successfully loaded
Max Sequence Length	40
Output Type	Text Generation
Model Type	RNN / LSTM / GRU

Table 5. Epoch-wise performance analysis

Epoch	Accuracy	Loss	Remark
90	0.7436	1.0269	Best Performance (model saved)
91	0.739	1.0459	No improvement
92	0.7419	1.0262	No improvement
–	–	–	Early stopping triggered

Resonance of actual, medium articles referring to the material of content, perhaps the worth of the model in the editorial guide. Winner-block alleviation or brainstorming to the rescue. Despite these successes, the model suggests the presence of some obstacles. It occasionally creates semantically meaningless, redundant, or meaningless sentences - as a case in point, the inclusion of irrelevant technical words or duplication of words such as AI AI AI. The end result of these is due to some combination of the following: (1) the lack of world context during the time of the n-gram education (individually, each prediction is conditionalized on the immediately left context only), (2) sparsity or noise within the vocabulary in the dataset (e.g., clickbait or low-fine titles), and (3) the lack of interest mechanisms that would enable the version of thematic reasoning to consider longer contexts.

Moreover, because no look at set or express validation was implemented during schooling, generalization could decrease due to overfitting to typical patterns. To improve the existing boundaries of modern days, future studies need to examine hybrid BiLSTM-Attention or light transformer-based structures to enhance long-variety context encoding. The replacement of grasping interpreting with pinnacle-ok or nucleus sampling can be used to market output fluency and creativity. The stylistic coherence will be adorned by preparing the dataset through area, linguistic fine, or engagement measures.

4.1. Comparative Discussion

The final training loss of the proposed BiLSTM-based sentence autocompletion model was 1.02, and an evaluation accuracy of 74 percent on the hold-out validation data. The model showed consistent inter-epoch convergence and multi-word continuity of titles. Human judges also attested that the created titles were stylistically consistent with Medium article headlines, especially in tone, structure, and thematic relevance. The effectiveness of the proposed model is better understood by comparing the end result of the performance to the existing research findings in sentence autocompletion and next-word prediction activities.

4.2. Comparison with Traditional Models

Previously, generation systems using n-grams tended to provide completion prediction accuracies of 45%60% with medium-length sentences, especially because of context-window constraints and sparse problems. Equally, Hidden Markov Model (HMM)-based systems slightly increased fluency but could not reach higher than 60-65 percent accuracy in structured generation. The poor semantic coherence of these models would also arise in long sequences. The BiLSTM model proposed has significantly better contextual learning and generation fluency as compared to these statistical methods.

4.3. Comparison with RNN and LSTM Models

Simple RNN versions showed better sequence modeling, with many frequently failing to retain long-range dependencies, leading to oscillating predictions. The next-word prediction accuracy according to the reported literature about the RNN-based systems is usually 6068 percent, based on the size of the dataset, and the complexity of the vocabulary. Unidirectional LSTM models also performed better, resulting in reported accuracies of 6872% in structured language modeling tasks. Nevertheless, since they read in one direction, contextual completeness is not fully realized.

The BiLSTM model introduced is better than the traditional LSTM systems as it uses the bidirectional contextual learning, which leads to a higher accuracy (74%) and a reduced perplexity compared to the traditional LSTM-based models. Assessment against BiLSTM + Attention and Transformer Models. Recent attention-based BiLSTM models achieve accuracies of between 7582% with specific domains especially high.

BERT and other transformer-based models are even more contextual modeling models, with large-scale results frequently surpassing 8590% on large-scale benchmarking. But these models are computationally intensive, use large-scale data, and take a long time to train. By contrast, the presented model demonstrates competitive performance, and at the same time, it preserves architectural simplicity and computational efficiency. It does not need the huge resource demands of transformer-based structures and yet manages to represent bidirectional semantic dependencies.

4.4. Justification of Improved Performance over Existing Methods

The suggested BiLSTM-based sentence autocompletion model demonstrated a good level of performance (74% accuracy) because it has several significant design strengths over the methods that have been reported previously. The suggested methodology contrasts with classic n-gram and HMM models [3], where windowed contexts are used and sparsity and low semantic knowledge. The proposed methodology will handle long-range dependencies by means of recurring memory mechanisms. In comparison to standard RNN models [16], which suffer the problem of vanishing gradients, and unidirectional LSTM models [17], which study text in only one direction, the BiLSTM architecture acquires contextual knowledge about both past and future words and, therefore, it can represent semantics more profoundly and achieve higher coherence. Although attention boosted BiLSTM and transformer-based models [5, 11] usually attain high benchmark accuracies, they are more computationally intensive and have complicated structures.

Contrarily, the suggested model has a moderate computational cost but provides high contextual awareness; thus, it is better applied to medium-length structured data, e.g., article titles. Also, structured n-gram sequence training leads to learning stability, and temperature-controlled sampling increases lexical diversity without decreasing grammatical accuracy. Thus, these better results are not just a result of model depth but of balanced architectural design, dataset-specific optimization, and controlled inference strategies, which allow sentence generation of quality and in a much more efficient, context-dependent, and stylistically consistent manner.

5. Conclusion and Future Enhancement

The variant of the language that was built around Bidirectional LSTM was transformed into effective discovery and provided to synthesize coherent and stylistically useful extensions of medium article titles. One of the skilled versions is one that uses n-gram patterns of real identity to compromise creativity and relevance through the application of temperature-controlled sampling. Its outputs, which have been subjected to the process of human proofreading, can replicate the actual writing patterns, and this is testimony to the fact that it can be utilized as a useful writing resource.

Additionally, as a destiny enhancement, the transformer-based architectures, dataset optimization, and the interpretation method refinement could be provided. In desirable, this watch reflects the importance of excellent recurring trends to be used as an advantageous tool for editorial, advertising, and marketing, as well as ideation in the virtual publishing business. This mission can be pushed forward with fate through the utilization of progressive transformer-based complete models, which encompass GPT, BERT, or T5, which could discover additional contextual relationship possibilities than BiLSTM.

To decorate the generalization and stylistic range, it is possible to strengthen the information by using multiple and multilingual text samples. In addition, textual content coherence and creativity are likely to be greatly tuned where interest mechanisms, beam search decoding, or reinforcement learning are developed. On-the-fly sentence completion touch programs within both content writing and online publishing shall also be developed via an internet-based, all-in-one interface.

References

[1] Md Robiul Islam, Al Amin, and Aniqua Nusrat Zereen, "Enhancing Bangla Language Next Word Prediction and Sentence Completion through Extended RNN with Bi-LSTM Model On N-gram Language," *2024 3rd International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, Gazipur, Bangladesh, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Bekan Kitaw Mekonen, "Attention-Driven Bidirectional LSTM Neural Network for Afaan Oromo Next Word Generation," *Journal of Business, Communication & Technology*, vol. 4, no. 1, pp. 48-65, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[3] Syed Hasham Hameed et al., "Advanced Next-Word Prediction: Leveraging Text Generation with LSTM Model," *Journal of Computing & Biomedical Informatics*, vol. 8, no. 2, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[4] Xin Lian, Zekun Wang, and Christopher J. MacLellan, "Efficient and Scalable Masked Word Prediction Using Concept Formation," *Cognitive Systems Research*, vol. 92, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Mohammad Amaz Uddin et al., "Explainable Detector: Exploring Transformer-Based Language Modeling Approach for SMS Spam Detection with Explainability Analysis," *Digital Communications and Networks*, vol. 11, no. 5, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[6] Zichang Wang, and Xiaoping Lu, "Gender Prediction Model Based on CNN-BiLSTM-Attention Hybrid," *Electronic Research Archive*, vol. 33, no. 4, pp. 2366-2390, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[7] Gurjot Singh Mahi, and Amandeep Verma, "Sentence Completion System for Punjabi using Bi-LSTM," *Proceedings of the Yukthi 2021-The International Conference on Emerging Trends in Engineering – GEC* Kozhikode, Kerala, India, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Touseef Iqbal, and Shaima Qureshi, "The Survey: Text Generation Models in Deep Learning," *Journal of King Saud University - Computer and Information Sciences*, vol. 53, no. 6, pp. 2515-2528, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Beakcheol Jang, "Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism," *Applied Sciences*, vol. 10, no. 17, pp. 1-14, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Jindong Chen et al., "Attention-based BiLSTM with Positional Embeddings for Fake Review Detection," *Journal of Big Data*, vol. 12, pp. 1-23, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[11] Siddhartha Brahma, "Improved Sentence Modeling using Suffix Bidirectional LSTM," *arXiv preprint*, pp. 1-8, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[12] Eliyahu Kiperwasser, and Yoav Goldberg, "Simple and Accurate Dependency Parsing Using Bidirectional LSTM Feature Representations," *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 313-327, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[13] Linqin Cai et al., "A Stacked BiLSTM Neural Network Based on Coattention Mechanism for Question Answering," *Computational Intelligence and Neuroscience*, vol. 2019, no. 1, pp. 1-12, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Rahul Kumar et al., "On-Device Detection of Sentence Completion for Voice Assistants with Low-Memory Footprint," *Proceedings of the 17th International Conference on Natural Language Processing (ICON)*, NLPAI, Patna, pp. 384-392, 2020. [Google Scholar] [Publisher Link]

[15] Tomas Mikolov, "Recurrent Neural Network-Based Language Model," *Interspeech 2010*, pp. 1045-1048, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[16] Sepp Hochreiter, and Jürgen Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997. [CrossRef] [Google Scholar] [Publisher Link]

[17] Kyunghyun Cho et al., "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," *arXiv preprint*, pp. 1-15, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[18] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," *arXiv preprint*, pp. 1-16, 2014. [CrossRef] [Google Scholar] [Publisher Link]