International Research Journal on Advanced Engineering Hub
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 80 - 88
www.irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0011

A Comprehensive Review on Advances in Large Language Models for Topic Modeling

P. Haritha[1], P. Shanmugavadivu[2]
[1,2]Department of Computer Science and Applications, The Gandhigram Rural Institute (Deemed to be University), Gandhigram, Tamil Nadu, India.
EMail Id: *7haricsa@gmail.com[1], psvadivu@ruraluniv.ac.in[2]*

Abstract
Topic Modeling (TM) is an unsupervised machine learning (ML) approach that extracts latent patterns/ topics/ words from a given large text dataset, to accomplish semantic text summarization. TM plays a vital role in extracting topics from the text datasets pertaining to various domains including bioinformatics, economics, healthcare and social media data analysis. This literature review has analyzed the ML-based approaches devised for topic modeling. This article specifically provides deeper insights into the chosen baseline and advanced approaches using Large Language Model (LLM). The article highlights the potential of LLM integration with Bidirectional Encoder Representations from Transformers (BERT) guided by the clustering-based approaches, for semantic clustering of topics and associated documents. A thorough investigation on those methods for TM is performed and documented, as a source of reference. In addition, the compiled details on the evaluation metrics for TM, presented in this article may serve as ready-reference for those researchers interested in TM. Further, this article highlights the limitations of the reviewed articles which will prompt the researcher to develop novel methodologies and metrics for efficient TM.
Keywords: *LLM, Topic Modeling, BERT, Text Summarization, ChatGPTAgriculture.*

1. Introduction
Topic modeling is an inevitable element of Natural Language Processing (NLP) while performing tasks like text summarization, sentiment analysis on public domain text data. Topic modeling deals with the extraction of latent topics from the given input text data. There are several baseline approaches available to achieve topic modeling, viz, Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), Combined Topic Models (CTM), LDA with Products of Experts (PLDA), Embedded Topic Model (ETM), Neural Variational Document Model (NVDM), Neural Topic Model (NTM), Neural Sinkhorn Topic Model (NSTM) and Contrastive Learning Neural Topic Model (CLNTM). These models primarily use the principles of matrix factorization, probability and neural networking. Among these contemporary models, Bidirectional Encoder Representations from Transformers (BERT) and their embedding approach models such as Sentence Transformers, to achieve semantically rich results, compared to the traditional approaches. Due to the advancement and shift towards LLM, the integration of BERT provides efficient extraction of topics with relevant short summaries. This article is designed to provide a comprehensive literature review and analysis on the conventional and advanced approaches using LLM models for efficient topic modeling [1,2,3].

2. Motivation
The authors [4] implemented a prominent LDA and optimized LDA to deal with topic modeling in an unsupervised learning environment. However, the performance is remarkable, the conventional

International Research Journal on Advanced Engineering Hub
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 80 - 88
www.irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0011

methods have a poor understanding of medical datasets and also yield poor semantic outcomes. The traditional approaches for topic modeling face challenges in terms of data representation in numerical formats using cosine similarity-based matrices. In addition, the baseline approaches also failed to address the retrieval of sub topics rather than the single topics from the documents on efficient TM. To address all the drawbacks the study made detailed literature review on the LLM models integrated with BERT leading to clustering.

3. Research Questions

This review is designed to address the following research questions on TM approaches and the evaluation metrics.

- Are conventional approaches for TM are efficient?
- Which conventional approach aligns with LLM for TM semantically?
- What are the prompting techniques of LLM? What are all the prominent sentence transformers chosen to integrate with LLM?
- Are the clustering approaches ensure model interpretability and stability?
- What are the prominent datasets used for performing TM?
- What are the metrics used for the efficient evaluation of TM?

4. Review of Literature

The researchers [5] proposed a method called LimTopic that analyzes the limitation sections of various Association for Computational Linguistics (ACL) scientific research papers, such as Short Papers, Long papers and finding papers, to extract topics and the associated summaries. The extraction of topics and summaries were done using three approaches: LLM prompt, fine- tuned LLM and LLM integrated with BERT. LLM integrated with BERT provided better results than other approaches. Furthermore, LLM is implemented using Mistral, Llama and ChaGPT versions GPT 3.5 and GPT 4.0. Out of all, GPT 4.0 is focused on providing better coherence score. Initially, BERT extracted topic sentences associated with each topic and LLM, with GPT version 4.0 provided better topic summaries of each topic. The semantics of generated text summaries is evaluated using ChatGPT 4 and Claude 3.5. The article faced limitations due to the restrictions of the GPU which led to experimenting with the dataset on Llama 7b and 13b models. LimTopic suffered due to bias, as a consequence of inherent demographic and language biases in LLM. Moreover, the model also faces bias over prominent topics and sentences. The article [6] presents a method named Topic Evolution That You See (TETYS) to analyze the scientific articles related to Sustainable Development Goals (SDG) under five topics. The automated pipeline utilized BERT topic modeling enhanced with LLM. The article incorporates the hyperparameters optimization for efficient clustering and has launched a web portal (http://gmql.eu/tetys/). TETYS is evaluated on datasets for which abstracts, ranging from 25000 to 35000 per topic, are extracted from Elsevier Scopus using an Application Programming Interface (API) key. The Density-Based Clustering Validation (DBCV) score of 0.76 represents the highest score, providing cohesiveness and effective topic extraction. This technique requires CPU computation and could not attain global hyperparameter optimization. The researchers [7] proposed TopicGPT, which aims to address the issues in conventional Topic Models, such as LDA, that provide a bag of words, allowing researchers minimal control over formatting and extracting latent topics. TopicGPT provides a prompt-based LLM framework with two stages: topic generation and topic assignment. The Topic generation stage extracts topics and removes duplicates as well as infrequent topics. The output of the topic generation stage is fed as the input to the topic assignment stage, which assigns topics along with short summaries. TopicGPT is evaluated on 14K articles from Wikipedia and 32k bill summaries against the approaches, namely LDA, seeded LDA and BERT Topic. The TopicGPT achieves 0.74% harmonic

International Research Journal on Advanced Engineering Hub
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 80 - 88
www.irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0011

mean purity compared to the existing methodologies, which is 0.64%. The other metrics, Adjusted Rand Index and Normalized Mutual Infromation, listed in Table I, deal with the validation of clusters that determine the values nearer to 0 denote the randomness in cluster assignments and 1 for consistent assignments. In article [8] the three topic modeling techniques LDA, BERT and RoBERT are implemented on reviews collected from health applications MyChart, Replica and Teledoc. Initially, the datasets were preprocessed by removing stop words, special symbols and lowercase conversion of letters. The preprocessed reviews are transformed into a dictionary with the inclusion of unique words with relevant indexes. The reviews are represented as bigrams and identifiers to feed them into LDA, and the numbers of topics are specified to optimize the model. The model is evaluated with the coherence score that is best suited for the applications with user's interaction. LDA provides the topic distribution as the outcome. Robustly Optimized BERT Pretraining Approach (RoBERTa) is proposed to obtain the numeric values from the processed reviews and the hyperparameters of BERT-base embedding were fine-tuned. Next to the embedding phase, Uniform Manifold Approximation and Projection (UMAP) is used for dimensionality reduction, providing context-based features that are fed to the clustering phase uses HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) to cluster topics along with Silhouette score reduces cluster's size. Finally, Term Frequency is used generate the topics. Thus, the article suggested that RoBERT representing LLM has outperformed LDA. The researchers [9] evaluated the LLM with GPT-4o mini and Gemini 2.0 Flash on datasets collected from Q&A forums of Tor Network. The forums relate the Dark Web on the varieties of Deepweb Q&A, Deep Answers and Respostas Ocultas that deals with hidden answers. All these datasets are properly crawled and compiled as datasets. Single-word topics are extracted from the datasets through the combined methodology of LLM with GPT and Gemini against TF-IDF (scikit-learn) and LDA (Genism). The similarity score of GPT and Gemini are 0.724 which was greater than that of Genism with 0.477. In addition, Levenshtein Distance (LD) used to evaluate the quality of clustering in TM. Lower the LD results are, better the results of TM are. The obtained LD with mentioned dataseets provides 13.49 for GPT and Gemini and 18.22 for scikit-learn and 17.97 for the combination of scikit-learn and genism respectively. In article [10] an integrated LLM framework with encoder-decoder is proposed to extract highly clustered topics with relevant sentences. Initially, Fine-tuned LAnguage Net-T5 (FLAN-T5) is added with CNN encoder and decoder achieves substantial dimensionality reduction on the fed input data. The processed data is further fed into two autoencoders. The first autoencoder deals with extracting topic distributions and corresponding bag of words; the topic distributions are fed as input to the second autoencoder that extracts embedding from the inputs. Thus the proposed DeTiME (Diffusion-Enhanced Topic Modeling using Encoder-Decoder based LLM) model is designed by incorporating all those mentioned components. The authors also suggested the use of models Flan-T5-XL and LORA the Pretraining with Fine-Tuning (PEFT) as a future enhancement. The perceived limitations of the DeTiME are that its performance was not compared with Bidirectional and Auto-Regressive Transformers (BART) and Low-Rank Adaptation (LORA). In addition, the FLAN-T5 model needs to be scaled and lack of fine-tuning the dimensions of CNN and other architectures, such as Long Short-Term Memory (LSTM), Multi-Layer Perceptron (MLP) needs to be used for efficient performance. DeTiMe was evaluated using metrics namely Top-Purity, Top-NMI, Km Purity, Km NMI, Diversity and Coherence range from 0 to 1, with higher values indicating better topic extraction. In article [11] LLM is reported to outperform on the data extracted using Web of Science and LexisNexis than the conventional topic modeling approaches, namely

International Research Journal on Advanced Engineering Hub
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 80 - 88
www.irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0011

LDA, Non-negative Matrix Factorization (NMF) and Combined Topic Models (CTM). The hyperparameters of BERT integrated with Sentence Transformers are tuned with multi-qa-miniLM-L6-cosv1 for better topic extraction. In addition, the hyperparameters were tuned in the UMAP and HDBSCAN phases. Topic Diversity (TD), Coherence Score, Normalized Pointwise Mutual Information (NPMI) and Point Wise Mutual Information (PMI) were the metrics used to evaluate the performance of the proposed technique. The researchers [12] used the Reddit Mental Health Dataset for the evaluation of thire proposed LLM model against the conventional topic modeling approaches LDA, BERT, Mistral 7B instruct and GPT-3.5. Among all those approaches LLM with GPT-3.5 provided the highest accuracy with 20 billion parameters whereas Mistral 7B provided accuracy close to GPT-3.5 with 7.3 billion parameters. Hence, the proposed LLM is observed to be computationally simple and faster. But the conventional approaches LDA and BERT failed to understand the medical -based topics. The proposed method evaluated with Purity, Adjusted Rand Index (ARI) and Model Interpretability and Stability (MIS) determines the performance of the model; The higher the value is , the more efficient the model is. The article [13] has analyzed the implementation of traditional topic modeling approaches namely LDA, LDA with Products of Experts (PLDA), Embedded Topic Model (ETM), Neural Variational Document Model (NVDM), Neural Topic Model (NTM), Neural Sinkhorn Topic Model (NSTM) and Contrastive Learning Neural Topic Model (CLNTM) on 20 Newsgroup dataset that contains long documents and DBpedia having short documents. All the conventional models are compared with the proposed Word Agreement with Language Models (WALM) that focuses on qualitative representations of topics with higher semantic order. The metrics considered for evaluation were Semantic Overlap (Soverlap), Synset-based Similarity (Ssynset), Semantic or Ontology Alignment (Soa), Some overall topic metric or coherence (Sot). The specified metrics provided values were given in Table 1, which determines the semantic quality. The researchers have mentioned the limitations of the proposed WALM model can be biased with usage of LLM and was not evaluated using traditional metrics such as NPMI, coherence and diversity. The researchers [14] have attempted to integrate BERT with LLM, which was experimented on Three Twitter and X datasets. The proposed framework was designed to have three phases namely Topic Generation, Topic Reduction and Evaluation. The topic generation phase uses BERT, the extracted topics are fed to the Topic Reduction phase that uses four language models GPT-4o mini, Llama-3 8B, Gemma-3-12b, Qwen3-30b-a3b. The evaluation phase uses NPMI and TD as metrics. BERT configuration is embedded with all-miniLM-L6-v2 pretrained model, used UMAP for dimensionality reduction, HDBSCAN for clustering as well as CountVectorizer and TF-IDF for data representation. The proposed LLM with four language models reduces the computation overhead. In article [15], the implementation of LLM with short texts is discussed, for which the conventional topic models perform low in extracting the topics due to word co-occurrences. The researcher proposed a framework that uses LLM to extend the short texts before feeding them into topic models. Furthermore, prefix-tuning of a smaller language model coupled with a Variational Autoencoder (VAE) to train and extract topics from short texts. Initially, the short texts are transformed into longer texts using probabilistic and NTM models, due to which the topics may turn out to be diverse. To address this issue Prefix-tuned Variational Topic Model (PVTM) is used. This model is initially built with ProdLDA that uses VAE to represent data as the Bag-of-Words. However, this approach failed to capture the semantics. Thus, the existing approach was replaced by Prefix Tuning without modifying the pre-trained weights, to achieve appreciable topic inference. The datasets TagMyNews, Google News and Stackoverflow were used to evaluate the proposed

International Research Journal on Advanced Engineering Hub
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 80 - 88
www.irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0011

method against other baseline topic models mentioned on the previous literature and coherence score & Inverted Rank-Biased Overlap (IRBO) were computed for topic coherence. The researchers [16] have attempted to integrate LLM with the proposed Qualitative Insights Tools (QualIT) for the extraction of topics from the news articles. The QualIT is designed to have three phases: Key-Phrase Extraction, Hallucination Check and Clustering. The Key-Phrase Extraction phase overcomes BERT's single topic extraction by using LLM that extracted multiple key phrases from a document. The phase Hallucination Check tagged "Hallucination" for the key phrases below 10% of the coherence score. The clustering phase segregated the documents based on the extraction of main and sub topics. The proposed QualIT method was evaluated on 20 Newsgroup dataset comprised of 20,000 news articles. Initially the dataset was preprocessed to remove tokens with length of 3 and lemmatized the processed tokens. The proposed approach is compared with BERT in terms of the number of topics, topic coherence and topic diversity that achieved extracting 20 topics with topic coherence of 70% and 95.5% in topic diversity, where BERT performed 65.5% in topic coherence and topic diversity with 80%. SciTopic, a proposed framework that uses enhanced LLM led by clustering for the extraction of topics from scientific data. Initially, the proposed framework encodes the metadata of documents combined with embeddings to form features. All the documents were grouped with clusters of ambiguous contexts included to form triplet tasks that prompt the LLM to cluster the most coherent documents and provide deeper insights into the relationship among topics. Finally, the refined documents were fine-tuned with embedding models for better clustering. The scholarly articles from various datasets, namely Database systems and Logic Programming (DBLP v10), Artificial Intelligence for Diabetes Mellitus or AI-assisted Dynamic Modeling (AI-DM), Ninth Annual Conference on Neural Information Processing Systems (NeurIPS), arXiv and PubMed were used to evaluate the proposed framework against the traditional approaches namely LDA, NMF and ProdLDA. SciTopic achieves improvement on 21.8% of Topic coherence, Topic diversity with 14.6% and 5.61% Calinski-Harabasz Index and Davies-Boundin Index as clustering evaluation metrics [17].

Table 1 Extracts of Literature Review on Topic Modeling

Author(s)	Objective	Source	Topic Models	Prompting Techniques	Clustering	Number of Topics	Metrics	Results	Limitations
Azher et al. [2024]	Extraction of the limitation section from Scientific Articles	ACL Scientifc Articles	✓ LLM integrated ✓ BERT (all-MiniLM-L6-v2)	✓ GPT 4 ✓ GPT 3.5 ✓ LIama ✓ Mistral	HDBSCAN	10	Silhouette Score	0.588	✓ Limited GPU Constraints ✓ Faces Biases namely Demographic, Language and Methodolgoical towards prominent topics
							Coherence Score	0.617	
Invernici et al. [2025]	Analyzing and grouping of SDG based scientific articles	Abstracts of SDG based Articles from Elsevier Scopus	✓ Enhanced BERT with LLM	Not Mentioned	HDBSCAN	30	DBCV	0.76	✓ High GPU Requirements ✓ Lack of using global optimum during hyperparameter tuning
Pham et al. [2024]	Prompt-based Framework for Topic Extraction from text	Articles from Wikipeida (Wiki) and bill summaries	✓ LDA ✓ BERT ✓ SeededLDA ✓ TopicGPT	Mistral 7B	Not Mentioned	79	Harmonic Purity	0.74 (Wiki) 0.57 (Bills)	✓ Lack of information on architectural design and tuning of closed source models
						22	ARI	0.60 (Wiki)	

e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 80 - 88
www.irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0011

	collection	from Bills					NMI	0.70 (Wiki)	✓ TopicGPT does not handled in non-english datasets
						24	ARI	0.40 (Bills)	
							NMI	0.49 (NMI)	
El-Gayyar et al. [2024]	Comparision of Topic Modeling Techniques	Reviews from Health Applications	✓ LDA ✓ BERT ✓ RoBERT	Not Mentioned	HDBSCAN	8	Silhouette Score	BERT / RoBERT	✓ LDA lacks to cover semantic information ✓ Need to explore probabilistic based models and LLM for TM
								0.954 / 0.966	
						11		0.947 / 0.958	
						12		0.936 / 0.956	
						94	Coherence Score	0.608	
						97		0.628	
						81		0.612	
De-Marcos et al. [2025]	Compares LLM with Traditional TM Techniques	Dark web Q&A forums	✓ LDA ✓ LLM	✓ GPT (gpt-4o-mini) ✓ Gemini (Gemini-2.0-flash)	Not Mentioned	5	Semantic Similarity Score	0.724	✓ Computational cost dealing with LLM ✓ Potential Bias in LLM ✓ Evaluation Metrics NPMI and UMass (University of Massachusetts) are not included for robustness ✓ The techniques focused two-word topics than multi-word topics due to short length texts ✓ Phrase detection techniques or domain specific fine-tuning need to be done
							Levenshtein Distance	13.49	
Xu et al. [2023]	An Encoder-Decoder based LLM is proposed to ensure clusterability and semantic coherence	AG News 20 Newsgroup BBC News	✓ NTM ✓ DeTIME	Not Mentioned	Not Mentioned	20	Top-Purity	0.4577	✓ Other encoder techniques namely BART, PEFT and LORA are not used ✓ The model lacks in utilizing more parameters of Flan T5 ✓ Model lacks in optimizing the CNN encoder outputs ✓ The model uses standard architectures CNN, LSTM and MLP than advanced
							Top-NMI	0.2983	
							Km Purity	0.5929	
							Km NMI	0.3463	
							Diversity	0.6913	
							Coherence	0.7203	
Jung et al. [2024]	Exploring TM techniques aligning LLM	Web of Science and LexisNexis	✓ LDA ✓ NMF ✓ CTM ✓ BERT (multi-qa-miniLM-L6-	-	HDBSCAN	10	Coherence	0.450	✓ Data regarding social media are uncovered ✓ Deep Learning based TM techniques are not implemented

IRJAEH

International Research Journal on Advanced Engineering Hub
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 80 - 88
www.irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0011

			cos-v1)						
Zhao et al. [2025]	Identification of Mental discourse using LLM Framwork	Reddit	✓ LDA ✓ BERT ✓ LLM	✓ GPT 3.5 turbo ✓ Mistral 7B	HDBSCAN	110	Purity	0.7514	✓ Not Mentioned
							ARI	0.5014	
							MIS	0.6672	
Yang et al. [2025]	Evaluation Method for TM on semantic quality	20 Newsgroup DBpedia	✓ NTM ✓ LDA ✓ LLM	LLaMA-3-8B-Instruct	Not Mentioned	50	$S_{overlap}$	0.52	✓ Model biased towards LLM ✓ Model needs to be evaluated with other traditional metrics such as NPMI, Coherence, Diversity
							S_{synset}	0.58	
							S_{oa}	0.56	
							S_{ot}	0.68	
Janssens et al. [2025]	Integration of BERT with LLM on Topic Reduction	Twitter/X	✓ BERT (all-MiniLM—L6-v2) ✓ LLM	✓ GPT-4o-mini ✓ Llama-3-8-B ✓ Gemma-3-12B ✓ Qwen-3-30B-A3B	HDBSCAN	25	NPMI	0.1907	✓ Advanced Prompting strategies need to be designed to address the overlapping topics ✓ Need to Optimize the parameters of BERT
						50		0.1610	
						15		0.0534	
						25	Topic Diversity	0.8920	
						50		0.9255	
						15		0.8850	
Akash et al. [2024]	LLM for extending short texts before Topic Modeling	TagMyNews Google News Stack Overflow	✓ LDA ✓ BERT ✓ NTM ✓ DeTiME ✓ PVTM	Not Mentioned	Not Mentioned	20	Coherence Score (Tag MyNews)	0.632	✓ Text Generation may lead to addition of Irrelevant Information
						50		0.585	
						20	IRBO (Tag MyNews)	1.000	
						50		1.000	
Kapoor et al. [2024]	Integration of LLM with clustering based TM approaches	20 Newsgroup	✓ BERT ✓ LDA ✓ QualIT	Not Mentioned	K-Means	20	Coherence Score	0.75	✓ The approach lacks in using of HDBSCAN based clustering to handle complex data patterns.
							Topic Diversity	0.95	
Li et al. [2025]	Enhanced LLM focusing on thematic relevance and contextual differences in TM	AI-DM DBLP V10 NeurIPS arXiv PubMed	✓ LDA ✓ NMF ✓ ProdLDA ✓ NTM ✓ BERT ✓ SciTopic	Not Mentioned	Not Mentioned	100	Coherence Score (NeurIPS)	0.657	✓ Need to evaluate with diverse datasets to ensure clustering interpretability
							Topic Diversity (NeurIPS)	0.973	
							CHI (NeurIPS)	11.049	

5. Research Gap

Based on the thorough investigation of the existing literature on the topic modeling approaches, it is evident that the following challenges were identified, which invites for research progression in TM.

International Research Journal on Advanced Engineering Hub
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 80 - 88
www.irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0011

- Implementation of approaches for detecting subtopics or multi-word topics need to be considered
- Approaches to Topic Reduction need to be addressed
- Analysis of short texts and parameters to control the extended texts before TM
- Need to consider the approaches for extended Topic generation
- Need to analyze and optimize the architectural design and Hyperparameters of BERT, integrating with LLM
- Bias in implementing the advanced approach, such as LLM and its variants

Conclusion

The article provides a detailed insight into both conventional and advanced approaches for topic modeling along with the respective metrics. From this literature review, it is evident that advanced LLM models can handle both longer and shorter texts. It is apparent that appropriate integration of LLM with BERT induced with Autoencoders and Prefix-tuning have given for better topic extraction. Such integrated TM approaches were applied on short texts and was inferred that the results were reasonably be semantically rich even when dealing with datasets of various fields of data viz., medical and healthcare. This review article shall serve as a comprehensive guide for those researchers interested to develop robust and novel topic modelling techniques.

Acknowledgement

The authors gratefully acknowledge the financial support by National Fellowship for Other Backward Classes (NFOBC) scheme of Ministry of Social Justice & Empowerment and computational facilities granted by UGC-Special Assistance Programme (SAP), Image Processing Laboratory, Department of Computer Science and Applications, The Gandhigram Rural Institute (Deemed to be University).

References

[1]. Abdelrazek, A., Eid, Y., Gawish, E., Medhat, W., & Hassan, A. (2023). Topic modeling algorithms and applications: A survey. Information Systems, 112, 102131.
[2]. Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. Information Systems, 94, 101582.
[3]. Churchill, R., & Singh, L. (2022). The evolution of topic modeling. ACM Computing Surveys, 54(10s), 1-35.
[4]. Haritha, P., &Shanmugavadivu, P. (2023, December). Optimized Latent-Dirichlet-Allocation Based Topic Modeling–An Empirical Study. In International Conference on Speech and Language Technologies for Low-resource Languages (pp. 412-419). Cham: Springer Nature Switzerland.
[5]. Azher, I. A., Seethi, V. D. R., Akella, A. P., &Alhoori, H. (2024, December). Limtopic: Llm-based topic modeling and text summarization for analyzing scientific articles limitations. In Proceedings of the 24th ACM/IEEE Joint Conference on Digital Libraries (pp. 1-12).
[6]. Invernici, F., Curati, F., Jakimov, J., Samavi, A., &Bernasconi, A. (2025). Capturing research literature attitude towards sustainable development goals: an LLM-based topic modeling approach. Journal of Big Data, 12(1), 139.
[7]. Pham, C. M., Hoyle, A., Sun, S., Resnik, P., &Iyyer, M. (2024, June). Topicgpt: A prompt-based topic modeling framework. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 2956-2984).
[8]. El-Gayar, O., Al-Ramahi, M., Wahbeh, A., Nasralah, T., &Elnoshokaty, A. (2024). A comparative analysis of the interpretability of lda and llm for topic modeling: The case of healthcare apps.

International Research Journal on Advanced Engineering Hub
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 80 - 88
www.irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0011

[9]. De-Marcos, L., &Domínguez-Díaz, A. (2025). LLM-Based Topic Modeling for Dark Web Q&A forums: A Comparative Analysis with Traditional Methods. IEEE Access.
[10]. Xu, W., Hu, W., Wu, F., &Sengamedu, S. (2023). DeTiME: Diffusion-enhanced topic modeling using encoder-decoder based LLM. arXiv preprint arXiv:2310.15296.
[11]. Jung, H. S., Lee, H., Woo, Y. S., Baek, S. Y., & Kim, J. H. (2024). Expansive data, extensive model: Investigating discussion topics around LLM through unsupervised machine learning in academic papers and news. Plos one, 19(5), e0304680.
[12]. Zhao, C., & Chen, Y. (2025, June). LLM-powered Topic Modeling for Discovering Public Mental Health Trends in Social Media. In IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 119-132). Cham: Springer Nature Switzerland.
[13]. Yang, X., Zhao, H., Phung, D., Buntine, W., & Du, L. (2025). Llm reading tea leaves: Automatically evaluating topic models with large language models. Transactions of the Association for Computational Linguistics, 13, 357-375.
[14]. Janssens, W., Bogaert, M., &Poel, D. V. D. (2025). LLM-assisted topic reduction for BERTopic on social media data. arXiv preprint arXiv:2509.19365.
[15]. Akash, P. S., & Chang, K. C. C. (2024). Enhancing short-text topic modeling with LLM-driven context expansion and prefix-tuned VAEs. arXiv preprint arXiv:2410.03071.
[16]. Kapoor, S., Gil, A., Bhaduri, S., Mittal, A., & Mulkar, R. (2024). Qualitative insights tool (qualit): Llm enhanced topic modeling. arXiv preprint arXiv:2409.15626.
[17]. Li, P., Wang, Z., Zhang, X., Zhang, R., Jiang, L., Wang, P., & Zhou, Y. (2025). Scitopic: Enhancing topic discovery in scientific literature through advanced llm. arXiv preprint arXiv:2508.20514.