Topic Models Towards the Modern: A Comprehensive Survey of Topic Models in the Era of Large Language Models

Anonymous ACL submission

Abstract

LLMs have become foundational across many NLP applications, driving a shift from an algorithm-centric to a context-centric paradigm. As an important task in text mining, the landscape of topic modeling (TM) is similarly being reshaped by a growing body of LLM-driven research. From the perspective of LLM involvement, we review recent TM developments and categorize existing methods into three groups: Classical, LLM-Assisted, and LLM-Centric. For Classical methods, we refine prior taxonomies and highlight recent advances. For the LLM-Assisted and LLM-Centric settings, we introduce a new taxonomy that emphasizes the role of LLMs and the design of end-to-end workflows, respectively. We further discuss how the TM paradigm may evolve—spanning task boundaries, input choices, and distributional definitions—and outline key challenges and prospective future directions. We hope this survey catalyzes deeper integration between TM and LLMs—advancing topic modeling that is easier to use, more powerful, and more flexible.

1 Introduction

Topic modeling (TM), as a classic unsupervised learning task in the field of Natural Language Processing (NLP), aims to reduce high-dimensional text data to a low-dimensional topic space, automatically extracting topic information and parsing topic structures from a massive number of documents. With the rapid advancement of powerful large language models (LLMs) transforming NLP, the field of TM has increasingly been reshaped by a growing body of LLM-driven research, including studies that integrate LLMs with traditional models as well as those that employ LLMs exclusively as standalone solutions.

Prior surveys (Abdelrazek et al., 2023; Wu et al., 2024a; Hankar et al., 2025) have detailed a wide range of techniques—such as probabilistic generative modeling (Blei et al., 2003), autoencoding variational inference (Srivastava and Sutton, 2017), and clustering algorithms (Sia et al., 2020), along with mature toolkits like LDA (Yut et al., 2017), Top2Vec (Angelov, 2020), and neural topic models (NTMs) (Wu et al., 2024c). They have also covered NTMs optimized for specific scenarios and requirements, including short-text settings, hierarchical topics, and multilingual or multimodal contexts. However, most prior reviews emphasize innovations at the NTM level and do not focus on the involvement of LLMs.

The rise of LLMs is reshaping the landscape of TM. TM is increasingly being approached as an LLM-assisted or LLM-centric task, despite appearing to lag behind other NLP tasks such as named entity recognition, summarization, and machine translation. Nevertheless, the field continues to evolve rapidly, as evidenced by a growing body of recent research. In LLM-assisted settings, LLMs inject rich semantic priors into topic modeling across data augmentation, representation, inference, and evaluation: they expand short texts and summarize content (Akash and Chang, 2024), replace bag-of-words with robust embeddings (Grootendorst, 2022), and inspire post-processing frameworks (pariskang, 2025) that improve coherence, diversity, and interpretability. They also enable evaluation protocols that jointly assess document and topic semantics, with LLM-based metrics (Yang et al., 2025a) showing stronger alignment with human judgments than traditional automated scores. In LLM-Centric settings, many studies have proposed effective solutions in which the LLM serves as the primary inference mechanism for discovering topics, encompassing both single-turn (Xu et al., 2025a) and multi-turn (Hsu et al., 2024) interactions, as well as sequential and parallel LLM reasoning (Doi et al., 2024).

Our contributions are summarized as follows:

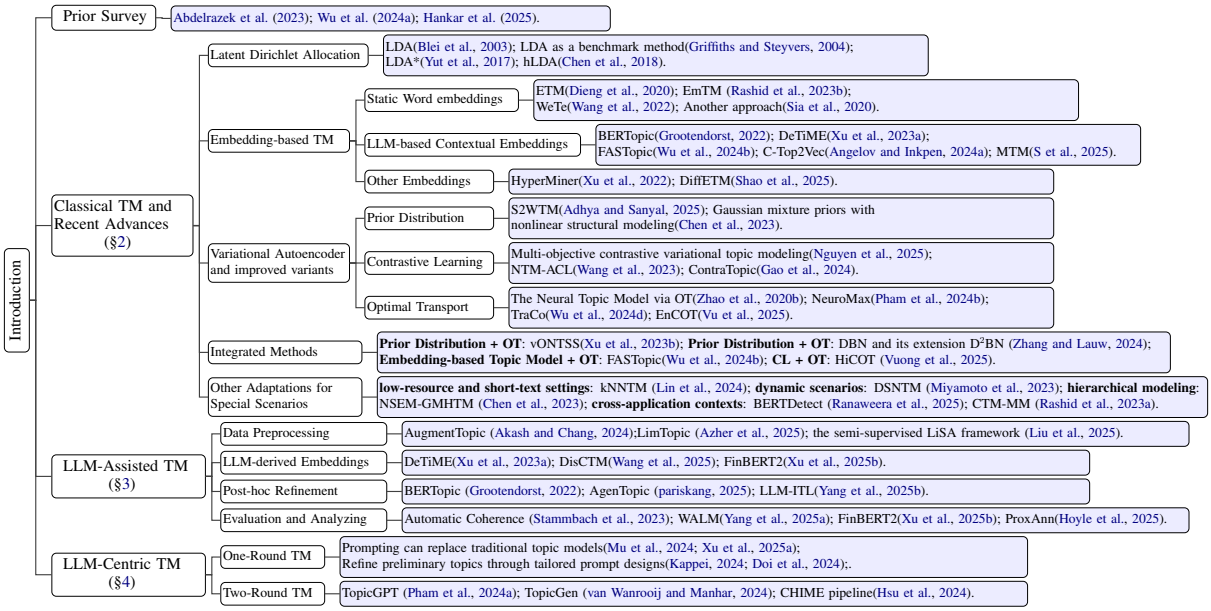

Figure 1: Taxonomy of Topic Modeling methods.

1. Updated perspective on TM in LLM era: We survey TM through the lens of LLM involvement and categorize methods into *Classical*, *LLM-Assisted*, and *LLM-Centric*. For *Classical* TM, we refine prior taxonomies with a special focus on advances from the past three years.

2. Systematic taxonomies for LLM-Assisted and LLM-Centric TM: We propose classification frameworks that clarify LLM roles (e.g., prior injection, supervision, constraint/feedback) and emphasize end-to-end workflow design.

3. Insightful discussions about paradigm evolution, challenges, and directions: We analyze how TM is transforming with LLMs—spanning task boundaries, input choices, and topic-distribution definitions—and identify key challenges and future research avenues.

2 Classical Topic Models and Recent Advances (Appendix B.1)

2.1 Latent Dirichlet Allocation (Tables 2)

Classical topic modeling originated with Latent Dirichlet Allocation (LDA) (Blei et al., 2003), which represents documents as mixtures of latent topics and models each topic as a probability distribution over words. Owing to its mature theoretical foundation, LDA has long served as a benchmark method in topic modeling research (Griffiths and Steyvers, 2004). Additionally, large-scale implementations such as LDA* (Yut et al., 2017) have been proposed to provide robust and efficient distributed inference across massive corpora. Furthermore, hierarchical extensions, such as hierarchical LDA (hLDA) (Chen et al., 2018), leverage the nested Chinese Restaurant Process to capture topic hierarchies and have recently been scaled up to industrial-scale datasets. However, LDA's bag-of-words assumption misses long-range dependencies, limiting performance on unstructured or noisy data. As probabilistic generative models, LDA variants also struggle on sparse short texts and depend heavily on manual preprocessing.

2.2 Embedding-based Topic Model (Tables 3)

Embedding-based topic models map words, documents, and topics into a continuous vector space, use pretrained word/sentence embeddings (e.g., Word2Vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014), BERT (Devlin et al., 2019)) to capture semantic similarity, and discover topics via clustering or probabilistic modeling in that space.

Static Word embeddings. A line of research builds topic models on static word embeddings, most often trained with Euclidean embeddings such as Word2Vec or Glove. The Embedded Topic Model (ETM) (Dieng et al., 2020) embeds both texts and topics into the latent semantic space, thus capturing semantic similarity more effectively than traditional bag-of-words approaches. Another approach (Sia et al., 2020) involves clustering pre-trained word embeddings combined with document-level information to perform weighted clustering and re-rank candidate keywords. Re-

cent works also follow this setting with additional improvements. EmTM (Rashid et al., 2023b) addresses the challenges of short text modeling by first embedding words with Word2Vec and then clustering them to form topics, which helps overcome data sparsity and improves interpretability in short documents. WeTe (Wang et al., 2022) models each document as a mixture of word embeddings and each topic as a distribution of topic embeddings in the same space, learning fine-grained topics by minimizing their semantic distance.

LLM-based Contextual Embeddings. Transformer-based models such as BERT provide contextual embeddings that capture semantic meaning beyond individual words. Building on this idea, BERTopic (Grootendorst, 2022) combines BERT-derived document embeddings with HDBSCAN clustering and class-based TF–IDF to produce interpretable topics in static and dynamic corpora. Xu et al. (2023a) proposed the DeTiME framework, which integrates encoder–decoder LLM-based embeddings with diffusion-based augmentation to optimize clustering quality and enable topic-related content generation, addressing the suboptimal performance of traditional neural topic models (NTMs). FASTopic (Wu et al., 2024b) introduces a clean encoding–alignment framework, using pretrained Transformer embeddings and Dual Semantic Reconstruction to align documents with learnable topic/word embeddings for efficient topic discovery. C-Top2Vec (Angelov and Inkpen, 2024a) leverages contextual token embeddings to build hierarchies, detect in-document topic spans, and label topics with phrases, easing choices of topic number and segmentation. MTM (S et al., 2025) extends to multi-view contextual embeddings and adds an embedding clustering regularization loss to improve topic stability and semantic diversity across views.

Other Embeddings. Beyond the aforementioned Euclidean embeddings, HyperMiner (Xu et al., 2022) introduces hyperbolic embeddings to model topic taxonomies, leveraging the tree-like properties of hyperbolic space to capture hierarchical semantic structures and integrate prior knowledge into topic modeling. More recently, DiffETM (Shao et al., 2025) introduces a diffusion-regularized representation enhancement that samples hidden variables directly from document embeddings—injecting document semantics into the latent space while gradually introduces noise conforming to the normal distribution into the document representations.

2.3 Variational Autoencoder (Tables 4-5)

Neural Topic Models (NTMs) emerged by fusing probabilistic topic modeling with deep neural networks—especially variational autoencoders (VAE) and neural decoders—to overcome bag-of-words and linear prior limits in classical models like LDA; they retain latent variables for document–topic and topic–word structure while using neural networks for amortized inference and generation. A typical representative is ProdLDA (Srivastava and Sutton, 2017), which reformulates Latent Dirichlet Allocation as a VAE by using an encoder-decoder structure, where the encoder approximates the posterior over topic proportions and the decoder generates word distributions. Building on this paradigm, subsequent research has advanced along several directions.

Prior Distribution. A key line of research refines the VAE prior to alleviate posterior collapse and strengthen latent topic representations. Chen et al. (2023) integrate Gaussian mixture priors with nonlinear structural modeling to induce more semantically meaningful topics and to organize them into coherent hierarchies. S2WTM (Adhya and Sanyal, 2025) adapts the latent space to a hyperspherical geometry and employ the Spherical Sliced-Wasserstein distance to better align the aggregated posterior with the prior.

Contrastive Learning. Contrastive learning (CL) has been applied to the VAE framework to improve topic quality such as topic coherence, topic diversity, and downstream performance. NTM-ACL (Wang et al., 2023) embeds contrastive signals through a cycle adversarial training scheme, while ContraTopic (Gao et al., 2024) attaches topic-wise contrastive regularization to the VAE loss, leveraging corpus statistics to enhance intra-topic coherence and inter-topic diversity. Nguyen et al. (2025) proposed a set-oriented CL approach to capture shared semantics across documents and formulate contrastive topic modeling as gradient-based multi-objective optimization to reach a Pareto-stationary balance between the ELBO and the contrastive objective.

Optimal Transport. Optimal transport (OT) is a mathematical framework for comparing and transforming probability distributions with minimal "cost." which has been incorporated into NTM to better align latent distributions and improve topic interpretability. The Neural Topic Model via

OT (Zhao et al., 2020b) uses transport-based objectives to capture document–topic relationships more accurately. With a context-aware decoder, TraCo further improves the coherence and rationality of hierarchical topic structures. OT-based objectives thus also help LLMs maintain semantic and structural consistency over long contexts. NeuroMax (Pham et al., 2024b) further optimizes the encoder by maximizing mutual information and introduces group-level regularization via OT, enhancing the informativeness and separation of latent topics. For hierarchical modeling, the Transport Plan and Context-aware Hierarchical Topic Model (TraCo) (Wu et al., 2024d) introduces sparse balanced transport plans to model parent–child dependencies while maintaining diversity among sibling topics. For short texts, Enhancing Global Clustering with OT in Topic Modeling (EnCOT) (Vu et al., 2025) applies a dual alignment mechanism, simultaneously matching documents with clusters and topics with clusters, which alleviates sparsity and strengthens topic separation.

2.4 Integrated Methods (Tables 6)

Some methods draw on the strengths of multiple strategies mentioned above. **(Prior Distribution + OT)** The vONTSS (Xu et al., 2023b) framework adopts a vMF prior while introducing an OT-based semi-supervised loss, aligning topic–word relations with external knowledge. **(Prior Distribution + OT)** The Dirichlet OT Barycenter Network (DBN) and its extension D^2BN (Zhang and Lauw, 2024) integrate Dirichlet priors with OT barycenters and graph neural networks, bridging semantic interpretability and structural modeling. **(Embedding-based Topic Model + OT)** FASTopic (Wu et al., 2024b) presents a clean and efficient framework combining encoding with OT whose Dual Semantic Reconstruction (DSR) mechanism directly models the semantic relationships between document embeddings from pre-trained Transformers and learnable topic/word embeddings, enabling effective discovery of latent topics. **(CL + OT)** HiCOT (Vuong et al., 2025) further unifies OT with hierarchical clustering and contrastive objectives, reinforcing both coherence and diversity in topic discovery.

2.5 Other Adaptations for Special Scenarios (Tables 7)

Several advances tailor topic modeling to challenges such as sparsity, temporal evolution, hierarchical structure, and multimodality, extending its applicability to diverse and demanding scenarios. In **low-resource and short-text settings**, kNNTM (Lin et al., 2024) tackles the problem of label sparsity by augmenting documents with nearest neighbors, which enriches the ELBO reconstruction target and makes VAE training more stable. For **dynamic scenarios**, DSNTM (Miyamoto et al., 2023) employs a self-attention mechanism within the VAE framework to capture temporal dependencies among topics, while citation-based regularization improves the alignment of evolving themes across time. In **hierarchical modeling**, NSEM-GMHTM (Chen et al., 2023) extends the prior distribution to a Gaussian mixture and integrates nonlinear structural equations, allowing the latent space to better represent both hierarchical and symmetric topic dependencies in sparse corpora. In **cross-application contexts**, VAE-inspired models such as BERTDetect (Ranaweera et al., 2025) and CTM-MM (Rashid et al., 2023a) adapt neural topic modeling to domains like Android malware detection and multimodal social media analysis, showing that the generative flexibility of VAEs can enhance semantic coherence under domain-specific constraints.

3 LLM-Assisted Topic Model (Appendix B.2)

In LLM-assisted topic modeling, the LLM serves an auxiliary role for specific subtasks, while the core inference over topics and document–topic distributions is performed by a non-LLM model, typically a NTM.

3.1 Data Preprocessing (Tables 8)

The multitask capabilities of LLMs greatly facilitate corpus preprocessing for adapting to different scenarios and models. Many preprocessing techniques such as summarization, expansion, rewriting/cleaning, element extraction (e.g., NER), and label generation can be employed to improve corpus quality and provide auxiliary supervision, thereby enhancing the effectiveness of topic modeling. Akash and Chang (2024) combine LLM-Centric context expansion with prefix-tuned variational autoencoders for short-text topic modeling. More advanced approaches, such as LimTopic (Azher et al., 2025) and the semi-supervised LiSA framework (Liu et al., 2025), leverage LLMs to generate candidate topic words and perform semantic-aware clustering, addressing the limitations of prior

methods that overlook fine-grained document semantics.

3.2 LLM-derived Embeddings (Tables 9)

This part leverages LLM-derived embeddings as semantically informed inputs to classic topic modeling, replacing bag-of-words with stronger priors and initializations. It overlaps with the advanced embedding-based methods in Section 2.2, including notable examples such as BERTopic and FASTopic.

Some innovations focus on refining LLM-derived embeddings for more efficient representations—for example, by leveraging more advanced LLM architectures, applying disentanglement, and enabling domain adaptation. Xu et al. (2023a) introduced DeTiME, which integrates encoder–decoder LLM embeddings with diffusion-based augmentation to improve clustering quality and support topic-related content generation, addressing the suboptimal performance of traditional NTMs. For domain adaptation, Wang et al. (2025) proposed DisCTM, a Transformer-based model with a topic disentanglement mechanism that better captures thematic structure and alleviates sparsity in short-text corpora. The Fin-Topicmodel in FinBERT2 (Xu et al., 2025b) uses fine-tuned embeddings in financial datasets as more efficient representations.

3.3 Post-hoc Refinement (Tables 10)

Works in this part typically leverage LLMs to post-process topic word lists generated by NTMs. Through refinements such as rewriting, correcting, merging, LLMs can enhance topic coherence, diversity, and readability. For instance, the widely used BERTopic (Grootendorst, 2022) library already provides functionality to incorporate custom prompts and call LLM APIs for regenerating topic representations. AgenTopic (pariskang, 2025), for example, is designed to analyze scientific literature by integrating GPT-4 with cache modules, whose post-processing with LLMs enables iterative refinement, which reduce redundancy among topics, decreases errors in document–topic assignments, and improves interpretability. Similarly, Yang et al. (2025b) proposed the LLM-in-the-loop (LLM-ITL) framework, which integrates LLMs with NTMs: NTMs learn global topics and document representations, while LLMs refine these topics via an OT-based alignment objective, dynamically adjusted according to the LLM's confidence in suggesting topic words.

3.4 Evaluation and Analyzing (Tables 11)

LLM-based subjective judging has become common and shows clear advantages in tasks such as summarization and reasoning. Recent work also highlights the role of LLMs in evaluating and analyzing topic modeling results. Stammbach et al. (2023) show that LLM-based evaluations correlate more strongly with human judgments than traditional metrics and can guide researchers in selecting an appropriate number of topics. Yang et al. (2025a) propose WALM (Word Agreement with Language Model), which jointly assesses document representations and topic quality, aligning more closely with human evaluation and complementing existing automated measures. The topic modeling part in FinBERT2 (Xu et al., 2025b) prompts LLMs to evaluate topic representations across three dimensions: coherence, conciseness, and informativeness. Hoyle et al. (2025) present the ProxAnn framework, where LLM-based proxy annotations are statistically indistinguishable from human annotations, suggesting LLMs can serve as reliable substitutes in automated evaluation.

However, studies note that LLMs struggle on large-scale and domain-specific corpora. Li et al. (2025) find that they often yield generic themes and miss domain-specific topics. Limited context windows also introduce hallucinations and scalability issues, constraining their standalone use for large-scale topic modeling.

4 LLM-Centric Topic Model (Appendix B.3)

In LLM-centric topic modeling, the LLM serves as the primary inference mechanism for discovering topics, generating topic assignments/word lists via prompting and others. Prompts of methods mentioned in this section are summarized in Appendix B.3.

4.1 One-Round LLM-Centric Topic Model

One-Round LLM-Centric Topic Model is a paradigm where the entire corpus is processed by an LLM in a single inference pass to produce the topic structure (often labels or hierarchies). Any subsequent steps that do not invoke additional LLM inference—such as keyword matching, clustering, or other non-LLM postprocessing—are allowed. Reasoning over, summarizing, or reorganizing the collected output within that same pass is still considered one round.

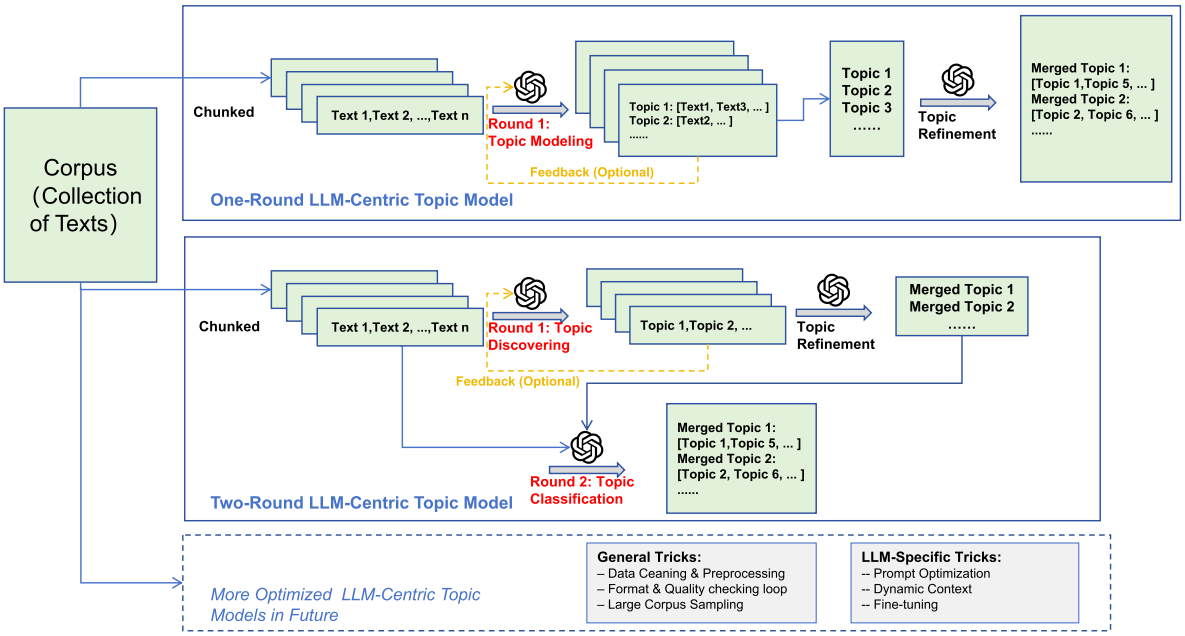

Figure 2: Illustration of Two Types of LLM-Centric Topic Models

Studies (Mu et al., 2024; Xu et al., 2025a) show that prompting can replace traditional topic models in a single pass, yielding human-aligned topics: both work position LLMs as out-of-the-box topic generators, with the latter casting topic modeling as long-form generation and demonstrating zero-shot parity or advantages over NTMs. Studies (Doi et al., 2024; Kappei, 2024) refine preliminary topics through tailored prompt designs (e.g., parallel or sequential prompting), improving coherence and diversity while reducing hallucinations—especially on small corpora ($N < 1000$) or very short texts.

4.2 Two-Round LLM-Centric Topic Model

Two-Round setup uses two LLM inference passes over the corpus. The first pass typically proposes or organizes candidate topics (labels, hierarchies). The second pass focuses on assigning documents to topics or refining the mapping between documents and topics. Two-Round LLM-Centric Topic Models are commonly utilized to handle large corpora, compensate for the current limitations of LLMs, and satisfy higher accuracy demands.

TopicGPT (Pham et al., 2024a) illustrates this paradigm: it produces candidate topics, removes redundant or infrequent ones, and assigns each document with supporting evidence. TopicGen (van Wanrooij and Manhar, 2024) targets small-data scenarios, introducing a three-phase pipeline of preliminary topic generation, consolidation for thematic clarity, and document classification, achieving competitive results compared to neural baselines. Similarly, Hsu et al. (2024) developed CHIME, an LLM-assisted framework for hierarchical scientific study organization to support literature reviews. Although TopicGPT and CHIME add an extra round to generate seed topics or apply extraction constraints, we regard this as preprocessing and still categorize them under the two-round framework.

5 Discussions

5.1 The Paradigm Shift Brought by LLM-Centric Topic Model

5.1.1 Clarifying Task Boundaries

We first delineate the boundaries between topic modeling and other topic-related tasks. Traditionally, topic models are clearly scoped and cannot address other tasks. In contrast, LLM-Centric topic modeling is closely connected to semantic extraction and organization tasks—such as topic classification, discovery, extraction, segmentation, hierarchical modeling, and corpus organization. LLMs' multi-task abilities enable tighter integration across tasks. For example, topic modeling can be decomposed into "topic discovery + topic classification," and corpus organization can be seen as a multi-round extension of topic modeling. Each of these tasks is described below, with a detailed breakdown in Table 1.

Task	Input	Output	Topic Representation	Supervision	Computation	LLM Applicability
TC	Single doc, predefined topics	Topic label	Short phrase	Sup.	Low	Well-suited
TD	Doc. collection	Topic set	Keywords/phrases	Unsup.	Med.	Partial
TE	Single doc	Keywords	Phrases	Unsup.	Med.	Suitable
TS	Long doc	Segments + labels	Phrases/sentences	Weak/Unsup.	Med.	Suitable
TM	Doc. collection	Topic list + assignment	Word dist.	Unsup.	High	Somewhat
DOA	Doc. collection	Tree/graph catalog	Multi-dim. repr.	Opt.	High	Partial

Table 1: Comparison of different topic-related tasks and their characteristics. The table summarizes common tasks in topic modeling and organization, showing their input/output forms, topic representations, level of supervision, computational requirements, and the applicability of LLMs. Abbreviation: Sup. = Supervised, Unsup. = Unsupervised, Med. = Medium, Hier. = Hierarchical, Repr. = Representation.

Topic Classification (TC): Assigns a document to one or more predefined topics. It is essentially a supervised text classification problem, where each label corresponds to a semantic category.

Topic Discovery (TD): Identifies latent themes from a collection of documents without pre-specified topic labels. This unsupervised task aims to reveal hidden semantic structures.

Topic Extraction (TE): Extracts representative keywords or phrases from a single document that summarize its main theme, capturing salient topical cues directly from the text.

Topic Segmentation (TS): Divides a long document into coherent segments, each associated with a topic, in order to capture topic shifts within the same document.

Topic Modeling (TM): Learns latent topics from a document collection, representing each topic as a probability distribution over words and each document as a distribution over topics.

Document Organization & Analysis (DOA): Organizes a large document collection into more structured thematic catalogs (trees or graphs), combining topical semantics with additional analytical dimensions such as sentiment, stance, or temporal patterns.

5.1.2 Input Granularity and Length Choices

Beyond topic-centric tasks, topic modeling is also closely related to high-level semantic generation tasks that shorten text, such as summarization, title generation, and key entity extraction. These allow long-document topic modeling to be decomposed into short-text generation plus short-text topic modeling. Traditional TM favors long documents and large corpora to learn low-dimensional topics, while short texts require special handling due to sparse co-occurrences. In contrast, LLM-based TM reframes the task as "summarizing a collection." Although inputs may vary in granularity (titles, abstracts, full documents, or document collections), the core challenge lies at the title and abstract levels. Other granularities can be reduced to these via extraction or summarization during preprocessing, which increases flexibility with respect to input corpora.

5.1.3 Flexibility Across Modalities, Languages, and Distributions.

Multi-modal and multilingual inputs also benefit from LLM-Centric approaches, since LLMs are strong zero-shot learners that transfer knowledge effectively from their pretraining corpora. In contrast, traditional NTMs are heavily dependent on training corpora and are prone to failure when training–test distribution shifts occur. Moreover, LLMs can leverage out-of-distribution token sampling and knowledge reasoning to flexibly generate customized topic representations. As with other NLP tasks, LLMs can be applied to topic modeling via zero-shot prompting with a custom prompt, or through tailored pipelines inspired by specific goals or constraints.

5.1.4 Distributions in Classical Topic Models and Their Counterparts in LLM-Centric Setting

In traditional topic modeling methods such as LDA, the latent semantic structure of a document collection is characterized by document-topic distribution and topic-word distribution:$\{\theta_i\}, \{\beta_k\}$:

$$\theta_i \in \mathbb{R}^{1\times K}, \quad i = 1, \ldots, N$$

which represents the probability distribution of document X_i over K topics.

$$\beta_k \in \mathbb{R}^{1\times V}, \quad k = 1, \ldots, K$$

which represents the probability distribution of topic z_k over the vocabulary of size V.

Entropy of Topic Distribution Given Document Context. Given a document (or corpus) context X, let $P(z_k \mid X)$ denote the normalized distribution over candidate topics $\{z_k\}_{k=1}^{K}$ induced by

the language model. We define the entropy of the context-conditioned topic distribution as

$$H_{\text{topic}|X} \;=\; -\sum_{k=1}^{K} P(z_k \mid X)\,\log P(z_k \mid X),$$

which quantifies the model's uncertainty about which topics are salient under the given context. Lower entropy indicates a few dominant topics (clear topical focus), while higher entropy suggests diffuse or mixed topical signals.

Entropy of Keyword Distribution Given a Topic. For a given topic z_k, let $P_k(w)$ denote the normalized salience distribution over representative keywords $w \in \mathcal{V}$. We define the entropy of the topic–word distribution as

$$H_{\text{word}|z_k} \;=\; -\sum_{w\in\mathcal{V}} P_k(w)\,\log P_k(w),$$

which measures how concentrated the keywords are under topic z_k. Lower entropy reflects a sharper, more coherent set of core terms; higher entropy indicates a more diffuse or ambiguous keyword profile.

5.2 Challenges of LLM-Centric TM

TM is a typical "long-input + long-output" task: it requires processing ultra-long inputs and generating extensive topics and others. For this, LLMs must capture global information from thousands of documents and produce topic structures, leading to the following key challenges:

Challenge of Global Attention. LLM-Centric TM needs position-invariant attention across hundreds/thousands of documents while maintaining global thematic coherence. It must avoid over-attending to early/contiguous segment and aggregate cross-document evidence in ultra-long contexts—ensuring balanced coverage, de-duplication, and a consistent global topic set/hierarchy.

Challenge of Faithful Generations. The system must generate large topic inventories with readable, non-redundant, corpus-faithful document assignments. Key requirements include: consistency across long outputs, duplication/coverage control, traceable evidence for each topic, balanced granularity/distinctness.

Challenge of qualified Benchmarks. There is a lack of widely accepted benchmarks and standardized metrics tailored to LLM-centric TM Beyond traditional measures, evaluation should capture interpretability, structural consistency (including hierarchies), coverage, and faithfulness to source documents.

5.3 Towards More Optimized LLM-Centric TM in Future

Future LLM-centric topic modeling can be further optimized from multiple perspectives. These improvements aim to balance computational efficiency, thematic accuracy, and interpretability.

General Tricks. Preprocessing plays a crucial role in adapting diverse text corpora to LLM capabilities. Long documents can be summarized to fit model context limits, while short texts can be expanded or clustered to alleviate sparsity. Moreover, enforcing strict format and quality constraints—such as fixing the number of documents per topic or keywords per topic—ensures consistency and interpretability. For large-scale corpora, efficient sampling strategies can be applied: topics are inferred from sampled subsets, and remaining documents are assigned using lightweight clustering or C-TF-IDF methods, thereby reducing cost without sacrificing overall structure quality.

LLM-Specific Tricks. Prompt design can substantially enhance topic generation quality. Incorporating seed topics, few-shot exemplars, or user-specified granularity helps LLMs generate more accurate and coherent topic representations. Dynamic retrieval-based context loading—selecting clusters of documents around an anchor text—can further improve token efficiency and promote fine-grained topic discovery in diverse corpora. Finally, fine-tuning LLMs with task-specific datasets—through instruction tuning or reinforcement learning from human feedback (RLHF)—can enhance instruction-following ability and mitigate hallucination. Such targeted adaptation can make LLM-centric topic modeling pipelines more robust, controllable, and generalizable across domains.

6 Limitations

In this survey, several limitations remain. First, the categorization partly relies on subjective judgments of involvement/coupling depth, making boundary cases debatable. Second, our coverage is largely restricted to the past three years, limiting longer-term historical perspective. Third, some claims may overestimate the impact of integrating LLMs on TM while underemphasizing foundational probabilistic modeling and its theoretical depth.

References

Aly Abdelrazek, Yomna Eid, Eman Gawish, Walaa Medhat, and Ahmed Hassan. 2023. Topic modeling algorithms and applications: A survey. *Information Systems*, 112:102131.

Suman Adhya and Debarshi Kumar Sanyal. 2025. S2WTM: Spherical sliced-Wasserstein autoencoder for topic modeling. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 23211–23225, Vienna, Austria. Association for Computational Linguistics.

Pritom Saha Akash and Kevin Chen-Chuan Chang. 2024. Enhancing Short-Text Topic Modeling with LLM-Driven Context Expansion and Prefix-Tuned VAEs. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 15635–15646, Miami, Florida, USA. Association for Computational Linguistics.

Dimo Angelov. 2020. Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*.

Dimo Angelov and Diana Inkpen. 2024a. Topic modeling: Contextual token embeddings are all you need. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 13528–13539, Miami, Florida, USA. Association for Computational Linguistics.

Dimo Angelov and Diana Inkpen. 2024b. Topic Modeling: Contextual Token Embeddings Are All You Need. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 13528–13539, Miami, Florida, USA. Association for Computational Linguistics.

Ibrahim Al Azher, Venkata Devesh Reddy Seethi, Akhil Pandey Akella, and Hamed Alhoori. 2025. 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*, JCDL '24, pages 1–12, New York, NY, USA. Association for Computing Machinery.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

HeGang Chen, Pengbo Mao, Yuyin Lu, and Yanghui Rao. 2023. Nonlinear structural equation model guided Gaussian mixture hierarchical topic modeling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10377–10390, Toronto, Canada. Association for Computational Linguistics.

Jianfei Chen, Jun Zhu, Jie Lu, and Shixia Liu. 2018. Scalable training of hierarchical topic models. *Proc. VLDB Endow.*, 11(7):826–839.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.

Adji B Dieng, Francisco JR Ruiz, and David M Blei. 2020. Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439–453.

Tomoki Doi, Masaru Isonuma, and Hitomi Yanaka. 2024. Topic Modeling for Short Texts with Large Language Models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 21–33, Bangkok, Thailand. Association for Computational Linguistics.

Xin Gao, Yang Lin, Ruiqing Li, Yasha Wang, Xu Chu, Xinyu Ma, and Hailong Yu. 2024. Enhancing topic interpretability for neural topic modeling through topic-wise contrastive learning. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 584–597.

Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl_1):5228–5235.

Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *Preprint*, arXiv:2203.05794.

Mustapha Hankar, Mohammed Kasri, and Abderrahim Beni-Hssane. 2025. A comprehensive overview of topic modeling: Techniques, applications and challenges. 628:129638.

Alexander Miserlis Hoyle, Lorena Calvo-Bartolomé, Jordan Lee Boyd-Graber, and Philip Resnik. 2025. ProxAnn: Use-Oriented Evaluations of Topic Models and Document Clustering. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15872–15897, Vienna, Austria. Association for Computational Linguistics.

Chao-Chun Hsu, Erin Bransom, Jenna Sparks, Bailey Kuehl, Chenhao Tan, David Wadden, Lucy Wang, and Aakanksha Naik. 2024. CHIME: LLM-Assisted Hierarchical Organization of Scientific Studies for Literature Review Support. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 118–132, Bangkok, Thailand. Association for Computational Linguistics.

Nakano Kappei. 2024. A Novel Approach to Topic Modeling Using Large Language Models (LLMs).

Raymond Li, Felipe Gonzalez-Pizarro, Linzi Xing, Gabriel Murray, and Giuseppe Carenini. 2023. Diversity-aware coherence loss for improving neural topic models. In *Proceedings of the 61st Annual*

Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 1710–1722, Toronto, Canada. Association for Computational Linguistics.

Zongxia Li, Lorena Calvo-Bartolomé, Alexander Miserlis Hoyle, Paiheng Xu, Daniel Kofi Stephens, Juan Francisco Fung, Alden Dima, and Jordan Lee Boyd-Graber. 2025. Large Language Models Struggle to Describe the Haystack without Human Help: A Social Science-Inspired Evaluation of Topic Models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7583–7604, Vienna, Austria. Association for Computational Linguistics.

Yang Lin, Xinyu Ma, Xin Gao, Ruiqing Li, Yasha Wang, and Xu Chu. 2024. Combating label sparsity in short text topic modeling via nearest neighbor augmentation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 13762–13774, Bangkok, Thailand. Association for Computational Linguistics.

Jianghan Liu, Ziyu Shang, Wenjun Ke, Peng Wang, Zhizhao Luo, Jiajun Liu, Guozheng Li, and Yining Li. 2025. LLM-Guided Semantic-Aware Clustering for Topic Modeling. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 18420–18435, Vienna, Austria. Association for Computational Linguistics.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Nozomu Miyamoto, Masaru Isonuma, Sho Takase, Junichiro Mori, and Ichiro Sakata. 2023. Dynamic structured neural topic model with self-attention mechanism. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5916–5930, Toronto, Canada. Association for Computational Linguistics.

Yida Mu, Chun Dong, Kalina Bontcheva, and Xingyi Song. 2024. Large Language Models Offer an Alternative to the Traditional Approach of Topic Modelling. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 10160–10171, Torino, Italia. ELRA and ICCL.

Thong Nguyen, Xiaobao Wu, Xinshuai Dong, Cong-Duy T Nguyen, See-Kiong Ng, and Anh Tuan Luu. 2025. Topic modeling as multi-objective contrastive optimization. *Preprint*, arXiv:2402.07577.

pariskang. 2025. Pariskang/AgenTopic.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Chau Minh Pham, Alexander Hoyle, Simeng Sun, Philip Resnik, and Mohit Iyyer. 2024a. 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)*, pages 2956–2984, Mexico City, Mexico. Association for Computational Linguistics.

Duy-Tung Pham, Thien Trang Nguyen Vu, Tung Nguyen, Linh Van Ngo, Duc Anh Nguyen, and Thien Huu Nguyen. 2024b. NeuroMax: Enhancing neural topic modeling via maximizing mutual information and group topic regularization. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 7758–7772, Miami, Florida, USA. Association for Computational Linguistics.

Hamed Rahimi, Hubert Naacke, Camelia Constantin, and Bernd Amann. 2024. Antm: aligned neural topic models for exploring evolving topics. In *Transactions on Large-Scale Data-and Knowledge-Centered Systems LVI: Special Issue on Data Management-Principles, Technologies, and Applications*, pages 76–97. Springer.

Nishavi Ranaweera, Jiarui Xu, Suranga Seneviratne, and Aruna Seneviratne. 2025. Bertdetect: A neural topic modelling approach for android malware detection. In *Companion Proceedings of the ACM on Web Conference 2025*, WWW '25, page 1802–1810, New York, NY, USA. Association for Computing Machinery.

Junaid Rashid, Jungeun Kim, and Usman Naseem. 2023a. Coherent topic modeling for creative multimodal data on social media. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 3923–3927, New York, NY, USA. Association for Computing Machinery.

Junaid Rashid, Jungeun Kim, and Usman Naseem. 2023b. Incorporating embedding to topic modeling for more effective short text analysis. In *Companion Proceedings of the ACM Web Conference 2023*, WWW '23 Companion, page 73–76, New York, NY, USA. Association for Computing Machinery.

Junaid Rashid, Jungeun Kim, and Usman Naseem. 2023c. Incorporating Embedding to Topic Modeling for More Effective Short Text Analysis. In *Companion Proceedings of the ACM Web Conference 2023*, pages 73–76, Austin TX USA. ACM.

Abilasha S, Rafika Boutalbi, and Stéphane Delliaux. 2025. Multi-view topic modeling using multi-text representations. In *Companion Proceedings of the ACM on Web Conference 2025*, WWW '25, page 1278–1282, New York, NY, USA. Association for Computing Machinery.

Wei Shao, Mingyang Liu, and Linqi Song. 2025. Diffetm: Diffusion process enhanced embedded topic model. *Preprint*, arXiv:2501.00862.

Suzanna Sia, Ayush Dalmia, and Sabrina J. Mielke. 2020. Tired of topic models? clusters of pretrained word embeddings make for fast and good topics too! In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1728–1736, Online. Association for Computational Linguistics.

Akash Srivastava and Charles Sutton. 2017. Autoencoding variational inference for topic models. *arXiv preprint arXiv:1703.01488*.

Dominik Stammbach, Vilém Zouhar, Alexander Hoyle, Mrinmaya Sachan, and Elliott Ash. 2023. Revisiting Automated Topic Model Evaluation with Large Language Models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9348–9357, Singapore. Association for Computational Linguistics.

van Wanrooij and Omendra Kumar Manhar. 2024. Topic modeling for small data using generative llms. In *Proceedings of the 36th Benelux Conference on Artificial Intelligence and the 33rd Belgian Dutch Conference on Machine Learning (BNAIC/BeNeLearn 2024)*, Utrecht, Netherlands. Benelux Association for Artificial Intelligence (BNVKI) and Netherlands Research School for Information and Knowledge Systems (SIKS).

Tu Vu, Manh Do, Tung Nguyen, Linh Ngo Van, Sang Dinh, and Thien Huu Nguyen. 2025. Topic modeling for short texts via optimal transport-based clustering. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 7666–7680, Vienna, Austria. Association for Computational Linguistics.

Hoang Tran Vuong, Tue Le, Tu Vu, Tung Nguyen, Linh Ngo Van, Sang Dinh, and Thien Huu Nguyen. 2025. HiCOT: Improving neural topic models via optimal transport and contrastive learning. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 13894–13920, Vienna, Austria. Association for Computational Linguistics.

Boyu Wang, Linhai Zhang, Deyu Zhou, Yi Cao, and Jiandong Ding. 2023. Neural topic modeling based on cycle adversarial training and contrastive learning. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 9720–9731, Toronto, Canada. Association for Computational Linguistics.

Dongsheng Wang, Dandan Guo, He Zhao, Huangjie Zheng, Korawat Tanwisuth, Bo Chen, and Mingyuan Zhou. 2022. Representing mixtures of word embeddings with mixtures of topic embeddings. *Preprint*, arXiv:2203.01570.

Rui Wang, Xing Liu, Yanan Wang, Shuyu Chang, Yuanzhi Yao, and Haiping Huang. 2025. Mining Topics towards ChatGPT Using a Disentangled Contextualized-neural Topic Model. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, pages 539–548, Hannover Germany. ACM.

Xiaobao Wu, Thong Nguyen, and Anh Tuan Luu. 2024a. A survey on neural topic models: Methods, applications, and challenges. *Artificial Intelligence Review*, 57(2):18.

Xiaobao Wu, Thong Nguyen, Delvin Zhang, William Yang Wang, and Anh Tuan Luu. 2024b. Fastopic: Pretrained transformer is a fast, adaptive, stable, and transferable topic model. *Advances in Neural Information Processing Systems*, 37:84447–84481.

Xiaobao Wu, Fengjun Pan, and Anh Tuan Luu. 2024c. Towards the TopMost: A Topic Modeling System Toolkit. *Preprint*, arXiv:2309.06908.

Xiaobao Wu, Fengjun Pan, Thong Nguyen, Yichao Feng, Chaoqun Liu, Cong-Duy Nguyen, and Anh Tuan Luu. 2024d. On the affinity, rationality, and diversity of hierarchical topic modeling. *Preprint*, arXiv:2401.14113.

Weijie Xu, Wenxiang Hu, Fanyou Wu, and Srinivasan Sengamedu. 2023a. DeTiME: Diffusion-Enhanced Topic Modeling using Encoder-decoder based LLM. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 9040–9057, Singapore. Association for Computational Linguistics.

Weijie Xu, Xiaoyu Jiang, Srinivasan Sengamedu Hanumantha Rao, Francis Iannacci, and Jinjin Zhao. 2023b. vontss: vmf based semi-supervised neural topic modeling with optimal transport. In *Findings of the Association for Computational Linguistics: ACL 2023*, page 4433–4457. Association for Computational Linguistics.

Xuan Xu, Haolun Li, Zhongliang Yang, Beilin Chu, Jia Song, Moxuan Xu, and Linna Zhou. 2025a. Topic modeling as long-form generation: Can long-context llms revolutionize ntm via zero-shot prompting? *Preprint*, arXiv:2510.03174.

Xuan Xu, Fufang Wen, Beilin Chu, Zhibing Fu, Qinhong Lin, Jiaqi Liu, Binjie Fei, Yu Li, Linna Zhou, and Zhongliang Yang. 2025b. Finbert2: A specialized bidirectional encoder for bridging the gap in finance-specific deployment of large language models. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, page 5117–5128, New York, NY, USA. Association for Computing Machinery.

Yishi Xu, Jianqiao Sun, Yudi Su, Xinyang Liu, Zhibin Duan, Bo Chen, and Mingyuan Zhou. 2023c. Context-guided embedding adaptation for effective topic modeling in low-resource regimes. *Advances in Neural Information Processing Systems*, 36:79959–79979.

Yishi Xu, Dongsheng Wang, Bo Chen, Ruiying Lu, Zhibin Duan, and Mingyuan Zhou. 2022. Hyperminer: Topic taxonomy mining with hyperbolic embedding. *Preprint*, arXiv:2210.10625.

Xiaohao Yang, He Zhao, Dinh Phung, Wray Buntine, and Lan Du. 2025a. LLM Reading Tea Leaves: Automatically Evaluating Topic Models with Large Language Models. *Transactions of the Association for Computational Linguistics*, 13:357–375.

Xiaohao Yang, He Zhao, Weijie Xu, Yuanyuan Qi, Jueqing Lu, Dinh Phung, and Lan Du. 2025b. Neural Topic Modeling with Large Language Models in the Loop. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1377–1401, Vienna, Austria. Association for Computational Linguistics.

Lele Yut, Ce Zhang, Yingxia Shao, and Bin Cui. 2017. Lda*: a robust and large-scale topic modeling system. *Proc. VLDB Endow.*, 10(11):1406–1417.

Delvin Ce Zhang and Hady W. Lauw. 2024. Topic modeling on document networks with dirichlet optimal transport barycenter. *IEEE Transactions on Knowledge and Data Engineering*, 36(3):1328–1340.

He Zhao, Dinh Phung, Viet Huynh, Trung Le, and Wray Buntine. 2020a. Neural topic model via optimal transport. *arXiv preprint arXiv:2008.13537*.

He Zhao, Dinh Q. Phung, Viet Huynh, Trung Le, and Wray L. Buntine. 2020b. Neural topic model via optimal transport. *arXiv: Information Retrieval*.

A E1 Information About Use Of Ai Assistants

We use GPT-based assistance within scope below:

1. Draft refinement: grammar, clarity, and conciseness edits on author-written text.
2. Brainstorming: suggested taxonomy wording variants and potential limitation statements;
3. Appendix preparation: helped structure and summarize selected papers in a standardized tabular/sectioned format in the appendix; inclusion/exclusion decisions and final summaries were verified and revised by the authors.
4. Other trivialities such as typos checking and latex formatting.

The AI assistant was not used to fabricate results, run experiments, annotate data, conduct quantitative analysis, or write the related work without human verification. All AI generated content is verified by authors.

B Appendix

B.1 Summaries of Classical Topic Models and Recent Advances

Table 2: Summaries of Topic Models based on Latent Dirichlet Allocation

Method Name	Dataset	Evaluation Metric	Methods	Major Contribution
LDA (Blei et al., 2003)	20Newsgroups, TREC, NIPS Corpus	Perplexity, Topic Coherence, Classification Accuracy	1. Proposes the **LDA** model, a probabilistic generative model that represents each document as a mixture of latent topics. 2. Defines each topic as a multinomial distribution over words, governed by Dirichlet priors α and β. 3. Employs variational inference and collapsed Gibbs sampling for approximate posterior estimation.	1. Establishes the theoretical foundation for probabilistic topic modeling and remains the benchmark for later variants. 2. Enables interpretable topic discovery and dimensionality reduction in large text corpora. 3. Inspires numerous extensions, including hierarchical (hLDA), supervised (sLDA), and scalable (LDA*) models.
LDA* (Yut et al., 2017)	Massive text corpora (e.g., Wikipedia, ClueWeb)	Inference Efficiency, Scalability, Topic Coherence	1. Extends classical Latent Dirichlet Allocation (LDA) to a **robust distributed system** for large-scale probabilistic inference. 2. Implements data and model parallelism to handle billions of tokens efficiently. 3. Optimizes memory allocation and synchronization for distributed Gibbs sampling.	1. Enables efficient, fault-tolerant topic modeling on industrial-scale corpora. 2. Addresses scalability bottlenecks in traditional LDA while maintaining topic quality and stability across massive datasets.
Hierarchical LDA (hLDA) (Chen et al., 2018)	20Newsgroups, Wikipedia, News Datasets	Perplexity, Topic Coherence, Tree Depth Quality	1. Models hierarchical topic structures using the **nested Chinese Restaurant Process (nCRP)** to capture multiple abstraction levels. 2. Represents documents as paths along topic trees, allowing parent–child semantic relationships. 3. Employs collapsed Gibbs sampling for level and path inference.	1. Provides a probabilistic foundation for hierarchical topic modeling, enabling interpretable multi-level topic structures. 2. Serves as the basis for scalable variants such as hLDA-c and distributed HTMs, though limited by sampling efficiency and local optima.

Table 3: Summaries of Embedding-based Topic Models

Method Name	Dataset / Domain	Evaluation Metrics	Core Methodology	Major Contribution
EmTM (Rashid et al., 2023c)	Web Snippets (10k), News Titles (12k)	Classification Accuracy, PMI-based Topic Coherence	1. Uses Word2Vec embeddings to represent words in dense vector space. 2. Applies PCA for dimensionality reduction and HDBSCAN for hierarchical clustering. 3. Constructs probabilistic document–topic and word–topic associations post-clustering.	1. Addresses data sparsity in short-text topic modeling through semantic embeddings. 2. Outperforms LDA, SATM, and PDMM on both coherence and accuracy. 3. Demonstrates scalability and interpretability in real-world short-text datasets.
WeTe (Wang et al., 2022)	20NG, DBpedia, Web Snippets	Topic Coherence, Diversity, Specificity (TS), Clustering Purity, NMI	1. Models documents as mixtures of word embeddings and topics as embeddings in the same space. 2. Learns topic distributions via Conditional Transport (CT) loss combining semantic distance and topic proportion. 3. Supports pretrained or learned embeddings (e.g., GloVe).	1. Reinterprets topic modeling as a distribution alignment problem in embedding space. 2. Achieves state-of-the-art topic and clustering quality on short texts. 3. Avoids approximate inference, enabling efficient large-scale optimization.
Meta-CETM (Xu et al., 2023c)	20NG, Yahoo Answers, Amazon Reviews	Perplexity (PPL), Topic Coherence, Topic Diversity, Few-shot Classification Accuracy	1. Constructs task-specific semantic graphs using dependency parsing. 2. Employs Variational Graph Autoencoder (VGAE) to generate context-adaptive word embeddings. 3. Introduces Gaussian Mixture prior in latent space for clustering-based topic inference.	1. First to integrate semantic graph encoding and GMM priors for few-shot topic modeling. 2. Outperforms baselines (LDA, ETM, CombinedTM) in low-resource scenarios. 3. Demonstrates strong adaptability and interpretability across heterogeneous tasks.
HyperMiner (Xu et al., 2022)	20NG, TMN, Wiki, RCV2	Topic Coherence (NPMI), Topic Diversity, Clustering Purity, NMI	1. Employs hyperbolic (Poincaré/Lorentz) embeddings to capture hierarchical semantics. 2. Models topic–word relations via hyperbolic distance and integrates contrastive learning with WordNet priors. 3. Jointly optimizes ELBO and hierarchical contrastive loss.	1. Introduces hyperbolic geometry to model tree-like topic hierarchies. 2. Incorporates external knowledge graphs for structure-preserving supervision. 3. Achieves superior coherence and diversity while reducing embedding dimensionality.
C-Top2Vec (Angelov and Inkpen, 2024b)	20 Newsgroups, Yahoo Answers	CNPMI, CSBERT, AMI, ARI	1. Uses SBERT embeddings with sliding window for multi-vector document representation. 2. UMAP + HDBSCAN for topic discovery. 3. Phrase-based topic labeling; BERTScore evaluation.	1. Supports document-level topic segmentation. 2. Short-phrase topic labels; improved coherence and representativeness. 3. Efficient runtime.
BERTopic (Grootendorst, 2022)	20 NewsGroups, BBC News, Trump Tweets, UNGA	NPMI, Topic Diversity, Runtime	1. Sentence-BERT embeddings + UMAP + HDBSCAN clustering. 2. Class-based TF-IDF (c-TF-IDF) for topic representation. 3. Supports dynamic topic modeling.	1. Flexible embedding usage; robust across datasets. 2. High topic coherence and diversity; supports evolving topics.
MTM (S et al., 2025)	20NG, BBC News, AG News, NYT, DBpedia	NPMI, Topic Diversity, NMI	1. Multi-view embeddings (Word2Vec, GloVe, BERT, RoBERTa, XLNet). 2. Optimal Transport regularization for topic-word alignment. 3. Dynamic attention to weight embeddings.	1. Mitigates embedding dependency; produces stable document-topic separation. 2. Improves interpretability and topic coherence.

Table 4: Summaries of Recent Advances of VAE-based Topic Models following the Optimization Direction of Prior Distribution and Contrastive Learning

Method Name	Dataset	Evaluation Metric	Methods	Major Contribution
S2WTM (Adhya and Sanyal, 2025)	20Newsgroups, Reuters-21578, NYTimes	Perplexity, Topic Coherence, NPMI	1. Introduces **Spherical Sliced-Wasserstein Autoencoder** for topic modeling. 2. Replaces KL divergence with SSW distance to align aggregated posterior with prior. 3. Employs spherical latent space with vMF and MvMF priors to prevent posterior collapse.	1. Effectively alleviates posterior collapse in VAE-based topic models. 2. Improves coherence and diversity across multiple datasets. 3. Achieves higher NPMI and lower perplexity than standard Neural Topic Models (NTMs).
NSEM-GMHTM (Chen et al., 2023)	WikiText-103, NYTimes, 20Newsgroups	Topic Coherence, Clustering Accuracy, Perplexity	1. Proposes **Neural Semantic Entropy Maximization** for hierarchical topic modeling. 2. Combines Gaussian Mixture priors with hierarchical latent structure. 3. Uses entropy regularization to enhance topic distinctiveness.	1. Achieves better hierarchical interpretability with fine-grained subtopics. 2. Reduces redundancy and improves semantic separation between topics. 3. Demonstrates strong scalability on large corpora.
Diversity-Aware Coherence Loss (DCL) (Li et al., 2023)	20Newsgroups, Reuters, WikiText	Topic Coherence (NPMI), Topic Diversity	1. Introduces a **diversity-aware coherence loss** for neural topic models. 2. Jointly optimizes topic coherence and diversity via contrastive learning. 3. Applies differentiable mutual exclusivity regularization among topics.	1. Significantly enhances diversity without sacrificing coherence. 2. Avoids mode collapse in neural topic generation. 3. Outperforms ETM and ProdLDA in coherence-diversity tradeoff.
NeuroMax (Pham et al., 2024b)	ArXiv, DBLP, ACL Anthology	Perplexity, NPMI, Reconstruction Error	1. Develops a **maximum-entropy constrained neural topic model**. 2. Uses mutual information maximization to maintain latent informativeness. 3. Employs adaptive reparameterization to prevent over-regularization.	1. Retains fine-grained semantic features while improving model stability. 2. Achieves lower reconstruction error and improved topic diversity. 3. Demonstrates better convergence than VAE-based baselines.
Multi-Objective Contrastive Topic Modeling (Nguyen et al., 2025)	20NG, IMDb, Wiki, AG News	NPMI, Topic Diversity, F1 Score for Classification	1. Introduces **Setwise Contrastive Learning** to capture shared semantic information across document sets. 2. Reformulates topic modeling as a **multi-objective optimization problem** balancing ELBO and contrastive loss. 3. Uses KKT-derived Pareto gradient adjustment to dynamically balance objectives.	1. Improves topic coherence and diversity simultaneously. 2. Avoids interference from low-level document statistics. 3. Demonstrates higher F1 in downstream classification compared to baselines.
NTM-ACL (Wang et al., 2023)	NYT, GRL, DBP, 20NG	C_A (Average Topic Consistency), C_P (Pointwise PMI), NPMI	1. Combines **Cycle-Adversarial Training** with contrastive learning to enhance theme-word generation. 2. Introduces **RMR (Reconstruct Minimal Replacement)** data augmentation for topic distributions. 3. Designs dual contrastive losses: self-supervised (generator) and discriminative (classifier).	1. Overcomes VAE limitations by applying contrastive learning to generative process. 2. Produces higher topic consistency and improved topic-word alignment. 3. Demonstrates superior performance when combining adversarial cycle training with contrastive learning.
ContraTopic (Gao et al., 2024)	20NG, Yahoo Answers, NYTimes	Topic Coherence (Avg NPMI), Topic Diversity, Purity & NMI for Document Clustering	1. Introduces a **topic-wise contrastive learning** framework for enhancing interpretability. 2. Uses Gumbel-Softmax sampling to efficiently generate positive/negative word pairs for contrastive regularization. 3. Optimizes a combined loss: reconstruction + KL divergence + contrastive regularization.	1. Enhances both intra-topic coherence and inter-topic distinction. 2. Generates semantically coherent and diverse topics outperforming multiple baselines. 3. Provides thorough ablation studies validating contribution of contrastive components.

Table 5: Summaries of Recent Advances of NTM following the Optimization Direction of Optimal Transport

Method Name	Dataset	Evaluation Metric	Methods	Major Contribution
NSTM (Zhao et al., 2020a)	20NG, WS, TMN, Reuters, RCV2	Topic Coherence (TC, NPMI), Topic Diversity, Purity, NMI	1. Represents documents with both word distributions and dense topic embeddings. 2. Minimizes Optimal Transport (OT) distance between word and topic distributions. 3. Uses Sinkhorn algorithm for efficient OT computation.	1. Introduces OT framework for topic modeling, improving topic coherence/diversity. 2. Embedding-based approach enhances semantic modeling, especially for short texts. 3. Simplifies training compared to VAE-based NTMs.
ANTM (Rahimi et al., 2024)	20NG, IMDB, Yahoo	Purity, NMI, ARI, Label-Topic Alignment	1. Chain graphical model: $y \rightarrow z \rightarrow x$, label-conditioned prior p(zly). 2. Soft indicator and variational inference ensure label-specific topic activation. 3. Label-topic relations modeled as entropy-regularized OT transport plan with pseudo-documents.	1. Explicitly aligns discovered topics with labels. 2. Uses OT in embedding space to capture structured label-topic relations. 3. Enhances clustering and interpretability in supervised scenarios.
EnCOT (Vu et al., 2025)	GoogleNews, SearchSnippets, StackOverflow, Biomedical	Topic Coherence, Topic Diversity, Purity, NMI	1. Applies OT to align documents with cluster centers (LOTDG) and topics with cluster centers (LOTTG). 2. Integrates with GloCOM embeddings and global clustering for representation enhancement.	1. Solves short text sparsity and topic/document separation issues. 2. OT-based dual alignment improves clustering and topic quality. 3. Demonstrates robustness to hyperparameters and large topic numbers.
FASTopic (Wu et al., 2024b)	20NG, WoS, NYT	Topic Coherence, Topic Diversity, Purity, NMI, Classification Accuracy	1. Dual Semantic Relation (DSR) framework reconstructs document-topic-word relations. 2. Embedding Transport Plan (ETP) ensures correct document-topic and topic-word assignments. 3. Pretrained Transformer for document embeddings, Sinkhorn for ETP optimization.	1. Achieves high efficiency and stability in topic modeling. 2. Reduces bias in topic embeddings and improves downstream clustering/classification. 3. Supports transfer across datasets and scales to large corpora.
TraCo (Wu et al., 2024d)	NeurIPS, Arxiv, 20NG, ACM, DBLP	Topic Coherence, Topic Diversity, Parent-Child Correlation (PCC), Sibling Diversity (SD)	1. Transport Plan Dependency (TPD) models hierarchical dependencies via sparse OT plan. 2. Context-aware decoders (CDD) ensure semantic granularity separation across layers.	1. First to apply OT theory to hierarchical topic modeling. 2. Improves parent-child affinity, sibling diversity, and hierarchical rationality. 3. Enhances downstream text classification and retrieval performance.

Table 6: Summaries of Recent NTM Advances with Integrated Methods

Method Name	Dataset	Evaluation Metric	Methods	Major Contribution
DBN / D²BN(Zhang and Lauw, 2024)	Cora (DS, ML, PL), Aminer, Web	Micro/Macro F1, NMI, AUC, Topic Coherence, Topic Diversity	1. Combines GNN embeddings with OT-based topic modeling. 2. Uses Dirichlet prior via rejection sampling as OT prior. 3. D²BN employs double OT barycenter to model semantic and structural spaces. 4. Loss combines OT distance, log-likelihood, KL divergence.	1. Unified framework modeling document content and network structure. 2. OT barycenter approach preserves semantic interpretability and network info. 3. Dirichlet OT prior improves topic quality and diversity. 4. Achieves SOTA in classification, clustering, link prediction, and topic metrics.
vONTSS(Xu et al., 2023b)	AgNews, 20News, DBLP	Km-Purity, Km-NMI, Top-Purity, Top-NMI	1. Uses von Mises-Fisher (vMF) distribution instead of Gaussian as latent prior. 2. Introduces temperature function and learnable concentration parameter . 3. Semi-supervised OT loss (LOT) aligns keywords with topics. 4. Two-stage training strategy for stability.	1. vMF prior alleviates latent space collapse and enhances clustering. 2. OT loss ensures keywords semantically match topics; equivalent to cross-entropy at global optimum. 3. Achieves high accuracy and diversity in semi-supervised settings. 4. Demonstrates stable and efficient performance across datasets.
HiCOT(Vuong et al., 2025)	20NG, AGNews, IMDB, SearchSnippets, GoogleNews	Topic Coherence (CV15), Topic Diversity (TD15), Purity, NMI	1. Document embeddings projected to topic space via MLP and OT alignment. 2. Hierarchical Agglomerative Clustering (HAC) of topics. 3. Contrastive learning: intra-cluster (CLT) and inter-cluster (CLC) losses. 4. Two-stage training: optimize base TM + OT loss, then hierarchical clustering + contrastive loss.	1. Integrates OT, hierarchical clustering, and contrastive learning for enhanced topic quality. 2. Improves document-topic alignment and clustering metrics. 3. Efficient inference, reducing reliance on large PLMs. 4. Produces stable and interpretable topic embeddings, validated via visualization and LLM evaluation.

Table 7: Summaries of Other Adaptations for Special Scenarios

Method Name	Dataset	Target Challenge	Methods	Major Contribution
kNNTM (Lin et al., 2024)	20Newsgroups, AG News, TweetEval	Low-resource / Short-text sparsity	1. Augments each document with nearest neighbors to expand contextual representation. 2. Incorporates neighbor information into the ELBO reconstruction objective. 3. Stabilizes VAE training and enhances topic consistency under sparse supervision.	1. Effectively mitigates label sparsity and data insufficiency. 2. Achieves higher topic coherence on short-text datasets. 3. Provides a general strategy for low-resource topic modeling.
DSNTM (Dynamic Self-Attention Neural Topic Model) (Miyamoto et al., 2023)	ACL Anthology, ArXiv Abstracts, DBLP Corpus	Temporal topic evolution	1. Introduces self-attention into the VAE framework to capture temporal dependencies. 2. Uses citation-based regularization to align evolving topics over time. 3. Models inter-temporal dynamics among latent topics.	1. Captures topic transitions more accurately in evolving corpora. 2. Improves continuity and interpretability of temporal topic evolution. 3. Extends VAE-based topic models to longitudinal data.
NSEM-GMHTM (Chen et al., 2023)	Wikipedia, Scientific Abstracts, News Commentary	Hierarchical topic structure	1. Extends the prior distribution to a Gaussian Mixture model. 2. Integrates nonlinear structural equation modeling (SEM) into the topic hierarchy. 3. Learns both symmetric and hierarchical dependencies among latent topics.	1. Enhances representation of multi-level and non-linear topic relations. 2. Improves coherence and diversity under sparse corpus conditions. 3. Demonstrates interpretability for hierarchical thematic structures.
BERTDetect (Ranaweera et al., 2025)	Drebin, AndroZoo, MalGenome (Android Malware)	Cross-domain application (Security)	1. Adapts VAE-style neural topic modeling for malware detection. 2. Utilizes contextualized BERT embeddings for semantic representation. 3. Detects latent threat-related topics in code and text features.	1. Bridges topic modeling with cybersecurity applications. 2. Demonstrates domain adaptability of VAE-based topic models. 3. Improves interpretability and detection performance in malware datasets.
CTM-MM (Rashid et al., 2023a)	Twitter, Instagram, YouTube Captions	Multimodal topic modeling	1. Extends contextualized topic modeling (CTM) to multimodal inputs (text + images). 2. Uses shared latent spaces for multimodal alignment within the VAE framework. 3. Jointly optimizes textual and visual semantics for topic inference.	1. Enhances topic coherence across heterogeneous modalities. 2. Enables joint analysis of text–image corpora. 3. Demonstrates flexibility of neural topic modeling beyond text-only inputs.

B.2 Summaries of LLM-Assisted Topic Models

Table 8: LLM-Assisted Topic Modeling Frameworks for Data Preprocessing

Method Name	Dataset	Evaluation Metric	Methods	Major Contribution
LLM-Centric Context Expansion + Prefix-tuned VAE (Akash and Chang, 2024)	Short-text datasets (e.g., StackOverflow, Tweets, News Titles)	Topic Coherence, Diversity, NPMI, Perplexity	1. Proposes a hybrid framework that integrates **LLM-based context expansion** with a **prefix-tuned variational autoencoder** (VAE). 2. Uses LLMs to enrich short-text inputs by generating semantic context before VAE encoding. 3. Employs prefix tuning to adapt pretrained VAEs without full fine-tuning.	1. Significantly enhances topic coherence for short texts by mitigating sparsity. 2. Demonstrates that context expansion effectively bridges LLM knowledge with probabilistic topic models. 3. Provides an interpretable, resource-efficient alternative to large-scale retraining.
LimTopic (Azher et al., 2025)	20 Newsgroups, AG News, Reddit, ArXiv Abstracts	Topic Coherence, Topic Diversity, Clustering Accuracy	1. Introduces **LimTopic**, an LLM-based topic modeling framework with **semantic limiting mechanisms**. 2. Uses LLMs to propose candidate topic words and filters them via semantic similarity constraints. 3. Clusters document embeddings based on semantic-aware topic anchors.	1. Balances coherence and diversity by constraining topic boundaries using LLM-derived semantics. 2. Addresses redundancy and overlap in traditional unsupervised topic extraction. 3. Provides controllable topic generation with minimal supervision.
LiSA: LLM-guided Semantic-Aware Clustering Framework (Liu et al., 2025)	Wikipedia Subset, News Articles, Research Abstracts	Silhouette Score, Topic Coherence, Human Evaluation	1. Proposes a **semi-supervised topic modeling framework** that integrates LLM-generated topic words with semantic-aware clustering. 2. LLMs generate candidate topic terms and weak labels, guiding unsupervised clustering models. 3. Incorporates human feedback through selective refinement loops.	1. Achieves fine-grained topic separation by aligning LLM-generated semantics with cluster formation. 2. Bridges the gap between zero-shot LLM inference and structured clustering. 3. Demonstrates robust performance under both labeled and unlabeled corpus settings.

Table 9: LLM-Embedding-Enhanced Topic Modeling Frameworks

Method Name	Dataset	Evaluation Metric	Methods	Major Contribution
DeTiME (Xu et al., 2023a)	20Newsgroups, AG News, ArXiv Abstracts	Topic Coherence, Clustering Accuracy, Diversity	1. Proposes **DeTiME**, integrating **encoder–decoder LLM embeddings** with a diffusion-based augmentation process. 2. Enhances topic clustering by generating smooth latent transitions between document embeddings. 3. Supports topic-related content generation through learned diffusion dynamics.	1. Addresses suboptimal performance of neural topic models (NTMs) by improving embedding structure and separability. 2. Demonstrates better cluster quality and interpretability across diverse corpora. 3. Bridges embedding learning and generative modeling for topic discovery.
DisCTM (Wang et al., 2025)	Short-text corpora (Tweets, Reddit posts, News headlines)	Perplexity, Topic Coherence, Sparsity Ratio	1. Introduces a **Transformer-based disentanglement mechanism** to separate semantic and topical information. 2. Uses LLM-derived embeddings as priors to initialize Transformer encoders. 3. Employs contrastive learning to enforce topic diversity and sparsity.	1. Mitigates topic overlap and improves interpretability on sparse short-text datasets. 2. Achieves domain adaptation by incorporating context-aware fine-tuned embeddings. 3. Outperforms traditional NTMs and embedding-only baselines under low-resource conditions.
Fin-Topicmodel (Fin-BERT2) (Xu et al., 2025b)	Financial news, SEC filings, Earnings reports	Topic Coherence, Domain Relevance, Cluster Stability	1. Builds **Fin-Topicmodel** upon the **Fin-BERT2** architecture with domain-specific embedding fine-tuning. 2. Uses fine-tuned LLM embeddings as semantic inputs to improve financial topic representation. 3. Combines clustering with hierarchical topic extraction to support financial analysis tasks.	1. Provides domain-adaptive topic modeling with enhanced interpretability for finance. 2. Demonstrates that LLM fine-tuning significantly boosts topic quality compared with generic embeddings. 3. Enables application of topic modeling to downstream finance-oriented analytics and event prediction.

Table 10: LLM-Assisted Post-Processing for Neural Topic Models

Method Name	Dataset	Evaluation Metric	Methods	Major Contribution
BERTopic with LLM Refinement (Grootendorst, 2022)	20Newsgroups, Twitter, Financial Reports, Wikipedia	Topic Coherence (CV), Diversity, Readability	1. Extends the **BERTopic** framework with optional **LLM-based post-processing** to refine topic word lists. 2. Allows users to define custom prompts and call LLM APIs for rewriting and relabeling topics. 3. Integrates semantic embeddings (BERT) with dynamic topic merging and labeling.	1. Improves coherence and interpretability of discovered topics without retraining models. 2. Demonstrates flexible integration of LLMs into existing neural topic modeling pipelines. 3. Serves as a widely adopted toolkit for hybrid embedding–LLM-based topic modeling.
AgenTopic (pariskang, 2025)	Semantic Scholar, ArXiv, PubMed	Topic Coherence, Redundancy Reduction, Interpretability	1. Combines GPT-4 with caching modules for iterative topic refinement. 2. Employs LLMs to post-process topic word lists—rewriting, merging, and correcting incoherent clusters. 3. Integrates citation graphs and metadata to enhance scientific topic linking.	1. Reduces redundancy and improves interpretability in large-scale academic corpora. 2. Demonstrates that LLM-driven post-processing can complement symbolic or embedding-based NTMs. 3. Provides an adaptive topic exploration system for literature analysis.
LLM-ITL (Yang et al., 2025b)	20Newsgroups, AG News, Reddit, Scientific Abstracts	Topic Coherence, Diversity, Alignment Loss	1. Introduces a **LLM-in-the-loop (LLM-ITL)** framework that combines NTMs with LLM refinement. 2. NTMs generate global topic structures, while LLMs iteratively adjust topic words via OT-based alignment. 3. The alignment objective dynamically adapts according to LLM confidence.	1. Enhances topic coherence and diversity while maintaining statistical grounding. 2. Bridges unsupervised NTMs with guided semantic alignment via LLM feedback. 3. Offers a flexible interface for integrating reasoning-based refinement into neural topic models.

Table 11: LLM-Based Evaluation and Analysis in Topic Modeling

Method Name	Dataset	Evaluation Metric	Methods	Major Contribution
WALM (Yang et al., 2025a)	20Newsgroups, ArXiv Abstracts, News Datasets	Word Agreement Score, Coherence, Alignment with Human Evaluation	1. Proposes WALM, an LLM-based evaluation framework assessing document–topic and word–topic alignment. 2. Computes word agreement by comparing model-generated topics with LLM contextual likelihoods. 3. Integrates both intrinsic and extrinsic evaluation components.	1. Achieves high alignment with human evaluation, complementing existing automated measures. 2. Provides a unified approach to assessing both topic coherence and representational quality. 3. Bridges semantic understanding and statistical consistency in evaluation.
ProxAnn (Hoyle et al., 2025)	20Newsgroups, Reddit, Wikipedia	Proxy Annotation Agreement, Statistical Indistinguishability, Human-Like Evaluation	1. Develops ProxAnn, a framework where LLM-based proxy annotations emulate human evaluation. 2. Uses statistical tests to verify indistinguishability between LLM and human judgments. 3. Evaluates topic quality and use-oriented measures via proxy scores.	1. Validates that LLM-based evaluations can replace human annotations with high reliability. 2. Provides scalable, consistent topic evaluation without costly human labeling. 3. Lays groundwork for automatic benchmarking pipelines using LLM proxies.

B.3 Summaries of Prompts in LLM-Centric Topic Model

Stage	Core instruction description (keep placeholders in)	Example output
Basic Prompt (Round 1)	Read the text below and list up to 3 topics. Each topic should contain fewer than 3 words. Ensure you only return the topic and nothing more. The desired output format: "Topic 1: xxx, Topic 2: xxx, Topic 3: xxx". {text}	Topic 1: climate policy Topic 2: carbon pricing Topic 3: emission targets
Basic Prompt + Seeds Topic (Another choice of Round 1)	Consider the previous extracted topics: {existing_topics}. Read the text below and list up to 3 topics. Each topic should contain fewer than 3 words. Ensure you only return the topic and nothing more. The desired output format: "Topic 1: xxx, Topic 2: xxx, Topic 3: xxx". {text}	Topic 1: vaccine uptake Topic 2: adverse events Topic 3: trial design
Prompt for Summarization (merge topics) (Round 1)	Summarize and merge the following list of topics into {Fixed_Number} final topics. {List_of_Topics}	Final Topic 1: public health policy Final Topic 2: vaccine safety Final Topic 3: clinical trials

Table 12: Structured Summaries of paper (Mu et al., 2024)

Stage	Core instruction description (keep placeholders in)	Example output
Simulate topic modeling on subsets (ParTM Round 1)	Write the results of simulating topic modeling for the following documents, each starting with "#". Assume you will identify K topics and use M top words for each topic. NOTE: Outputs must always be in the format "Topic k: word word word word word" and nothing else. Input is a list of short documents or headlines, one per line, each prefixed by "#".	Topic 1: word1 word2 word3 word4 word5 Topic 2: word1 word2 word3 word4 word5 ... Topic K: word1 word2 word3 word4 word5
Merge topic-modeling results across subsets (ParMg Round 2)	We aim to identify topics for the entire document set by merging the topic modeling results for each subset. Write the results of merging the following topic modeling results for each subset of the document set. Each subset result starts with "- n" and its topics start with "#". Assume you will finally identify K topics and use M top words for each topic. Outputs should reflect the topics before merging as much as possible; include topics that often appear before merging and avoid ones that do not. NOTE: Outputs must be in the format "Topic k: word word word word word" only.	- 1 # topic words from subset 1 # topic words from subset 1 ... - 11 # topic words from subset 11 ... Merged result: Topic 1: word1 word2 word3 word4 word5 ... Topic K: word1 word2 word3 word4 word5
Aother Method SeqTM: sequentially update topics using previously identified topics (Round 1)	We aim to identify topics for the entire document set by sequentially updating tentative topics identified from each subset, considering topics identified just before from another subset. Write the results of simulating topic modeling for the following documents, each starting with "#". Make the most use of the following topics previously identified from another set of documents, each starting with "Topic k:". Assume you will finally identify K topics and use M top words for each topic; outputs should be the same as the previous topics as much as possible, changing minimally when the given documents do not include them much or when a new topic is needed. NOTE: Outputs must be in the format "Topic k: word word word word word" and nothing else.	Previously identified: Topic 1: seed1 seed2 seed3 seed4 seed5 ... Topic K: seed1 seed2 seed3 seed4 seed5 New documents (each line starts with "#"): # doc line ... # doc line ... Sequential result: Topic 1: word1 word2 word3 word4 word5 ... Topic K: word1 word2 word3 word4 word5

Table 13: Structured Summaries of prompts of paralleled and sequential methods in paper (Doi et al., 2024).

Stage	Core instruction description (keep placeholders in)	Example output
Dataset Preprocessing and Sampling (non-LLM operation) (Step 1)	Operate topic modeling in a semantic space with appropriate abstraction. Preprocess input documents via extraction or summarization to obtain titles/abstracts that meet text-unit length limits. Sample an appropriate number of text units according to the model's context window.	
Topics Generation (Step 2)	Please conduct thematic analysis on the provided text data to generate independent topics balancing generalization and specificity. IMPORTANT: For "Source Titles", ONLY copy exact titles from input data (lines like "Title: [actual title]"). Output pure JSON with the structure: ["Topic i": "Summary": "One-sentence topic summary", "Keywords": ["keyword1", ...], "Source Titles": ["Exact title 1", ...]]. Core requirements: at least 3 topics; 5–12 keywords per topic; 3–8 exact source titles per topic; semantic coherence; minimized repetition; no duplicated titles within the same topic.	["Topic 1": "Summary": "Short coherent description.", "Keywords": ["k1","k2","k3","k4","k5"], "Source Titles": ["Exact title 1","Exact title 2","Exact title 3"] , "Topic 2": ... , "Topic 3": ...]
Texts Assignment (non-LLM operation) (Step 3)	Assign documents to discovered topics by keyword matching. Each document can map to one or more topics.	

Table 14: Structured Summaries of the Methodology of paper (Xu et al., 2025a)

Stage	Core instruction description (keep placeholders in)	Example output
Candidate creation (Round 1)	Your task is to distill a list of topics from multiple documents: DOCUMENT: {document 1}, DOCUMENT: {document 2}, ..., DOCUMENT: {document N}. Topics should be neither too general nor too specific (e.g., "food" too general; "lemon cake" too specific). A topic does not need to appear in multiple documents. Output a JSON exactly in the format: { 'topics': ['topic1', 'topic2', 'topic3'] }.	{ 'topics': ['topic A', 'topic B', 'topic C'] }
Topic reduction (Round 1)	Your task is to distill a list of core topics from an indexed list: 0: {topic 1}, 1: {topic 2}, ..., NCT-1: {topic NCT}. Remove duplicates, merge topics that are too general, and merge topics that are too specific into an appropriate generalization. Aim to end with about NT topics. Output JSON exactly as: "topics": ["topic1", "topic2", "topic3"].	"topics": ["generalized topic A", "topic B", "topic C"]
Topic reduction iterative (Round 1)	Your task is to merge a pair of topics from the indexed list: 0: {topic 1}, 1: {topic 2}, ..., NCT-1: {topic NCT}. Select the most similar and most granular pair to merge first. Output JSON: { 'topic_pair': [idx1, idx2], 'new_topic': 'new_topic' }. The new topic should be a simple, generalized common denominator—not a concatenation like "A and B".	{ 'topic pair': [3, 7], 'new topic': 'broader topic name' }
Topic assignment (Round 2)	Classify the given document into one of the provided topics. Input: DOCUMENT: {document}. Return only the chosen topic label (or ID, if specified by the task context).	Selected topic: Topic X

Table 15: Structured Summaries of prompts of TopicGen (van Wanrooij and Manhar, 2024).

Stage	Core instruction description (keep placeholders in)	Example output
Generate first-level/flat topics (Round 1)	You will receive a document {Document} and a set of existing top-level topics {Top-level topics}. Identify GENERALIZABLE topics within the document that can act as top-level topics in the hierarchy. If relevant top-level topics are missing from the provided set, add them; otherwise return the existing relevant top-level topics as identified in the document. Topics must be broad, not document-specific, represent a SINGLE topic, and be able to accommodate future subtopics. If the document contains no topic, return "None". Output ONLY relevant or newly added topics at the top level. See example.	[Top-level topics] [1] Topic A: Brief description. [2] Topic B: Brief description. or None
Generate second-level subtopics (Round 2)	You will receive a top-level branch {Topic branch} and several documents {Documents}. Identify GENERALIZABLE second-level topics to act as subtopics under the provided top-level topic. If relevant subtopics already exist or the provided top-level topic is sufficiently specific, return the existing relevant/duplicate topic(s). Each subtopic must reflect a SINGLE coherent topic, be broad enough for future subtopics, and the number of proposed subtopics should not exceed the number of documents. ONLY add at the second level; DO NOT add first- or third-level topics. See example.	[1] TopTopic [2] Subtopic A (Document: 1,3): brief description. [2] Subtopic B (Document: 2): brief description.
Refine/Merge same-level topics (Round 2)	You will receive a list of topics at the same level {Topic List}. Merge topics that are paraphrases or near-duplicates. Perform as many merges as needed; if no modification is needed, return "None". When merging, output the level indicator, the updated label and description, followed by the original topics being merged. Operate ONLY within the same level; do not introduce new levels or concepts. See example.	[2] Technology: Discuss technology and its impact on society. ([2] Digital Literacy, [2] Telecommunications) or None
Assign topics to a document (Round 3)	You will receive a topic hierarchy tree and a document {Document}. Assign the document to the most relevant existing topic(s) in the hierarchy. Output the topic labels, assignment reasoning, and supporting quotes taken directly from the document. DO NOT invent new topics or quotes. If the assigned topic is not on the top level, also output the path from the top level to the assigned topic. See example.	[1] Trade [2] Tariff: Mentions adjusting taxes on mixtures containing Fluopyram ("...suspend temporarily the duty on mixtures containing Fluopyram...")

Table 16: Structured Summaries of prompts of TopicGPT (Pham et al., 2024a).

Stage	Core instruction description (keep placeholders in)	Example output
Claim generation	Title: {title}; Abstract: {abstract}. Task: Conclude new findings and null findings from the abstract in one sentence in atomic format. Do not separate new vs. null findings; the single claim must be relevant to the title and contain no extra information. Definition: A scientific claim is an atomic, verifiable statement about one aspect of a scientific entity or process that can be verified from a single source.	One-sentence atomic claim about the study's finding that is relevant to {title}.
Hierarchy proposal (Round 1)	Review Title: {systematic_review_title}. Frequent entities from study abstracts: {freq_entities}. Study Claim List: {claim_list}. Instruction: 1) Top-Level Aspect Generation: Use entities to identify up to 5 top-level aspects from clinical study claims; list as bullets and cite entities; output as [Response 1]. 2) Hierarchical Faceted Category Generation: For each top-level aspect in [Response 1], generate hierarchical faceted categories aligned with claims; keep sibling granularity similar; avoid unrelated info; cite supporting claims; output as [Response 2]. Remember: be precise; include only relevant data; cite claims as "(Claim i, j, ...)", and output as a numbered nested list.	[Response 1]: Aspect 1: Name (Entity A, Entity B) Aspect 2: Name (Entity C) ... [Response 2]: Aspect 1: Name (Claim 0, 2) 1: Subcategory A (Claim 0) 1.1: Subsubcategory (Claim 2) 2: Subcategory B (Claim 3) Aspect 2: Name (Claim 4, 5) 1: Subcategory ...
Sibling coherence judgment (Round 1)	Instruction: You will assess whether a given set of sibling categories logically belong together under their shared parent category ("coherence"). Label whether ALL siblings are coherent. If all fit well and have similar granularity, reply "[These sibling categories are coherent]"; otherwise reply "[These sibling categories are NOT coherent]". Base decisions solely on coherence. Follow format: provide step-by-step reasoning then the final bracketed label. Inputs: Parent category: {parent_category}; Sibling categories: {sibling_categories}.	Step-by-step reasoning: [analytical justification] Answer: [These sibling categories are coherent] or [These sibling categories are NOT coherent]
Category assignment (Round 2)	Instruction: Assess whether a specific claim belongs to a provided category and assign a binary label. Choose "The claim belongs to the category" if any part/aspect of the claim is relevant (including negations/opposites); otherwise choose "The claim does NOT belong to the category". Only use the requested output format. Provide step-by-step reasoning before the final label. Inputs: Claim: {claim}; Category: {category}.	Step-by-step reasoning: [why the claim is or is not relevant] Answer: [The claim belongs to the category] or [The claim does NOT belong to the category]

Table 17: Structured Summaries of prompts for CHIME (Hsu et al., 2024)