Towards Modern Topic Models: A Survey of Taxonomies and Paradigm Shifts from Algorithm-Centric to LLM-Centered Topic Analysis

Xuan Xu[1,4,*] **Zhongliang Yang**[1,*] **Haolun Li**[1,*] **Rui Tian**[2]
Beilin Chu[1] **Jia Song**[1,4] **Yu Li**[3] **Shaolin Tan**[4] **Linna Zhou**[1,†]
[1] Beijing University of Posts and Telecommunications
[2] University of Science and Technology Beijing
[3] Beijing Value Simplex Technology Co. Ltd
[4] Zhongguancun Laboratory
{sh22xuxuan,yangzl,lhldudu,beilin.chu,songj,zhoulinna}@bupt.edu.cn
u202343368@xs.ustb.edu.cn liyu@entropyreduce.com shaolintan@hnu.edu.cn

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. We review recent TM developments and categorize existing methods into three groups: Classical Algorithm-Centric, LLM-Assisted, and LLM-Centric. For traditional algorithm-centric 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 examine two key transformations brought by LLM-centric TM: expanded task scope and a shift from model-level improvements to system-level engineering. We also propose a future roadmap for more optimized LLM-Centric TMs and identify ongoing critical challenges. We aim for this survey to spur closer integration between TM and LLMs and to further drive the progress of modern TM.[1]

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 algorithm-centric models, as well as those that employ LLMs exclusively as the core of 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 algorithm-centric methods and do not focus on the involvement of LLMs. Beyond recent enhancements to the classical TM algorithm, the convergence of LLM and TM remains underexplored as an independent topic. We reviewed the recent fieldwork and present new taxonomies and our view of this paradigm shift.

Algorithm-Centric TM denotes relatively lightweight methods with clearly defined structures, such as probabilistic graphical models (LDA) and neural variational models (ProdLDA, VAE-style LDA variants). When LLMs are incorporated into a particular step of a conventional topic modeling pipeline to boost performance, but the overall framework is still driven by traditional algorithms, we categorize the approach as LLM-Assisted. In LLM-assisted settings, LLMs inject rich semantic priors into topic modeling across data augmentation, representation, inference, and evaluation. LLM-Centered topic analysis methods treat LLMs as the primary engine, shifting the

*These authors contributed equally.
†Corresponding author.
[1]A list of resources is available at https://github.com/Harrisonls2004/Towards-Modern-Topic-Models

Findings of the Association for Computational Linguistics: ACL 2026, pages 6536–6561

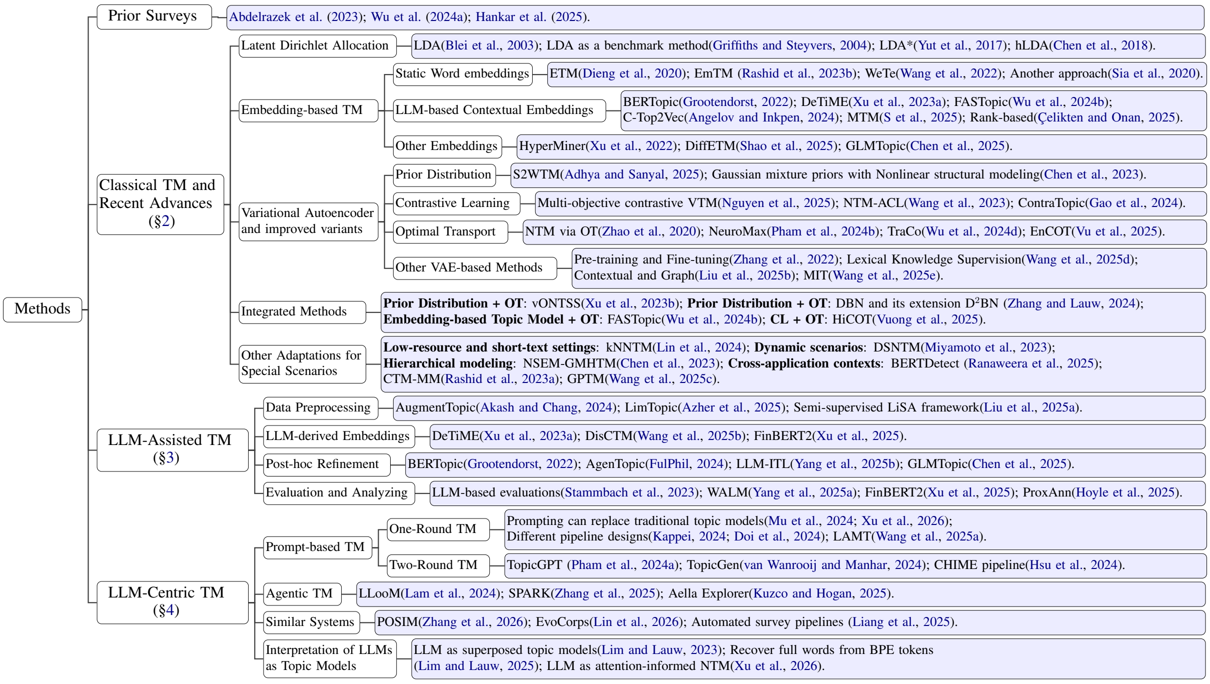

Figure 1: Taxonomy of diverse approaches.

research emphasis to prompt or instruction design, alignment, reasoning capabilities, and integration with external tools.

More importantly, the topic analysis paradigm shift brought about by LLMs, along with their prospective development trajectories and associated challenges, warrants deeper examination and discussion. We also present a rich set of perspectives and analyzes. Our main contributions are summarized as follows:

1. Updated review on TM in the LLM era: We survey TM through the lens of LLM involvement and categorize methods into *Classical Algorithm-Centric*, *LLM-Assisted*, and *LLM-Centric*. For *Classical Algorithm-Centric* 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 and method selection: We articulate the transformative impact of LLM-centric topic modeling, characterized by two key developments: a broadened and more inclusive task scope across multiple dimensions, and a shift from model-level improvements to system-level engineering. We also provide a contemporary comparative analysis for method selection.

4. Roadmap and open challenges for optimized LLM-Centric TM: We present a forward-looking roadmap for engineering more efficient and effective LLM-Centric TMs, and we enumerate critical ongoing challenges to catalyze tighter integration between TM and LLMs and accelerate progress in modern TM.

2 Classical Algorithm-Centric TM and Recent Advances (Appendix B.1)

2.1 Latent Dirichlet Allocation (Tables 4)

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 like LDA* (Yut et al., 2017) enable robust, efficient distributed inference on massive corpora. Hierarchical variants, e.g., hLDA (Chen et al., 2018), use the nested Chinese Restaurant Process to model topic hierarchies and have been scaled to industrial datasets. However, LDA's bag-of-words assumption overlooks long-range dependencies, hurting performance on unstructured/noisy data; as probabilistic generative models, LDA variants also struggle with sparse short texts and rely heavily on manual preprocessing.

2.2 Embedding-based Topic Model (Tables 5)

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. This line of work builds topic models on static embeddings (e.g., Word2Vec, GloVe). ETM (Dieng et al., 2020) jointly embeds texts and topics in a shared latent space, improving semantic similarity over bag-of-words. Clustering-based approaches (Sia et al., 2020) group pre-trained word vectors with document-level cues to re-rank keywords. Subsequent variants enhance this paradigm: EmTM (Rashid et al., 2023b) clusters Word2Vec embeddings to mitigate short-text sparsity and improve interpretability, and WeTe (Wang et al., 2022) models documents as mixtures of word embeddings and topics as distributions of topic embeddings to learn fine-grained topics.

LLM-based Contextual Embeddings. Transformer models (e.g., BERT) provide contextual embeddings that support clustering-based topic discovery. BERTopic (Grootendorst, 2022) combines BERT document embeddings, HDBSCAN, and class-based TF–IDF for interpretable topics. DeTiME (Xu et al., 2023a) augments encoder–decoder LLM embeddings with diffusion to improve clustering and enable topic-aware generation. FASTopic (Wu et al., 2024b) aligns documents with learnable topic/word embeddings via Dual Semantic Reconstruction for efficient discovery. C-Top2Vec (Angelov and Inkpen, 2024) leverages contextual token embeddings for hierarchies, in-document spans, and phrase labels, while MTM (S et al., 2025) uses multi-view contextual embeddings with clustering regularization to improve stability and diversity. Extending this work, Çelikten and Onan (2025) propose an ensemble framework that integrates rank-based aggregation and LLM-driven topic extraction to achieve robust modeling of both AI-generated and human-authored scientific texts.

Other Embeddings. Beyond Euclidean spaces, HyperMiner (Xu et al., 2022) employs hyperbolic embeddings to capture hierarchical semantics in topic taxonomies. DiffETM (Shao et al., 2025) injects document semantics into latent variables via diffusion-regularized enhancement, gradually adding noise that conforms to a normal distribution. GLMTopic (Chen et al., 2025) integrates adaptive community-enhanced graph embeddings to achieve high-level semantic representation.

2.3 Variational Autoencoder (Tables 6-8)

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 employs the Spherical Sliced-Wasserstein distance to better align the aggregated posterior with the prior.

Contrastive Learning. CL has been integrated into VAEs to improve topic coherence, diversity, and downstream performance. NTM-ACL (Wang et al., 2023) injects contrastive signals via cycle adversarial training; ContraTopic (Gao et al., 2024) adds topic-wise contrastive regularization guided by corpus statistics to boost intra-topic coherence and inter-topic diversity. Nguyen et al. (2025) formulates set-oriented CL with gradient-based multi-objective optimization to balance the ELBO and contrastive objectives at Pareto-stationary solutions.

Optimal Transport. OT aligns latent distributions to enhance interpretability in NTMs. The OT-based NTM (Zhao et al., 2020) models document–topic relations more accurately; NeuroMax (Pham et al., 2024b) maximizes mutual information and adds group-level OT regularization to improve topic informativeness and separation. For hierarchies, TraCo (Wu et al., 2024d) uses sparse balanced transport plans to encode parent–child structure while preserving sibling diversity. For short texts, EnCOT (Vu et al., 2025) dual-aligns documents/clusters and topics/clusters to mitigate sparsity and sharpen separation. OT-based objectives have also been explored to help LLMs maintain semantic and structural consistency over long contexts.

Other VAE-based Methods. Several studies under the VAE framework enhance topic modeling by integrating external knowledge and structural information. Zhang et al. (2022) employs a pre-training and fine-tuning strategy to effectively incorporate external knowledge with low computational overhead. Wang et al. (2025d) introduces a distributed keyword-guided topic model that aligns topics with user interests through lexical knowledge supervision. Liu et al. (2025b) improves topic coherence and diversity by incorporating contextual and graph information. Mutual Information Topic Model (MIT) (Wang et al., 2025e) enhances topic diversity by maximizing the mutual information between word and topic distributions.

2.4 Integrated Methods (Tables 9)

Some methods combine multiple strategies. **(Prior + OT)** vONTSS (Xu et al., 2023b) uses a vMF prior with an OT-based semi-supervised loss to align topic–word relations to external knowledge. **(Prior + OT)** DBN and D^{2}BN (Zhang and Lauw, 2024) couple Dirichlet priors with OT barycenters and GNNs, balancing semantic interpretability and structure. **(Embeddings + OT)** FASTopic (Wu et al., 2024b) unifies Transformer-based document embeddings with learnable topic/word embeddings via OT and Dual Semantic Reconstruction for efficient topic discovery. **(CL + OT)** HiCOT (Vuong et al., 2025) integrates OT with hierarchical clustering and contrastive objectives to improve coherence and diversity.

2.5 Other Adaptations for Special Scenarios (Tables 10)

Advances target sparsity, dynamics, hierarchy, and multimodality. In **low-resource/short-text** settings, kNNTM (Lin et al., 2024) augments documents with nearest neighbors to stabilize VAE training and enrich reconstruction. For **dynamic** corpora, DSNTM (Miyamoto et al., 2023) embeds self-attention in a VAE and adds citation-based regularization to track evolving themes. For **hierarchies**, NSEM-GMHTM (Chen et al., 2023) employs a Gaussian-mixture prior with nonlinear structural equations to represent hierarchical and symmetric dependencies. In **cross-application** contexts, BERTDetect (Ranaweera et al., 2025) and CTM-MM (Rashid et al., 2023a) adapt VAE-style topic models to Android malware and multimodal social media, improving coherence under domain constraints.Wang et al. (2025c) introduces a genre-aware personalized NTM for mining user preferences from online reviews.

3 LLM-Assisted TM (Appendix B.2)

LLM-assisted means that the LLM serves an auxiliary role for specific subtasks, while the core inference over topics and document–topic distributions is performed by traditional algorithms such as LDA or NTM.

3.1 Data Preprocessing (Tables 11)

LLMs facilitate corpus preprocessing via summarization, expansion, rewriting/cleaning, and label generation, improving quality and adding weak supervision for topic modeling. Akash and Chang (2024) combine LLM-based context expansion with prefix-tuned VAEs for short texts. More advanced methods—LimTopic (Azher et al., 2025) and LiSA (Liu et al., 2025a)—use LLMs to propose candidate topic words to guide unsupervised clustering models with semantic-aware clustering, addressing prior limitations in capturing fine-grained document semantics.

3.2 LLM-derived Embeddings (Tables 12)

LLM-derived embeddings offer two key advantages over traditional bag-of-words: they provide stronger priors or initializations, and they overlap with embedding-based TM discussed in Section 2.2 (e.g., BERTopic, FASTopic). Further refinements focus on improving efficiency and domain adaptation through advanced architectures, disentanglement, and adaptation techniques. For instance, DeTiME (Xu et al., 2023a) integrates encoder-decoder embeddings with diffusion augmentation to enhance clustering and enable topic-aware generation. DisCTM (Wang et al., 2025b) introduces topic disentanglement for domain-specific short texts, while Fin-Topicmodel in FinBERT2 (Xu et al., 2025) uses fine-tuned financial embeddings for more efficient representations.

3.3 Post-hoc Refinement (Tables 13)

LLMs post-process topic word lists via rewriting, correction, and merging to improve coherence, diversity, and readability. BERTopic (Grootendorst, 2022) already supports custom prompts for regenerating topics. Subsequent work, GLM-Topic (Chen et al., 2025), leverages LLM-powered representation tuning to refine clustered docs into more contextually relevant and interpretable topics. AgenTopic (FulPhil, 2024) integrates GPT-4 with caching for iterative refinement, reducing redundancy and assignment errors; LLM-ITL (Yang et al., 2025b) couples NTMs (learning global topics) with LLM-based OT alignment that adapts to LLM confidence.

3.4 Evaluation and Analyzing (Tables 14)

LLM-based evaluation correlates well with humans and aids analysis. Stammbach et al. (2023) reports stronger correlations than traditional metrics and provides guidance on topic number. WALM (Yang et al., 2025a) jointly assesses document representations and topic quality; FinBERT2 (Xu et al., 2025) prompts for coherence, conciseness, and informativeness. ProxAnn (Hoyle et al., 2025) shows that LLM proxy annotations can match human labels. Nonetheless, LLMs can produce generic themes and miss domain-specific topics in large corpora (Li et al., 2025); Moreover, limited context windows also induce hallucinations and scalability issues.

4 LLM-Centric TM (Appendix B.3)

LLM-centric means that the LLM serves as the primary inference mechanism for discovering topics, generating topic assignments and topic descriptions via prompting and other LLM reasoning methods. Prompts of each method mentioned in this section are summarized in Appendix B.3.

4.1 Level 1: Prompt-based TM (Tables 15-20)

At this level, core steps such as topic generation and refinement are accomplished with one or two rounds of LLM calls, relying on minimal external tools; in practice, this level translates to prompt/template engineering and simple, shallow pipelines (often a single pass with optional minor refinement).

One-Round Prompt-based TM One-Round Prompt-based TM 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.

Prior work shows that vanilla prompts in one pass inference can replace traditional topic models, yielding human-aligned topics (Mu et al., 2024; Xu et al., 2026), while tailored prompt designs improve coherence/diversity and reduce hallucinations, especially on small or short-text corpora (Doi et al., 2024; Kappei, 2024). However, one-round processing may lead to problems such as a lack of complete topic assignment and incomplete topic coverage. The existing research focuses on short texts or small corpora. Wang et al. (2025a) propose LAMT (LLM Aligned Multi-label Topic Model), a novel approach that introduces LLMs into multi-label topic modeling. LAMT leverages the semantic understanding capabilities of LLMs to achieve precise topic-text segment alignment, while incorporating word embeddings for knowledge-enhanced topic mining.

Two-Round Prompt-based TM The 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 Prompt-based TMs are commonly utilized to handle large corpora, compensate for the current limitations of LLMs, and satisfy higher accuracy demands.

Representative systems generate and consolidate topics, then classify them with evidence (TopicGPT (Pham et al., 2024a); TopicGen (van Wanrooij and Manhar, 2024)); hierarchical organization for literature reviews follows a similar spirit (CHIME (Hsu et al., 2024)). Although some add preprocessing or constraints, they fit the two-round paradigm.

4.2 Level 2: Agentic TM and Similar Systems (Tables 21-26)

Rather than relying on just one or two passes of text generation over a corpus, this level employs more sophisticated workflows that use diverse prompts for distinct functions, or coordinated agentic systems (Weng, 2023) that are integrated with knowledge bases, retrieval mechanisms, and iterative evaluation loops to en-

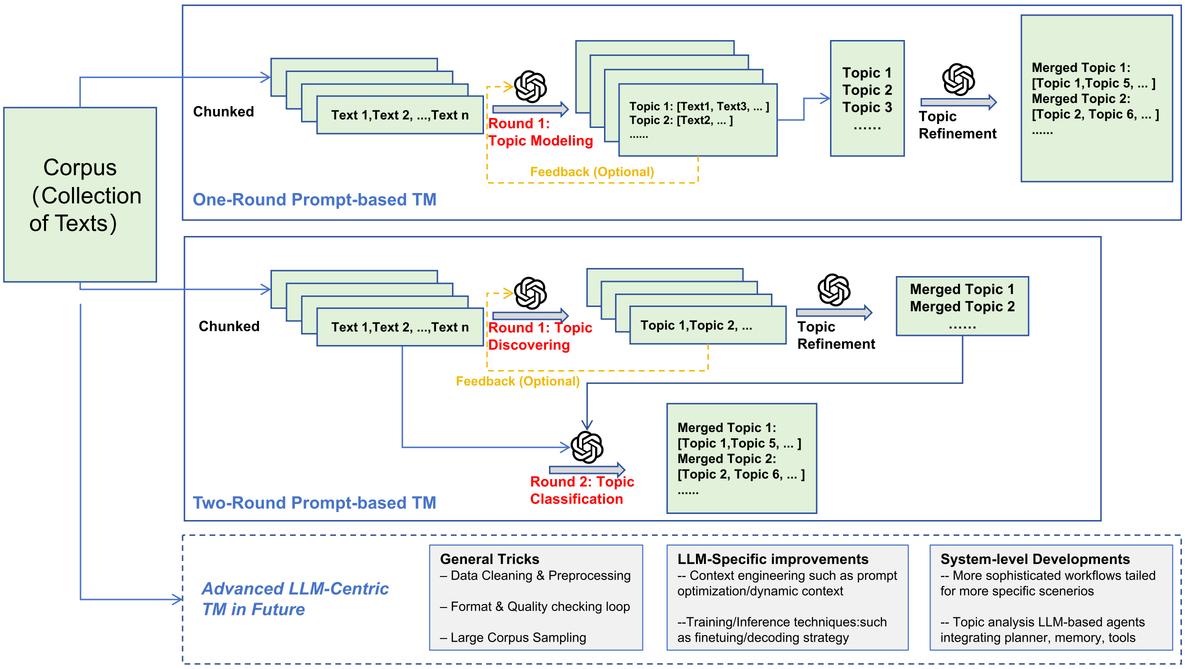

Figure 2: Illustration of Two Types of Prompt-based TMs and Advanced LLM-Centric TM in Future

able autonomous task solving. Agentic TM and similar systems converge on a unified collect/retrieve/aggregate → measure/induce/analyze → visualize/report pipeline. Practically, they combine short/long-term memory with multi-tool collaboration (RAG, graphs, NER/clustering), forming a tool-augmented, system-integrated workflow.

Agentic TM. The scope of the agentic TM remains limited to topic-level analysis. LLooM introduces concept induction, leveraging LLMs to iteratively synthesize and generalize text, elevating topic modeling from low-level keywords to high-level concepts with explicit inclusion criteria (Lam et al., 2024). SPARK proposes the first LLM-based multi-agent simulation framework that jointly models the co-evolution of topics and stances in online discourse through natural language interactions, capturing bidirectional dynamics that traditional topic modeling approaches cannot address (Zhang et al., 2025). Aella Explorer maps large scientific corpora via SPECTER2 embeddings, UMAP, and K-Means, with LLM-curated labels for interpretability and interaction (Kuzco and Hogan, 2025).

Similar Systems. Beyond topic analysis, similar text analysis systems also perform reasoning on sentiment, stance, and insight. POSIM (Public Opinion Simulator) is a multi-agent simulation framework that models the evolution of public opinion on social media for governance purposes (Zhang et al., 2026). EvoCorps frames discourse governance as a dynamic social game and coordinates roles for monitoring, planning, grounded generation, and multi-identity diffusion (Lin et al., 2026). Automated survey pipelines use multi-round retrieval, clustering, and Attribute-Tree distillation to drive RAG-based outlines and drafting, yielding auditable evidence-to-text outputs (Liang et al., 2025). Collectively, as a general engineering paradigm for LLM-based applications, agentic systems are capable of performing extended chains of reasoning and multi-dimensional analysis beyond traditional topic modeling.

4.3 Interpretation of LLMs as Topic Models

Recently, researchers have begun exploring the reinterpretation of LLMs as topic models to open the black box of their internal representations. Lim and Lauw (2023) proposed treating TLMs as "superposed topic models", demonstrating that coherent topics can be disentangled from the weights of GPT-2 and LLaMA, with effectiveness validated through zero-shot topic modeling. Subsequently, to address the evaluation dilemma caused by BPE tokenization, the same authors introduced a ranking-based method to recover full words from BPE tokens, enabling existing evaluation metrics to be applied to BPE-space topic models (Lim and Lauw,

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. Abbreviation: Sup. = Supervised, Unsup. = Unsupervised, Med. = Medium, Hier. = Hierarchical, Repr. = Representation.

2025). Building on this, Xu et al. (2026) proposed a unified framework: in the white-box setting, recovering document-topic and topic-word structures via attention probing; in the black-box setting, reformulating topic modeling as a long-input generation task with a post-generation signal compensation strategy. Together, these works reveal the mechanism by which LLMs implicitly perform topic modeling, offering new perspectives and methodological support for model interpretability research.

5 Discussions

5.1 Future Roadmap for More Optimized LLM-Centric TM

Future LLM-centric topic models can be optimized along several axes: not only through generalizable pre-/post-processing tricks tailored to different scenarios but also via training/reasoning methods that enhance the model's generative ability—and, more importantly, through agentic system blueprints that integrate tools, knowledge bases, and LLM capabilities. These improvements jointly 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 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 or chunking strategies can be applied: topics are inferred from selected subsets, and remaining documents are assigned using lightweight clustering or C-TF-IDF methods, thereby reducing costs without sacrificing overall structural quality.

LLM-Specific improvements. Prompt design can enhance the quality of topic generation. 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. At the decoding stage, diversity-promoting strategies—such as nucleus sampling, diverse beam search, or temperature scaling—can reduce repetitive or overly generic topic labels, encouraging the model to surface a broader and more representative set of latent themes. 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.

System-level Developments. From an agent systems engineering (Weng, 2023) perspective, we portray a holistic blueprint for agentic topic models, integrating memory, tools, and knowledge bases, along with clear design guidelines and practical development paths. For instance, a topic-aware memory and accumulation mechanism could allow agents to retain and refine discovered topic structures across sessions, enabling incremental knowledge consolidation rather than redundant re-computation. Moreover, self-evolution loops—where agents autonomously update their topic taxonomies based on incoming data—or human-in-the-loop evolution, where analyst feedback iteratively reshapes topic granularity and boundaries, can drive continuous model improvement beyond one-shot inference. A summary table of the added content is provided in Table 2.

5.2 Paradigm Shift Brought by LLM-Centric TM

5.2.1 Expanded and more Inclusive Task Scope across Multiple Dimensions

Clarifying Task Boundaries We distinguish traditional topic modeling from adjacent tasks (classification, discovery, extraction, segmentation, hier-

Layer	Module	Primary Responsibilities
Control	Coordinator / Planner	Task decomposition; stage orchestration; merge/split/refine/stop policies; parameter auto-adaptation
Execution	Preprocessing & Context engineering	Cleaning; chunking; summarization; retrieval; top-level classification
	Topic Generation	Topic discovery and refinement (merge/split/prune)
	Postprocessing & Output	Document assignment; quality evaluation/alignment; reporting/visualization
Memory	Short-term Memory	Incomplete/validated topics from current run; key instructions; intermediate human feedback
	Long-term Memory	Domain topic knowledge; accumulated topic library; glossary; evaluation records; user preferences
Tools	Knowledge Bases	Support with structured and unstructured knowledge
	Models & Ops Tools	NER/classification; lightweight inference; graph read/write; embeddings and clustering
	Evaluation & Alignment	Metric computation; human/LLM feedback to drive iteration

Table 2: A structured framework for advanced topical analysis agents, founded on the 4-component agent paradigm.

archies, and corpus organization). Brief task definitions appear in Appendix C, with a summary in Table 1. However, in the LLM-centric paradigm, multi-task abilities blur boundaries: topic modeling becomes a compositional process (e.g., "discover $\rightarrow$ classify $\rightarrow$ organize"), and more complicated tasks such as corpus organization emerge as an iterative extension of topic modeling. This unification reflects a paradigm shift rather than mere tooling.

Input Granularity and Length Choices Traditional TM favors long documents and large corpora due to BoW co-occurrence needs and struggles with short texts; LLM-centric workflows mitigate sparsity via generation and compression, aligning the core challenge to title/abstract-level semantics. In particular, topic modeling for long documents can first be transformed into short-text generation by summarization, title generation, and key-entity extraction. This process converts content at different granularities (titles, abstracts, full texts, and entire collections) into compact, comparable units, thereby improving the use of context and the adaptation of capabilities.

Flexibility Across Modalities, Languages, and Scenarios Traditional NTMs rely on in-domain training data and degrade under shifts. LLM-centric TM not only natively supports multilingual and multimodal inputs but also produces tailored topic representations for various scenarios involving distribution shifts. In engineering practice, the LLM-centric TM is implemented through well-designed prompts or pipelines (retrieval, memory, tools) that allow various scenarios to be applied appropriately.

5.2.2 From Model-level Improvements to System-level Engineering

Historically, NTMs focused primarily on model-level innovations—architectural designs, prior structures, and inference optimization. In contrast, the LLM-centric paradigm shifts emphasis toward system-level engineering: prompt design, retrieval-augmented pipelines, memory mechanisms, and tool integration. Building on this engineering foundation, LLM-centric TM is poised to evolve further—from a static, one-pass analysis tool into an autonomous, interactive, and adaptive system. Future models may actively engage users in a collaborative loop: generating preliminary topics, soliciting feedback (e.g., "split this into finer aspects"), and iteratively refining representations on the fly. Moreover, as LLMs become more agentic, topic modeling could integrate seamlessly with downstream tasks—such as summarization, knowledge graph construction, or trend forecasting—forming end-to-end pipelines where topic discovery directly drives decision-making. Ultimately, the boundary between "topic modeling" and "corpus understanding" may dissolve, giving rise to systems that not only describe what topics exist but also explain why they matter, how they evolve, and what actions they suggest.

5.3 Challenges of LLM-Centric TM

Given that LLM-centric TM remains relatively underexplored while traditional lightweight methods still offer clear benefits for certain tasks, we first present a qualitative comparison of the efficiency–effectiveness trade-offs across the three

Dimension	Traditional TM (e.g., LDA, NMF, NTM)	LLM-Assisted TM (traditional core + LLM)	LLM-centric (Prompt-based / Agentic TM)
Topic quality	Moderate; risks redundancy; BoW limits abstraction; improves with tuning	Moderate–good; LLM aids naming and merge/prune	**Good–excellent**; strong abstraction; rationales and evidence
I/O flexibility	Low; heavy preprocessing; word-list outputs	Medium; LLM summaries/embeddings support richer inputs and topic outputs	**High**; robust inputs; instruction-controlled outputs (labels, summaries, evidence)
Cost & latency	**Low–medium**; CPU-friendly; low latency	Low–medium; sparse LLM calls	Medium–high; token costs and orchestration introduce latency
Scalability & engineering	**High**; online/SVI; mature ecosystem	High; integrates easily; moderate engineering effort	Medium; token costs and no mature ecosystem; agentic setups add memory/tools/orchestration
Privacy & compliance	**Strong**; easy on-prem	Deployment dependent; on-prem strong	Weaker by default (API-based); on-prem feasible but costly
Transfer & multilingual	Medium; limited without special preparation	Good; multilingual embeddings/labels help	**Strong**; performs well across domains and languages

Table 3: A contemporary comparative analysis of different TM paradigms for method selection.

paradigms (Table 3). Building on this analysis, we then outline the overarching vision of LLM-centric TM and discuss the key challenges that must be addressed to advance its maturity.

Challenges of Global Attention and High Costs LLM-Centric TM is a typical "long-input + long-output" task: it requires processing ultra-long inputs and generating extensive topics and others. To achieve this, LLMs need to aggregate global information across thousands of documents and generate topic structures, which in turn demands robust long-range attention and results in an inference cost that grows quadratically. It must avoid over-attending to early contiguous segments and aggregate cross-document evidence in ultra-long contexts—ensuring balanced coverage, de-duplication, and a consistent global topic set hierarchy.

Challenges of Evaluation Benchmarks and Metrics Evaluating LLM-centric topic modeling introduces challenges beyond traditional metrics such as diversity and coherence. As a generation task, there is currently no widely recognized benchmark suite specifically designed for this paradigm. Beyond conventional dimensions like coherence and diversity, evaluation frameworks should also assess structural consistency (e.g., hierarchies), corpus coverage, and faithfulness to source documents. Furthermore, outputs must demonstrate practical usability, such as coverage control that avoids over-generation or under-generation, balanced granularity and distinctness across topics, and traceable evidence linking each topic back to its supporting docs. Establishing standardized benchmarks and metrics that capture these dimensions remains an urgent priority.

Challenges of Faithfulness and Interpretability While the natural language explanations offered by LLMs far surpass traditional probability distributions in readability and user-friendliness, they also introduce fundamental challenges related to faithfulness and interpretability. First, LLMs may hallucinate plausible but unsupported themes that do not actually exist in the corpus. Second, stochastic decoding undermines the stability and reproducibility required for reliable analysis. Third, and most fundamentally, there exists a gap between internal reasoning and surface explanation—the relationship between input and output remains opaque, and how to extract interpretable structures from the model's decision process is still underdeveloped. Addressing these challenges—through hybrid verification mechanisms, uncertainty quantification, or probing internal states for interpretable structures—is essential for building LLM-centric TM that is both powerful and trustworthy.

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, due to the authors' knowledge background and practical experience being primarily focused on LLMs, some claims may overestimate the impact of integrating LLMs on TM while underemphasizing foundational probabilistic modeling and its theoretical depth.

Acknowledgements

This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFC3305401 and in part by the National Natural Science Foundation of China under Grants 62172053, 62302059, and T2322023.

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. 2024. 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.

Weisi Chen, Walayat Hussain, and Junjie Chen. 2025. Glmtopic: A hybrid chinese topic model leveraging large language models. *Computers, Materials & Continua*, 85(1):1559–1583.

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.

FulPhil. 2024. Agentopic: Topic modeling with llm-based agent. `https://github.com/pariskang/AgenTopic`.

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. *Neurocomputing*, 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).

Sean Kuzco and Sam Hogan. 2025. Aella science dataset explorer. LAION research paper dataset visual explorer. MIT License. Stars: 696, Forks: 111 (accessed 2025-12-31).

Michelle S. Lam, Janice Teoh, James A. Landay, Jeffrey Heer, and Michael S. Bernstein. 2024. Concept induction: Analyzing unstructured text with high-level concepts using lloom. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, page 1–28. ACM.

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.

Xun Liang, Jiawei Yang, Yezhaohui Wang, Chen Tang, Zifan Zheng, Shichao Song, Zehao Lin, Yebin Yang, Simin Niu, Hanyu Wang, Bo Tang, Feiyu Xiong, Keming Mao, and Zhiyu li. 2025. Surveyx: Academic survey automation via large language models. *Preprint*, arXiv:2502.14776.

Jia Peng Lim and Hady Lauw. 2023. Disentangling transformer language models as superposed topic models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 8646–8666, Singapore. Association for Computational Linguistics.

Jia Peng Lim and Hady Lauw. 2025. Interpreting topic models in byte-pair encoding space. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 10810–10838, Abu Dhabi, UAE. Association for Computational Linguistics.

Ning Lin, Haolun Li, Mingshu Liu, Chengyun Ruan, Kaibo Huang, Yukun Wei, Zhongliang Yang, and Linna Zhou. 2026. Evocorps: An evolutionary multi-agent framework for depolarizing online discourse. *arXiv preprint arXiv:2602.08529*.

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. 2025a. 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.

Jiyuan Liu, Jiaxing Yan, Chunjiang Zhu, Xingyu Liu, Li Qing, and Yanghui Rao. 2025b. Neural topic modeling via contextual and graph information fusion. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 13247–13263, Suzhou, China. 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.

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*, 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, Ziang Li, Yi Zheng, Gongzhi Luo, Yongquan Zhou, and Yuxiang Zhou. 2025a. Topic modeling and alignment with large language models for multi-labeled text corpora. *Expert Systems with Applications*, 264:125737.

Rui Wang, Xing Liu, Yanan Wang, Shuyu Chang, Yuanzhi Yao, and Haiping Huang. 2025b. 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.

Rui Wang, Jiahao Lu, Xincheng Lv, Shuyu Chang, Yansheng Wu, Yuanzhi Yao, Haiping Huang, and Guozi Sun. 2025c. Mining user preferences from online reviews with the genre-aware personalized neural topic model. In *Proceedings of the ACM on Web Conference 2025*, WWW '25, page 1877–1888, New York, NY, USA. Association for Computing Machinery.

Rui Wang, Yanan Wang, Ziang Li, Haitao Cheng, and Guozi Sun. 2025d. Distributed keyword-guided topic model with lexical knowledge supervision. *ACM Trans. Knowl. Discov. Data*, 19(6).

Rui Wang, Deyu Zhou, Haiping Huang, and Yongquan Zhou. 2025e. Mit: Mutual information topic model for diverse topic extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2):2523–2537.

Lilian Weng. 2023. Llm-powered autonomous agents. *lilianweng.github.io*.

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, Fufang Wen, Beilin Chu, Zhibing Fu, Qinhong Lin, Jiaqi Liu, Binjie Fei, Yu Li, Linna Zhou, and Zhongliang Yang. 2025. 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.

Xuan Xu, Zhongliang Yang, Haolun Li, Beilin Chu, Rui Tian, Yu Li, Shaolin Tan, and Linna Zhou. 2026. Llm as attention-informed ntm and topic modeling as long-input generation: Interpretability and long-context capability. *Preprint*, arXiv:2510.03174.

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.

Bowen Zhang, Yi Yang, Fuqiang Niu, Xianghua Fu, Genan Dai, and Hu Huang. 2025. SPARK: Simulating the co-evolution of stance and topic dynamics in online discourse with LLM-based agents. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 23061–23073, Suzhou, China. Association for Computational Linguistics.

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.

Linhai Zhang, Xuemeng Hu, Boyu Wang, Deyu Zhou, Qian-Wen Zhang, and Yunbo Cao. 2022. Pre-training and fine-tuning neural topic model: A simple yet effective approach to incorporating external knowledge. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5980–5989, Dublin, Ireland. Association for Computational Linguistics.

Yongmao Zhang, Kai Qiao, Zhengyan Wang, Ningning Liang, Dekui Ma, Wenyao Sun, Jian Chen, and Bin Yan. 2026. POSIM: A multi-agent simulation framework for social media public opinion evolution and governance. *arXiv preprint arXiv:2603.23884*.

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

Tuğba Çelikten and Aytuğ Onan. 2025. Topic modeling through rank-based aggregation and llms: An approach for ai and human-generated scientific texts. *Knowledge-Based Systems*, 314:113219.

A E1 Information About Use Of AI Assistants

We use GPT-based assistance within the scope below:

1. Draft refinement: grammar, clarity, and conciseness edits of the 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 checking for typos 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 4: 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 5: Summaries of Embedding-based Topic Models

Method Name	Dataset / Domain	Evaluation Metrics	Core Methodology	Major Contribution
ETM (Dieng et al., 2020)	Scientific Texts, Heterogeneous Corpora (evaluated in the uploaded paper on topic counts 10–50)	Topic Coherence, Stability	1. Embeds both words and topics into a shared latent space. 2. Parameterizes topic-word distributions by a SoftMax over the dot products between topic and word embeddings. 3. Leverages pretrained GloVe embeddings to enhance semantic coherence.	1. Introduces a semantically informed neural topic model by coupling topic modeling with word embeddings. 2. Improves topic coherence and interpretability through shared embedding geometry. 3. Serves as a strong NTM baseline for later embedding-enhanced and hybrid topic models.
EmTM (Rashid et al., 2023b)	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.
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.
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.
C-Top2Vec (Angelov and Inkpen, 2024)	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.
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.
DiffETM (Shao et al., 2025)	20NewsGroup, New York Times (NYT-3000 / NYT-5000 / NYT-10000)	Topic Coherence, Topic Diversity, Topic Quality, Perplexity	1. Extends ETM by introducing a diffusion process into document-topic distribution sampling. 2. Uses a feed-forward network to obtain enhanced document representations, then progressively adds Gaussian noise with a linear noise scheduler. 3. Combines the diffusion module, document-topic distribution module, and topic-word distribution module, and optimizes reconstruction loss with KL divergence.	1. Relaxes the overly restrictive logistic-normal assumption in ETM and better captures complex document-topic distributions. 2. Significantly improves coherence, diversity, and perplexity over ETM and other strong neural topic model baselines. 3. Is among the first works to incorporate diffusion modeling into embedded topic modeling.
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.
RankAgg-LLM-TM (Çelikten and Onan, 2025)	Scientific Abstracts: Human-authored PubMed Radiology Abstracts, AI-generated Abstracts, AI-paraphrased Abstracts	Topic Coherence, Topic Diversity, Topic Stability	1. Uses LLaMA, Gemini, and GPT-4 to generate document embeddings, concatenates them, and applies UMAP for dimensionality reduction. 2. Integrates traditional and neural topic models, including LDA, CTM, VAE-based topic model, and ETM, to produce candidate topics. 3. Applies rank-based aggregation with Borda count over coherence, relevance, and stability, followed by redundancy filtering and final topic selection.	1. Proposes a robust ensemble framework for topic modeling across AI-generated, AI-paraphrased, and human-authored scientific texts. 2. Mitigates positional bias and instability in LLM outputs through rank-based aggregation. 3. Achieves stronger coherence, diversity, and stability than conventional and state-of-the-art neural topic modeling baselines.
GLMTopic (Chen et al., 2025)	Weibo (Chinese Social Media, 22,184 posts)	Coherence (c_v), Interpretability (Human Evaluation), Human Involvement	1. Proposes a hybrid topic modeling framework integrating LLMs with graph-based embedding (ACGE), UMAP dimensionality reduction, and HDBSCAN clustering. 2. Introduces LLM-based representation tuning to refine topic keywords and improve semantic consistency. 3. Reduces reliance on heavy preprocessing by leveraging community-enhanced embeddings for direct semantic extraction. 4. Enables automatic topic interpretation without requiring human experts for labeling.	1. Achieves higher topic coherence (c_v=0.610) than LDA and BERTopic, demonstrating stronger semantic consistency. 2. Produces more interpretable and contextually meaningful topics, especially for Chinese short-text data. 3. Significantly reduces human involvement in topic labeling compared to traditional methods. 4. Shows that LLM-enhanced hybrid models are particularly effective for non-English and short-text scenarios.

Table 6: 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 7: 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., 2020)	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 → z → x, label-conditioned prior p(z|y). 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 8: Summaries of Other VAE-based Methods

Method Name	Dataset	Evaluation Metric	Methods	Major Contribution
PT-NTM (Zhang et al., 2022)	20NG, NYTimes, Other Large Corpora	Topic Coherence, Topic Diversity	1. Pre-trains a neural topic model on a large corpus and then fine-tunes it on downstream datasets. 2. Uses a standard encoder-decoder neural topic model (MLP-based) with BoW inputs. 3. Incorporates Dirichlet prior via MMD-based distribution matching.	1. Proposes a simple yet effective pretrain-then-finetune paradigm for topic modeling. 2. Significantly improves coherence and diversity with low computational overhead compared to PLM-based methods. 3. Reduces data requirement, achieving strong performance even with limited labeled data.
CGTM (Liu et al., 2025b)	20NG, Other Benchmark Datasets	NPMI, Topic Coherence, Topic Diversity	1. Extends VAE-based topic modeling by fusing contextual (BERT) and BoW representations via topic-wise fusion. 2. Introduces topic alignment and topic sharpening constraints to improve informative topic learning. 3. Incorporates word co-occurrence graph information via graph information fusion and graph decoder.	1. Effectively combines contextual semantics and statistical word information for improved topic quality. 2. Enhances both informativeness and coherence of topics through fusion and constraint mechanisms. 3. Achieves superior performance over state-of-the-art baselines in both automatic and human evaluations.
DiskTM(-LK) (Wang et al., 2025d)	Multiple Benchmark Corpora	Topic Coherence, Topic Diversity	1. Introduces distributed keyword prior knowledge via Gaussian-based topic-word distributions. 2. Aligns topic-word distributions with keyword priors using column-wise constraints. 3. Extends model with lexical knowledge (e.g., WordNet) to improve interpretability.	1. Enables keyword-guided topic modeling aligned with user interests. 2. Improves topic interpretability and diversity compared to purely unsupervised models. 3. Incorporates external lexical knowledge for semantically meaningful topic discovery.
MIT (Wang et al., 2025e)	Benchmark Text Corpora	Topic Coherence, Topic Diversity	1. Maximizes mutual information (MI) between word distribution and topic distribution. 2. Uses encoder network to infer topic distributions and applies statistical two-sample tests (MMD/Energy) to match Dirichlet prior. 3. Introduces a statistic network for MI estimation and optimization.	1. First neural topic model purely based on mutual information maximization. 2. Significantly improves topic diversity while maintaining high coherence. 3. Achieves more stable training compared to adversarial topic models.

Table 9: 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 10: 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.
GPTM (Wang et al., 2025c)	Amazon Reviews (Books, Sports, Movies)	Topic Coherence, Topic Diversity, Personalized Metrics (PHR, PGC)	1. Incorporates genre-aware text representations using a pre-trained transformer to align topics with product categories. 2. Introduces dual inference networks for topic modeling and user preference modeling. 3. Utilizes genre-aware and user-aware contrastive learning to enhance topic discrimination and personalization. 4. Matches topic distributions to Dirichlet prior via MMD to ensure interpretability.	1. Enables joint modeling of topics and user preferences, generating personalized topic profiles. 2. Improves alignment between topics and external semantic categories (genres). 3. Achieves higher topic coherence and diversity while capturing user-specific interests. 4. Introduces new evaluation metrics (PHR, PGC) for personalized topic modeling.

B.2 Summaries of LLM-Assisted Topic Models

Table 11: 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., 2025a)	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 12: 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., 2025b)	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 (FinBERT2) (Xu et al., 2025)	Financial news, SEC filings, Earnings reports	Topic Coherence, Domain Relevance, Cluster Stability	1. Builds **Fin-Topicmodel** upon the **FinBERT2** 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 13: 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 (FulPhil, 2024)	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 14: 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.
LLM-Eval-TM (Stammbach et al., 2023)	NYT, WikiText	Topic Coherence (Rating, Intrusion), Correlation with Human Judgments, ARI (for topic number)	1. Uses large language models to automatically evaluate topic coherence via rating and intrusion tasks. 2. Proposes LLM-based scoring as an alternative to traditional metrics such as NPMI and Cv. 3. Introduces a document-level evaluation strategy using LLM-based label assignment and purity measurement. 4. Incorporates research-question-aware prompting to guide topic number selection.	1. Demonstrates that LLM-based evaluation correlates better with human judgments than traditional coherence metrics. 2. Shows that word-level coherence (WT) is insufficient; document-level purity (DT) is more reliable. 3. Provides a practical framework for automatically selecting the number of topics. 4. Bridges topic modeling evaluation with LLM-based semantic understanding.
BASS (Li et al., 2025)	Bills, Sci-Fi Dataset	Clustering Metrics (Purity, ARI, NMI), Answer Quality, Answer Consistency, Human Preference	1. Proposes a human-in-the-loop LLM-assisted topic modeling framework (BASS) combining LLMs and active learning. 2. Integrates LLM-generated topic suggestions with user refinement to align topics with user intent. 3. Uses active learning with LDA-based representations to scale topic assignment efficiently. 4. Conducts user studies to evaluate how topic models help users understand document collections.	1. Shows that unsupervised LLM topic models generate readable but overly generic topics, especially for domain-specific data. 2. Demonstrates that human supervision significantly improves topic quality and downstream understanding. 3. Finds that traditional LDA remains competitive in clustering quality despite weaker interpretability. 4. Highlights limitations of LLMs in scalability, hallucination, and domain-specific topic discovery.

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 15: 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 16: 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 17: Structured Summaries of the Methodology of paper (Xu et al., 2026)

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 18: 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 19: 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: Sub-subcategory (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 20: Structured Summaries of prompts for CHIME (Hsu et al., 2024)

Stage	Core instruction description (keep placeholders in)	Example output
Distill (Summarization)	Input: Text example $\{text\}$. Prompt: "Summarize the main point into {n_bullets} bullet points, each with {n_words} words. Return JSON: {bullets}."	Bullets: ["gender roles criticized", "dismiss women concerns"]
Synthesize (Concept Generation)	Input: Bullet summaries $\{bullets\}$. Prompt: "I have these bullet summaries: {bullets}. Write {n_concepts} unifying patterns. For each pattern, provide: (1) NAME, (2) classification PROMPT, (3) example_ids. Return JSON."	Concept: "Criticism of gender roles"
Score (Concept Application) Label: Yes	Input: Concept prompt $\{concept\_prompt\}$ and dataset. Prompt: "Given the concept definition, determine whether each text matches the concept. Provide label and reasoning."	Match score: 1.0
Loop (Refinement)	Iteratively re-run Distill–Cluster–Synthesize on uncovered data to produce higher-level concepts.	Higher-level concept: "Gender discourse conflict"

Table 21: Structured Prompts for LLooM Concept Induction(Lam et al., 2024)

Stage	Core instruction description (keep placeholders in)	Example output
Agent Initialization	Initialize agent with persona $\{persona\}$ including age, traits, and initial stance.	Agent: Age=35, Trait=Neutral
Memory Update	Prompt: "Previous memory: {long_memory}. Today's summary: {short_memory}. Update long-term memory."	Updated memory: "Concern about safety issues persists."
Stance Generation	Prompt: "In this role play, act as a social media user. Write a tweet reflecting your opinion. Indicate stance: -1/0/1."	Tweet: "Self-driving cars need better safety."
Topic Evolution	Prompt: "Generate a tweet including hashtags #EventName#." New hashtags introduce new subtopics.	New topic: #SafetyFirst
Simulation Loop	Agents interact $\rightarrow$ update stance $\rightarrow$ generate new topics $\rightarrow$ update topic tree.	Topic tree expands

Table 22: Structured Prompts for SPARK Multi-Agent Topic Evolution(Zhang et al., 2025)

Stage	Core instruction description (keep placeholders in)	Example output
Query Understanding	Input: User query $\{question\}$. Prompt: "Given the query, identify relevant aspects for exploration."	Query intent: "political themes"
Pattern Discovery	Prompt: "Analyze dataset and identify key themes or clusters relevant to {question}. Return structured themes."	Themes: [Policy, Economy, Trust]
Iterative Refinement	Prompt: "Refine themes based on user feedback: {feedback}."	Updated themes: [Policy debate, Social trust]
Interactive Loop	User → Agent → Update → Visualization cycle.	Dashboard updated
Visualization	Present structured insights in dashboards or tables.	Topic clusters displayed

Table 23: Structured Prompts for Aella Agentic Topic Exploration(Kuzco and Hogan, 2025)

Stage	Instruction description (grounded in paper)	Example output
Analyst Agent (Monitoring)	Continuously monitor incoming posts and evaluate core arguments, opinion extremity, sentiment distribution, and popularity signals. Trigger structured alerts when polarization risks are detected.	Alert: High extremity detected
Strategist Agent (Planning)	Formulate intervention strategies based on current alerts and historical action–outcome memory, including core arguments, leader style, and agent configuration.	Strategy: Promote balanced narrative
Leader Agent (Generation)	Instruction (from paper): Retrieve evidence from the knowledge base, generate multiple candidate responses, and select the most effective output via reflection and voting mechanisms (USC).	Generated post
Amplifier Agents (Diffusion)	Disseminate leader outputs through multiple role identities, generating diverse yet consistent responses to amplify content.	Multiple role-based posts
Feedback and Iteration	Evaluate intervention effectiveness against baseline metrics and update strategy via structured feedback reports.	Report: Extremity reduced

Table 24: Instruction-level abstraction of EvoCorps multi-agent framework(Lin et al., 2026)

Stage	Instruction description (derived from pipeline)	Example output
Keyword Expansion	Expand the input topic into multiple related keywords to improve coverage of literature retrieval.	Keyword list
Reference Retrieval	Retrieve and filter relevant papers from offline and online sources using expanded keywords.	Relevant paper set
AttributeTree Construction	Parse references into structured attribute representations (e.g., method, task, dataset) to enhance information density.	Structured attributes
Outline Generation	Generate a hierarchical outline based on processed references to ensure logical structure.	Outline
Survey Writing	Generate survey content grounded in retrieved references and structured outline.	Draft section
Refinement	Rewrite and refine draft to improve coherence, readability, and logical consistency.	Improved text

Table 25: Instruction-level abstraction of SurveyX survey generation pipeline(Liang et al., 2025)

Stage	Instruction description (derived from Social-BDI architecture)	Example output
Perception → Belief	Update hierarchical belief states (identity, psychological cognition, event opinion, emotion) based on environment input and memory.	Updated belief
Belief → Desire	Map beliefs to behavioral motivations such as expressing opinions, seeking recognition, or spreading information.	Desire: Express opinion
Desire → Intention	Translate motivations into actionable plans, including action type, expression strategy, and narrative style.	Planned action
Intention → Action	Generate content reflecting intention, including stance, emotion, and narrative structure.	Generated post
Environment Feedback	Agent actions interact with the environment through temporal dynamics and social networks, influencing future states.	Updated environment

Table 26: Instruction-level abstraction of POSIM Social-BDI agent framework (Zhang et al., 2026)

C Detailed Introduction of Different Topic-related Tasks

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.