

UDC 81'322

DOI: 10.18413/2313-8912-2025-11-3-0-5

Tatiana A. Litvinova^{1,2} 
Yury A. Ippolitov³ 
Pavel V. Seredin² 

**From STM to GPT: A Comparative
Study of Topic Modeling Methods for AI in Dentistry**

¹ Voronezh State Pedagogical University,
86 Lenin St., Voronezh, 394043, Russia
E-mail: centr_rus_yaz@mail.ru
ORCID: 0000-0002-6019-3700

² Voronezh State University,
1 University sq., Voronezh, 394018, Russia
E-mail: paul@phys.vsu.ru
ORCID: 0000-0002-6724-0063

³ Voronezh State Medical University named after N.N. Burdenko,
12 Studencheskaya Street, 394036 Voronezh, Russia
E-mail: dsygma@mail.ru
ORCID: 0000-0001-9922-137X

Received 13 August 2025; accepted 15 September 2025; published 30 September 2025

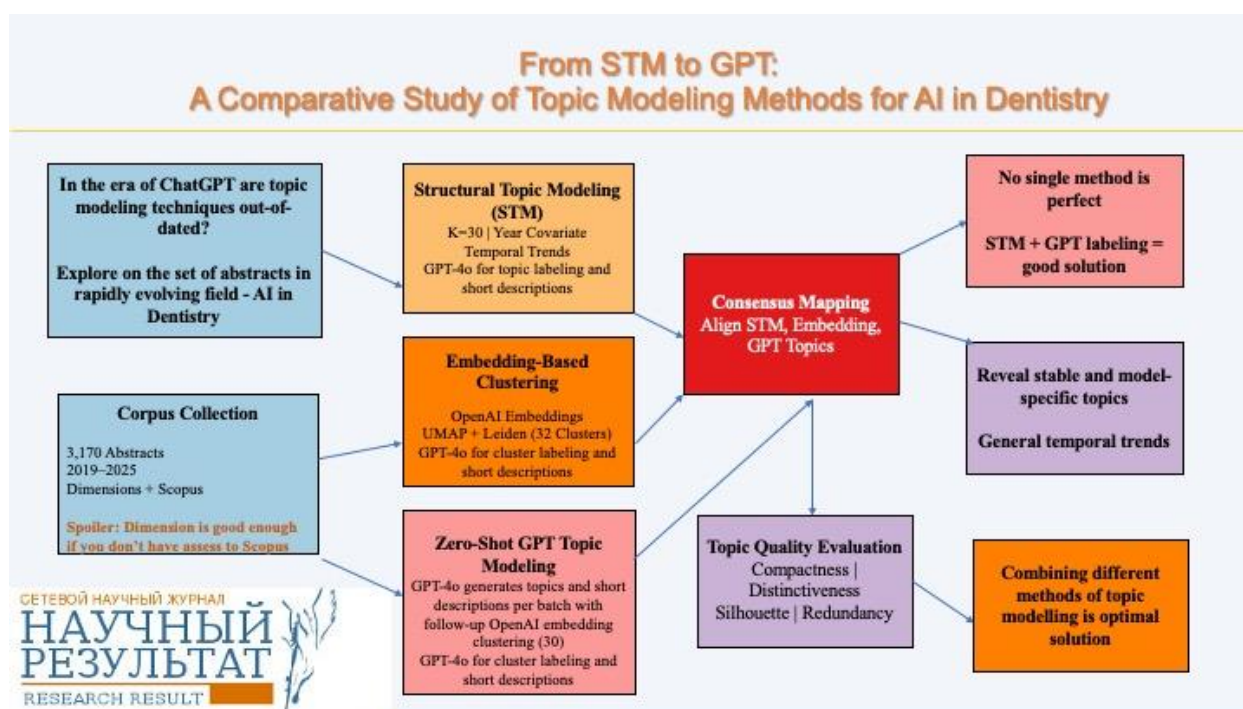
Abstract: This study presents a comprehensive topic modeling analysis of scientific abstracts in the field of artificial intelligence (AI) applied to dentistry. We compiled and analyzed 3,170 peer-reviewed abstracts published between 2019 and 2025 from the Dimensions and Scopus databases. Three complementary approaches were compared: (1) Structural Topic Modeling (STM), a probabilistic framework incorporating publication year as a covariate to enable temporal trend analysis; (2) embedding-based clustering using the Leiden algorithm on OpenAI text embeddings; and (3) zero-shot GPT-based topic modeling, in which GPT-4o generated topics, descriptions, and keywords directly from batches of abstracts without model training. Topic quality was evaluated using compactness, distinctiveness, silhouette scores, and label redundancy. STM consistently produced the most compact and well-separated topics, embedding-based clustering excelled in identifying discrete semantic groupings, and GPT-based modeling provided interpretable, human-readable labels but exhibited greater thematic overlap. To ensure comparability across methods, we introduced a two-layer alignment framework that integrates topic-level similarity with document-level consensus, enabling robust cross-model comparison. Using this framework, we identified stable topics consistently recovered across methods (e.g., caries detection, radiographic AI diagnostics) as well as method-specific themes. Temporal trend analysis in this shared space revealed a clear shift from foundational AI methods (e.g., image segmentation, image enhancement) toward applied and integrative areas, including large language model applications, patient-facing tools, and AI in clinical education. Our results underscore the value of combining classical probabilistic models with modern large language model (LLM) tools for optimal topic modeling performance. While GPT-4o enhances interpretability, it should not be used in isolation for mapping thematic structures in

scientific literature, at least not without pre-screening and prompt experimentation. Overall, our findings demonstrate the importance of hybrid topic modeling for mapping thematic structures in fast-evolving scientific domains, with dentistry serving as a case study.

Keywords: Topic Modeling; Structural Topic Model; GPT-4; Large Language Models; Dentistry; AI in Dentistry; LLM in Dentistry; Scientometric Analysis; Embedding-Based Clustering; Zero-Shot Topic Modeling; Bibliometric methods.




Acknowledgements: The study is supported by Russian Science Foundation, grant number 23-15-00060.

How to cite: Litvinova, T. A., Ippolitov, Y. A., Seredin, P. V. (2025). From STM to GPT: A Comparative Study of Topic Modeling Methods for AI in Dentistry, *Research Result. Theoretical and Applied Linguistics*, 11 (3), 85–121. DOI: 10.18413/2313-8912-2025-11-3-0-5



УДК 81'322

DOI: 10.18413/2313-8912-2025-11-3-0-5

Литвинова Т. А.^{1,2} 
Ипполитов Ю. А.³ 
Середин П. В.² 

От STM до GPT: сравнительный анализ методов
тематического моделирования на материале
предметной области «ИИ в стоматологии»

¹ Воронежский государственный педагогический университет
ул. Ленина, 86, Воронеж, 394043, Россия
E-mail: centr_rus_yaz@mail.ru
ORCID: 0000-0002-6019-3700

² Воронежский государственный университет
пл. Университетская, 1, Воронеж, 394018, Россия
E-mail: paul@phys.vsu.ru
ORCID: 0000-0002-6724-0063

³ Воронежский государственный медицинский университет имени Н.Н. Бурденко
ул. Студенческая, 12, Воронеж, 394036, Россия
E-mail: dsvgma@mail.ru
ORCID: 0000-0001-9922-137X

*Статья поступила 13 августа 2025 г.; принята 15 сентября 2025 г.;
опубликована 30 сентября 2025 г.*

Аннотация: Исследование представляет собой анализ научных аннотаций в области применения искусственного интеллекта (ИИ) в стоматологии с использованием различных методов тематического моделирования. Мы сформировали и проанализировали корпус из 3170 аннотаций научных статей, опубликованных в 2019–2025 гг. в изданиях, индексируемых в базах данных Dimensions и Scopus. Были сравнены три подхода к тематическому моделированию: структурное тематическое моделирование (STM) вероятностная модель, позволяющая анализировать временные тенденции; кластеризация на основе эмбедингов с использованием алгоритма Leiden – стабильная альтернатива BERTopic; моделирование с использованием GPT-4o без обучения модели. Для оценки качества тем была применена совокупность метрик. Показано, что алгоритм STM дает наиболее компактную и четко разделённую структуру тем; GPT оказался эффективным для создания названий тем и кратких описаний, но показал большее тематическое перекрытие и менее чёткие границы между темами. Мы также выполнили согласованное выравнивание тем в едином GPT-пространстве и выявили как стабильные, так и специфичные для моделей темы, а также общие временные тренды. Полученные результаты подчёркивают ценность комбинирования классических вероятностных моделей с возможностями LLM для достижения оптимального качества тематического моделирования. Хотя GPT-4o повышает интерпретируемость, его не следует использовать как единственный метод для анализа тем. Предложенный гибридный подход является масштабируемой и воспроизводимой стратегией для проведения обзоров литературы в быстро развивающихся областях исследований.

Ключевые слова: Тематическое моделирование; структурное тематическое моделирование; GPT-4; Большие языковые модели; ИИ в стоматологии; Выравнивание тем; Кластеризация на основе эмбедингов; Извлечение тем с использованием больших языковых моделей; Анализ научной литературы

Информация об источниках финансирования или грантах: Исследование выполнено в Воронежском государственном университете при поддержке Российского научного фонда, грант № 23-15-00060.

Информация для цитирования: Литвинова Т. А., Ипполитов Ю. А., Середин П. В. От STM до GPT: сравнительный анализ методов тематического моделирования на материале предметной области «ИИ в стоматологии» // Научный результат. Вопросы теоретической и прикладной лингвистики. 2025. Т. 11. № 3. С. 85–121. DOI: 10.18413/2313-8912-2025-11-3-0-5



1. INTRODUCTION

In recent years, the exponential growth of scientific publications has created unprecedented opportunities and challenges for understanding the thematic landscape of research. The volume, diversity, and interdisciplinarity of modern scholarly output render traditional narrative-based literature reviews insufficient for mapping the breadth and evolution of research topics, particularly in specialized and rapidly developing domains (Torres et al., 2025). One such domain is artificial intelligence (AI) in dentistry, where AI applications in diagnostic imaging, segmentation, risk prediction, and treatment

planning have attracted considerable scholarly attention. AI is considered to represent a paradigm shift in this field (Cosola et al., 2025), and transformation of the modern dental practice workflow through AI is anticipated in the next decade (Silveira, 2024). The rapid expansion of this literature (Shirani, 2025) mirrors the accelerating adoption of AI in clinical workflows; however, the thematic organization and temporal development of the field remain incompletely characterized.

Topic modeling has become a powerful computational technique for uncovering hidden themes within large collections of text.

By examining patterns of word co-occurrence and semantic relationships, it allows researchers to identify coherent groups of documents, measure the prevalence of themes, and track how topics evolve over time. Beyond descriptive analysis, topic modeling provides a scalable and replicable alternative to manual coding, facilitating systematic reviews in fields where the volume of publications exceeds the capacity of traditional synthesis methods. With advances in computational power and big data, topic modeling has been widely applied across various domains, including text classification, information retrieval, and recommendation systems, establishing itself as a fundamental technology in natural language processing and text mining (Hu et al., 2025).

Various algorithmic paradigms have been developed for topic modeling, each grounded in different linguistic and computational principles. Probabilistic models such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003) and Structural Topic Modeling (STM) (Roberts et al., 2019) represent topics as distributions over words and documents as mixtures of topics. STM advances this approach by incorporating document-level covariates, such as publication year, enabling the explicit modeling of temporal trends and the impact of metadata on topic prevalence. These models are valued for their interpretability, reproducibility, and strong statistical foundation. However, they require manual interpretation of topic-word distributions, careful selection of the number of topics, and domain expertise to ensure meaningful labeling (Lee et al., 2025; Sbalchiero and Eder, 2020).

Embedding-based clustering offers an alternative approach to topic extraction. By leveraging advances in transformer-based language models, such as BERT and its domain-specific variants, this method transforms documents into high-dimensional semantic vectors that capture contextual meaning beyond simple lexical similarity. Clustering algorithms are then applied to

these embeddings to group documents based on semantic proximity. Methods like BERTopic (Grootendorst, 2022), kBERT (Islam, 2025), and other implementations often produce fine-grained, semantically coherent clusters that can identify subtopics overlooked by probabilistic approaches. For instance, BERTopic (Grootendorst, 2022) utilizes transformer-based embeddings, dimensionality reduction (UMAP), and clustering (HDBSCAN) to discover topics and is widely used in scientific literature analysis. However, embedding-based methods have some limitations: they do not inherently model temporal dynamics and typically assume each document belongs to a single topic, which may not fully capture multi-topic documents (Islam, 2025).

A third, rapidly evolving approach leverages large language models (LLMs) such as GPT-4, which can perform topic induction and labeling in a zero-shot setting. These models are trained on vast text corpora using transformer architectures, enabling them to generate topic descriptions and thematic summaries directly from raw text without requiring training on the target corpus. The strength of LLMs lies in their ability to produce linguistically fluent and contextually rich topic labels. However, concerns persist regarding their stability, transparency, and vulnerability to biases inherited from training data, as well as their tendency to generate hallucinations (Mu et al., 2024a). Additionally, their outputs may exhibit greater topical overlap and weaker structural separation compared to probabilistic or embedding-based methods (Mu et al., 2024a). Other challenges introduced by LLMs include API costs, rate limits, and the dependence of topic extraction quality on prompt accuracy.

Despite the widespread adoption of topic modeling across various disciplines, few studies have systematically compared these methodological approaches specifically in the context of analyzing scientific literature. Most existing evaluations focus on general-purpose datasets or single modeling families, limiting our understanding of how different methods

perform on domain-specific corpora. Furthermore, although AI in dentistry represents a promising case for thematic mapping due to its interdisciplinary nature and rapid growth, it has yet to be examined through a comprehensive, multi-method topic modeling analysis.

To address these gaps, this study applies and compares three topic modeling approaches to a curated corpus of 3,170 peer-reviewed abstracts on AI in dentistry, published between 2019 and 2025. The methods include: 1) structural topic modeling (STM) to analyze topic prevalence and temporal trends; 2) embedding-based clustering using the Leiden algorithm to group semantically similar documents in high-dimensional space; and 3) zero-shot GPT-based topic modeling, which generates topics and their brief descriptions directly without requiring model training.

To evaluate and compare the outputs of these methods, we developed a blended similarity framework for aligning topics across models and assessing their quality. At the topic level, this framework combines two complementary similarity measures: TF-IDF-based cosine similarity, which quantifies lexical overlap between topics by weighting terms according to their frequency and distinctiveness within the corpus; and embedding-based semantic similarity, which captures contextual and conceptual relationships between topics by comparing their vector representations in a high-dimensional semantic space. By integrating these measures, we account for both surface-level lexical matches and deeper semantic connections, enabling more robust cross-model topic alignment.

At the document level, we introduce a consensus score that evaluates how consistently different models assign documents to corresponding topics. This metric combines structural overlap measured by the Jaccard index of document sets with semantic fit, assessed through embedding-based similarity between shared documents and candidate centroids. This approach

ensures that topics deemed “aligned” at the lexical-semantic level also demonstrate coherence in their document memberships.

This comparative design enables us to evaluate not only the internal coherence and distinctiveness of topics produced by each method but also their ability to capture temporal trends and domain-specific themes. The AI-in-dentistry corpus serves as a practical testbed for methodological insights with broader implications for topic modeling in specialized scientific fields. This study is guided by the following research questions:

1. How do probabilistic, embedding-based, and LLM-based topic modeling methods differ in the thematic structures they generate from domain-specific scientific abstracts?

2. Which methods are most effective for producing compact, distinct, and interpretable topics in the context of AI applications in dentistry?

3. To what extent do the topics identified by different methods overlap or diverge, and how can cross-model alignment of these models contribute to a consensus view of the field’s thematic landscape?

4. How do the identified topics evolve over time?

2. LITERATURE REVIEW

Topic modeling methods such as Latent Dirichlet Allocation (LDA), Non-negative Matrix Factorization (NMF), and Structural Topic Models (STM) are fundamental techniques in natural language processing (NLP). These methods have long been employed to automate literature reviews by identifying thematic structures within large corpora, enabling the rapid detection of research topics and their evolution over time. They offer scalable alternatives to manual coding, which is particularly valuable when analyzing hundreds or thousands of documents. Introduced by Blei et al. (2003), LDA remains one of the most widely used techniques, especially in systematic literature reviews and bibliometric analyses (Ogunleye et al., 2025).

One of the most popular traditional approaches to topic modeling is the Structural Topic Model (STM), which enhances LDA by incorporating document-level covariates (e.g., year, metadata) to model how topic prevalence and word usage vary across groups (Roberts et al., 2019). STM is actively used for trend analysis, which is especially important for scientific literature reviews (Şakar and Tan, 2025; Ogunleye et al., 2025).

BERTopic is widely used for topic modeling in scientific literature (Wu et al., 2024). Benz et al. (2025) compared the results of LDA and BERTopic applied to the same dataset of scientific articles ($n = 34,797$) and found that, despite differences in their methodologies, both LDA and BERTopic generate topic spaces with a similar overall structure. However, significant differences emerge when examining specific multidimensional concepts. Additionally, Jung et al. (2024) evaluated four topic modeling methods (LDA, Nonnegative Matrix Factorization, Combined Topic Models, and BERTopic) and demonstrated that the BERT-based model outperformed the others in terms of both topic diversity and coherence.

However, manual labeling of topics remains a bottleneck in scaling such reviews. Topic models generate outputs as keyword lists, which require manual interpretation for labeling. This process is labor-intensive, subjective, and difficult to scale, especially for dynamic corpora like scientific literature.

In response, researchers have explored LLM-powered summarization and topic assignment, demonstrating improved labeling quality and reduced subjectivity. For example, Kozłowski (2024) proposes assessing the reliability of three LLMs (Flan, GPT-4o, and GPT-4 Mini) for topic labeling in scientific literature. His findings indicate that both GPT models can accurately and precisely label topics based on the models' output keywords. Additionally, three-word labels are preferred to better capture the complexity of research topics.

With advancements in natural language processing and artificial intelligence, large language models, such as OpenAI's ChatGPT, have opened new avenues for analyzing extensive scientific literature. Numerous methods for topic extraction have been developed based on LLMs (for example, GPTopic by Reuter et al. (2024)). These tools facilitate the efficient extraction and analysis of information from publications, including the identification of study areas, research topics, and methodological approaches (Hu et al., 2025; Lee et al., 2024; Sharma and Wallace, 2025).

Riaz et al. (2025) compare LDA with advanced transformer-based models such as BERT and GPT. They propose integrating BERT with extra long-term memory networks, specifically XLNet, as a promising approach. Similarly, Mathis et al. (2024) evaluate the effectiveness of an open-source LLM against traditional human thematic analysis in processing qualitative interviews within a psychiatric context. Their study demonstrates that open-source LLMs can effectively generate robust themes from qualitative data, achieving a high degree of similarity to those produced by human analysts.

Despite growing interest, there remains limited research comparing the use of LLMs in thematic analysis compared to human coding and traditional topic modeling techniques. Furthermore, while LLMs, including those employed in GPTopic, offer flexibility and zero-shot capabilities, they often struggle to maintain topic granularity and tend to produce overlapping or redundant topics. In contrast, traditional methods, such as LDA and the BERT-based models remain consistent and reliable, generating coherent topics without hallucinations. This makes them highly adaptable across various domains without requiring extensive fine-tuning (Mu et al., 2024b).

Since no single method can fully capture the topical structure of a text collection, combining traditional and LLM-

based topic modeling techniques can be advantageous. However, this integrated approach remains underutilized in studies analyzing scientific abstracts. For instance, Meng et al. (2024) integrated LLMs and topic modeling methods, including GPT and BERTopic. Mahmoud et al. (2025) developed a method to identify trends in a large dataset of 76,689 research papers by first embedding the texts using SBERT, followed by dimensionality reduction with PCA and UMAP, and then hierarchical clustering. Subsequently, LDA was applied to detect topics, which were automatically labeled using the ChatGPT API. This combined approach produced semantically meaningful topics. A similar method was employed by Tarek et al. (2024), who used HDBSCAN on SBERT embeddings, followed by LDA applied to each cluster. The OpenAI “gpt-3.5-turbo-0125” model was utilized to summarize the abstracts.

In summary, integrating LLMs into topic modeling workflows marks a significant advancement in tackling the longstanding challenges of interpretability, consistency, and human validation. However, it is essential to compare their performance with traditional topic modeling methods in the analysis of scientific abstracts, as only a limited number of studies have focused on this type of text.

The volume of literature on AI in dentistry is rapidly increasing (Shirani, 2025; Xie et al., 2024); however, to our knowledge, no study has yet applied topic modeling analysis to this expanding field. Numerous systematic reviews exist – for example, Shirani (2025) showed that 244 out of 1,368 studies in his dataset are reviews – but most are general and do not focus on specific AI applications within particular disciplines (ibid.). These reviews are typically narrative in style and cover several decades of publications. A few bibliometric reviews have been conducted in this area (Shirani, 2025; Xie et al., 2024; Zatt et al., 2024), but they generally employ simple NLP methods, such as keyword co-occurrence analysis using VOSviewer (see also Allani et al., 2024; de

Magalhães and Santos, 2025), combined with manual coding (Shirani, 2025), or primarily analyze publication metadata like authorship, citations, and keywords. Traditional bibliometric approaches often struggle to effectively manage and analyze the vast volume of literature, especially when identifying specific research domains. In contrast, advanced techniques – including large language models and topic modeling – can better reveal evolving research patterns and priorities (Hu et al., 2025).

To the best of our knowledge, no prior study has systematically applied topic modeling methods to a curated dataset of scientific abstracts on AI in dentistry. In this context, dentistry offers a well-defined and timely testbed for evaluating topic extraction methods in a rapidly evolving applied domain. Our study contributes methodologically by demonstrating how various approaches – structural topic model (STM), embedding-based clustering, and GPT zero-shot topic extraction – can be employed to uncover key research themes and temporal trends within a specialized corpus. To ensure comparability across these diverse methods, we further introduce a two-layer alignment framework that integrates topic-level similarity with document-level consensus, enabling robust cross-model comparison. Although the empirical focus is on dentistry, the comparative framework we develop is generalizable to other fields facing similar challenges in mapping the thematic landscape of AI research.

3. MATERIALS AND METHODS

3.1. Dataset construction

To investigate the thematic structure and research trends in the application of artificial intelligence within dentistry, we compiled a domain-specific corpus by merging data from two major bibliographic databases: Scopus and Dimensions. This approach differs from most previous reviews in the field, which typically focus on a single database such as Web of Science (WoS) or Scopus. Scopus was selected for its status as one of the largest

peer-reviewed abstract and citation databases, providing high-quality metadata and abstracts. Dimensions, on the other hand, is a widely used open-access bibliometric platform that offers exportable metadata, including citation counts, affiliation details, and full abstract information.

The search strategy involved a comprehensive query incorporating common AI terms, with an emphasis on the most recent deep learning and NLP methods (e.g., “deep learning”, “large language model”, “multimodal fusion”) and dental research keywords (e.g., “oral health”, “implant dentistry”). Only English-language documents published between 2019 and 2025 were included. The search was conducted on July 23, 2025.

We selected 2019 as the starting point because, according to our preliminary analysis and other researchers who conducted bibliometric studies using WoS, most papers in this field were published after this year (Shirani, 2025; Xie et al., 2024). These studies reported a modest annual growth rate in publications from 2000 to 2018, followed by explosive growth beginning in 2019 (Xie et al., 2024).

The query used for both databases was designed to comprehensively capture literature intersecting AI and dentistry: ("**deep learning**" OR "**natural language processing**" OR **NLP** OR "**large language model**" OR "**LLM**" OR "**transformer model**" OR "**BERT**" OR "**GPT**" OR "**chatGPT**" OR "**multimodal model**" OR "**multimodal fusion**" OR "**multimodal learning**" OR "**vision language model**" OR "**image based AI**" OR "**3D reconstruction**" OR "**segmentation model**" OR "**object detection**" OR "**representation learning**" OR "**self-supervised learning**") AND ("**dentistry**" OR "**dental**" OR "**oral health**" OR "**oral cavity**" OR "**periodontology**" OR "**periodontitis**" OR "**prosthodontics**" OR

"endodontics" OR "**orthodontics**" OR "**dental caries**" OR "**tooth decay**" OR "**implant dentistry**" OR "**dental diagnosis**" OR "**dental radiology**" OR "**dental imaging**").

The Scopus dataset was exported as a CSV file with UTF-8 encoding, and the Dimensions data were similarly exported using the platform’s metadata export tools. All records were imported into R and cleaned using the readr, dplyr, and stringr packages. The two datasets were deduplicated based on DOI strings, which were normalized by converting to lowercase, trimming whitespace, and removing common prefixes (e.g., <https://doi.org/>) using regular expressions. Records without abstracts were excluded, and documents were filtered to include only those containing more than 50 words to ensure semantic richness.

It is important to highlight that Dimensions offers a valuable alternative for researchers without access to subscription-based databases. In our dataset (see Table 1), 1,857 abstracts (approximately 60.5%) were found in both Scopus and Dimensions, while 659 (about 21.3%) were unique to Scopus and 752 (around 24.5%) were unique to Dimensions. Based on these figures, we estimate that Dimensions covers roughly 73.8% of the content indexed by Scopus, whereas Scopus includes approximately 71.2% of the content indexed by Dimensions. These findings emphasize the complementary nature of the two sources and support their combined use to achieve a more comprehensive retrieval of AI-related literature in dentistry.

After deduplication and filtering, the resulting dataset contained 3,170 unique abstracts, each linked to its unique DOI, title, source, and publication year. It is freely available in the accompanying GitHub repository¹.

¹ <https://github.com/Litvinova1984/Topic-Modeling-for-AI-in-Dentistry/tree/main> (accessed on August 10, 2025).

Table 1. Summary Statistics of the Combined Dataset

Таблица 1. Статистика объединенного датасета

Source	Initial Records	With Abstracts	After Filtering for DOI	After Merging and Deduplication	After final filtering
Dimensions	2710	2629	2609	3191	3170
Scopus	2649	2649	2514		

This corpus serves as the foundation for all subsequent topic modeling analyses. It contains 478,257 tokens, with a mean document length of 143 tokens and a median length of 138 tokens.

It is should be noted that abstracts were selected rather than full texts because they are uniformly available across publishers, provide comparable summaries of research contributions, and allow for the construction of a large, consistent dataset suitable for systematic topic modeling comparison.

3.2. Topic modelling approaches

To thoroughly investigate and assess the underlying thematic structures in the literature on AI in dentistry, we utilized and compared three distinct topic modeling techniques. Each method is based on a different algorithmic framework, ranging from probabilistic generative models to neural embeddings and large language models. A detailed description of each approach is provided below.

3.2.1. Structural Topic Modeling

To analyze topic prevalence and content variation over time, we employed Structural Topic Modeling (STM) using the *stm* R package. We selected STM over other topic modeling methods because it integrates metadata, enabling us to examine how these factors influence topic prevalence. Additionally, incorporating metadata enhances the interpretability of the topics (Roberts et al., 2019). The publication year was included as a covariate in the prevalence model.

Before conducting topic modeling, all abstracts were thoroughly preprocessed to ensure consistency and improve the

interpretability of the model. The entire workflow was performed in R, using the *quanteda*, *stringi*, and *text2vec* packages.

Abstracts were tokenized and cleaned through several steps. First, all text was converted to lowercase, and numerical tokens were removed. Stopwords were eliminated using the standard English stopwords list provided by *quanteda*, which we enhanced with a domain-specific list of high-frequency terms that do not aid in topic discrimination.

To better capture the multi-word expressions common in our dataset, we identified and combined collocations (bigrams) using the *textstat_collocations* function from the *quanteda.textstats* package. Bigrams were selected based on a minimum frequency of 50 and a lambda value greater than 3, resulting in 252 compound terms. Examples include deep learning ($n = 3,071$, $\lambda = 7.55$), neural network ($n = 717$, $\lambda = 6.45$), and panoramic radiographs ($n = 558$, $\lambda = 6.13$). These were incorporated into the token stream using the *tokens_compound* function from *quanteda*.

We deliberately avoided stemming and lemmatization to preserve domain-specific terminology, particularly technical compounds and clinical expressions that are essential for accurate topic interpretation.

The final document-feature matrix (DFM) for STM modeling was generated from the compounded token object, including only features (unigrams and bigrams) that appeared in at least 20 documents. This frequency threshold yielded 2,623 unique terms, striking a balance between vocabulary richness and model sparsity.

We experimented with various numbers of topics ($K = 10$ to 50), selecting the optimal

value of K based on semantic coherence and exclusivity metrics.

The key outputs of the STM analysis include a ranked list of high-probability words for each topic, a set of FREX words that balance frequency and exclusivity to improve interpretability, and a list of ten representative abstracts most strongly associated with each topic.

The posterior document–topic probability matrix (θ) was used to assign each abstract to the topic with the highest probability.

We also calculated the average prevalence of each topic across the corpus, providing insights into their overall thematic prominence.

To enhance topic interpretability and reduce the need for manual labeling, we implemented GPT-based topic labeling, which has proven to be an effective strategy (Kozłowski, 2024). For each topic in the final STM model, we provided the language model (GPT-4o) with the top 10 FREX terms and the three most representative abstracts for that topic (as output by STM), along with a prompt requesting a short topic label (1–3 words) and a brief description (one sentence).

The labeling process employed a structured prompt to ensure consistency and clarity (see Figure 1). For reproducibility, the exact prompts and representative model outputs are available in the accompanying repository.

Figure 1. Prompt for topic labeling via GPT-4o (STM)

Рисунок 1. Промт для наименования топигов с использованием модели GPT-4o (STM)

```
get_stm_gpt_label <- function(topic_id, top_words, abstracts, total_topics = 30) {
  prompt <- paste0(
    "You are a scientific expert reviewing STM topic models in dentistry and AI.\n\n",
    "Your task is to assign:\n",
    "1. A short, highly specific label (1-3 words) that uniquely describes this topic, avoiding generic phrases like 'AI in Dentistry'.\n",
    "2. A distinctive one-sentence description of what this topic is about.\n\n",
    "This is topic ", topic_id, " out of ", total_topics, ". Your labels must not overlap in meaning with other topics.\n\n",
    "Top terms: ", top_words, "\n\n",
    "Here are representative abstracts:\n", paste(abstracts, collapse = "\n\n"),
    "\n\nFormat your reply as:\nTopic Label: <LABEL>\nDescription: <ONE SENTENCE>"
  )
}
```

3.2.2. Embedding-Based Topic Modeling Using Leiden Clustering

To complement the STM approach and offer an alternative embedding-based clustering method, we applied Leiden community detection to document embeddings. Each abstract was embedded using OpenAI's text-embedding-3-small model, producing a 1,536-dimensional vector per abstract. The embeddings were retrieved via API and cached locally. These embeddings capture semantic relationships and offer a high-dimensional representation ideal for clustering and similarity analysis.

To construct a neighborhood graph for clustering, we used Uniform Manifold Approximation and Projection (UMAP) with $n_neighbors$ set to 15, min_dist set to 0.1, as these parameters effectively balance the

preservation of local structure with global interpretability.

We performed Leiden clustering using the `leidenbase` package, accessed via the `igraph` and `uwot` UMAP interfaces, on a k -nearest neighbor graph constructed from the reduced embeddings. This approach parallels the BERTopic pipeline in structure but employs a graph-based clustering method instead of the density-based HDBSCAN algorithm used in BERTopic. The Leiden method offers improved resolution control and yields more stable clusters, particularly in high-dimensional spaces.

This method, similar to BERTopic, involves three stages: (i) grouping semantically similar documents within the embedding space, (ii) clustering these groups, and (iii) extracting representative keywords using class-based TF-IDF. The key difference

lies in the clustering step: while BERTopic utilizes density-based clustering (HDBSCAN), the Leiden algorithm identifies communities within a graph representation. This approach offers greater control over resolution and stability, particularly in high-dimensional spaces with varying density. Additionally, the Leiden method integrates seamlessly with our R-based workflow and enables direct comparison with STM by using the same GPT-generated topic labeling technique.

We chose Leiden clustering over BERTopic because it provides greater stability and better resolution control, especially when handling variable densities in high-dimensional embedding spaces (Shapurian, 2024).

While STM assigns each document a probabilistic mixture of topic proportions across 30 latent topics, the Leiden clustering method partitions the corpus into distinct groups, with each document assigned to exactly one semantically coherent cluster based on its position in the UMAP-projected embedding space. STM topics are model-

derived latent themes optimized to explain word co-occurrence patterns within the corpus, whereas Leiden clusters represent semantic proximity in the contextual embedding space, highlighting how documents relate to one another based on large language model representations.

Overall, Leiden clustering complements STM by providing an embedding-level perspective on document similarity that is independent of word frequency modeling. This approach facilitates the comparison and cross-validation of topics and offers an alternative framework for downstream tasks such as zero-shot labeling, cluster-level summarization, and alignment with GPT-generated topic descriptors.

We also performed GPT-based cluster labeling. For each topic in the final model, we provided GPT with the top 10 representative keywords calculated using class-balanced TF-IDF, three abstracts, and a prompt requesting a short label (1–3 words) along with a brief description (one sentence) for each cluster (see Figure 2).

Figure 2. Prompt for topic labeling via GPT-4o (embedding-based model)

Рисунок 2. Промт для наименования топигов с использованием модели GPT-4o (кластеризация эмбедингов)

```
get_gpt_label_distinct <- function(cluster_id, top_words, abstracts, total_clusters = 32) {
  prompt <- paste0(
    "You are a scientific expert reviewing scientific document clusters in dentistry and AI.\n\n",
    "Your task is to assign:\n",
    "1. A short, highly specific label (1-3 words) that uniquely describes this cluster, avoiding generic phrases like 'AI in Dentistry'.\n",
    "2. A distinctive one-sentence description of what this topic is about.\n\n",
    "You are working on cluster ", cluster_id, " out of ", total_clusters, ". Your labels must not overlap in meaning with other clusters.\n\n",
    "Top terms: ", top_words, "\n\n",
    "Here are representative abstracts:\n", paste(abstracts, collapse = "\n\n"),
    "\n\nFormat your reply as:\nTopic Label: <LABEL>\nDescription: <ONE SENTENCE>"
  )
}
```

3.2.3. Zero-shot LLM-based Topic Modeling

Inspired by TopicGPT (Pham et al., 2024), we utilized GPT-4o in a zero-shot setting to generate interpretable topics directly

from the corpus without requiring predefined topics or model training. The corpus was divided into 159 batches, each containing 20 abstracts, and each batch was submitted to GPT-4o using a structured prompt (Figure 3).

Figure 3. Prompt for topic extraction via GPT-4o

Рисунок 3. Промт для извлечения топигов с использованием модели GPT-4o

```
# === STEP 2: Make GPT prompt from abstracts ===
make_prompt <- function(texts_batch) {
  joined_texts <- paste(texts_batch, collapse = "\n\n")
  prompt_text <- paste(
    "You are a scientific topic modeling assistant.",
    "Analyze the following research abstracts and extract 3 coherent but diverse topics that best describes them.",
    "Provide:\n1. Topic label (max 5 words)\n2. A brief 1-2 sentence description\n3. 5-10 representative keywords\n\nAbstracts:\n\n",
    joined_texts
  )
  list(list(role = "user", content = prompt_text))
}
```

A total of 417 local topics were extracted. To minimize redundancy and facilitate comparison across methods, we embedded each GPT-generated topic profile (including label, keywords, and description) using 1,536-dimensional OpenAI embeddings. We then applied k-means clustering ($k = 30$) to these embeddings to consolidate them into global topics. Each resulting cluster was labeled by GPT and stored as a topic profile by combining the final label, keywords, and description.

Topic labels generated by each of the three approaches were independently evaluated by three human annotators (the authors). All labels received high ratings (4-5 on 1-5 scale) from each annotator and were retained for downstream analysis.

3.2.4. Cross-Approach Comparison

To facilitate cross-model comparison and downstream interpretation, we mapped topics from the STM and embedding-based clusters into a common GPT topic space. This approach enabled unified labeling, visual alignment, and consistent topic tracking across models. To compare and consolidate topic assignments from STM, GPT-extracted topics, and embedding-based clusters, we employed a two-stage process: (1) topic alignment and (2) document-level consensus labeling.

In the first stage, we focused on aligning the topics themselves, independent of document assignments. Using STM topics as the reference set, each topic was compared to

all topics from GPT and embedding-based clustering through three complementary similarity measures: TF-IDF cosine similarity, embedding-based cosine similarity, and a blended similarity score, as described below.

1. Topic Representation Construction. Each topic, derived from STM, Embeddings, and GPT, was represented in two ways: a) a textual profile, which combined the topic label, top keywords (such as FREX words or TF-IDF terms), and a description; and b) a semantic embedding profile, generated by computing a 1,536-dimensional OpenAI text-embedding-3-small vector based on the textual profile. This approach enables the calculation of high-dimensional semantic similarity between topic descriptions.

2. Similarity Metrics. To compare topic pairs across models, we calculated two types of similarity measures: (1) TF-IDF cosine similarity, where we constructed TF-IDF vectors for all topic profiles using a shared vocabulary and computed pairwise cosine similarities between each pair of approaches; and (2) embedding cosine similarity, where we used OpenAI embedding vectors for the topic profiles to compute semantic cosine similarities between topics. For each pair of approaches (STM↔GPT, STM↔EMB, GPT↔EMB), we computed the full all-pairs matrices of TF-IDF cosine and embedding cosine similarities between topic textual profiles and combined them into the Blended Topic Similarity Score (BlendSim) (Equation 1):

$$S_{\text{topic}}(t_i^{(A)}, t_j^{(B)}) = \frac{1}{2} (\cos_{\text{TF-IDF}}(t_i^{(A)}, t_j^{(B)}) + \cos_{\text{Embedding}}(t_i^{(A)}, t_j^{(B)})) \quad (\text{Equation 1})$$

where $t_i^{(A)}, t_j^{(B)}$ are topics i and j from approaches **A** and **B**, costf-idf is a cosine similarity between TF-IDF vectors of the textual profiles, cosEmbedding is a cosine similarity between embedding vectors of the textual profiles. The factor $1/2$ indicates equal weighting of the two components.

While Equation (1) provides a topic-level alignment score based solely on textual profiles, it does not guarantee agreement at the document level. Two topics may appear

lexically and semantically similar, yet group different sets of documents. To establish a document-level consensus, we therefore define a second metric, the **Document Consensus Score (DocCons)** S_{blend} , which combines structural overlap, measured by the Jaccard index of document sets, with semantic fit, measured as the mean cosine similarity between shared documents and the candidate topic's centroid (Equation 2):

$$S_{\text{blend}}(A, B) = w_J \cdot J(A, B) + w_{\text{cos}} \cdot \overline{\text{cos}}(A, B), \quad (\text{Equation 2})$$

Here, $J(A, B)$ denotes the Jaccard index of the document sets assigned to topics **A** and **B**. The semantic component $\text{cos}(A, B)$ is the mean cosine similarity between the L2-normalized embeddings of documents in the intersection $C_A \cap C_B$ and the centroid of topic **B** (defined as the normalized mean of its document embeddings). If no documents overlapped, the cosine term was set to zero. We fixed $w_J = w_{\text{cos}} = 0.5$ to weight both components equally, and each STM topic was aligned to the candidate topic with the highest score; documents then inherited this aligned label.

Importantly, Eq. (2) serves not only as a similarity measure but also as the operational basis for consensus topic assignment: it determines which candidate topic provides the best match, and all documents inherit this unified label for further analysis.

We then assessed the consistency of consensus topics across the three modeling approaches. At the document level, we identified three categories of alignment: topics were considered stable if the STM, GPT, and embedding-based assignments all converged on the same GPT-space label for a document; partially aligned topics were those where STM agreed with exactly one other model (either GPT or embedding-based), but not both; and method-specific topics were

those related to documents for which no alignment was observed across the methods.

To evaluate model performance, we employed several quantitative metrics to assess both internal quality and inter-topic distinctiveness. Compactness was measured using the mean Euclidean distance to each topic's centroid (mean_own_dist), representing the average distance between each document and its assigned topic centroid—lower values indicate tighter topic clusters. Additionally, we calculated the mean cosine similarity to the assigned centroid (mean_own_cos), where higher values reflect greater semantic coherence within topics. To assess topic separation (distinctiveness), we computed the mean nearest-neighbor centroid distance (mean_nn_other_dist), which is the average Euclidean distance from each topic centroid to its closest neighboring centroid; higher values suggest better topic separation. We also measured the mean margin (mean_margin), defined as the difference between a document's distance to its own topic centroid and to the nearest other centroid—higher margins indicate that documents are both close to their own topic and well separated from competing topics. Furthermore, we included standard metrics such as the Davies–Bouldin Index (DB) and the Dunn Index, which capture the ratio of intra-topic compactness to inter-topic separation.

In this context, a “centroid” does not represent a geometric figure with a homogeneous or symmetric shape but simply the average vector of all document representations assigned to a topic. Each topic thus corresponds to a cloud of points in high-dimensional space, which may be heterogeneous and anisotropic. The centroid serves as a representative point to summarize this set and to compute compactness and separation metrics. Distances were measured using Euclidean norms on L2-normalized vectors, which are directly related to cosine similarity. This approach provides an interpretable and computationally efficient approximation of intra-topic cohesion and inter-topic distinctiveness without assuming geometric symmetry of the clusters.

This multi-metric evaluation framework enabled us to objectively compare STM, embedding-based, and GPT-based topics in terms of both internal coherence and semantic distinctiveness.

3.2.5. Trend analysis across models

To evaluate how the research focus evolved over time, we applied a consistent procedure across all three modeling approaches: STM, embedding-based clustering, and GPT-derived topics.

For STM topics, we determined the number of documents in which each topic was dominant for each publication year, defining dominance as the highest STM probability within a document. For each topic, we constructed yearly time series of (i) dominant document counts and (ii) mean STM probabilities. Temporal changes were quantified by fitting simple linear regressions of each metric against publication year, with slope coefficients extracted as trend estimates. Trends were classified as increasing (\uparrow), decreasing (\downarrow), or stable ($-$) depending on whether slopes exceeded predefined thresholds (± 0.0005 for probability; ± 0.5 documents per year).

These thresholds were selected to capture changes that are small enough to be statistically detectable over our time span, yet

large enough to be practically meaningful (e.g., an average shift of at least half a document per year or a consistent increase of ≥ 0.5 percentage points in topic probability over a decade). A topic was defined as emergent if both STM probability and the dominant document counts exhibited positive slopes exceeding the threshold.

For embedding-derived clusters, topic prevalence was measured as the mean cosine similarity between each document and the centroid of its assigned cluster for each publication year, along with the number of documents assigned to each cluster. Cosine similarity captures the semantic alignment of documents with the cluster core, analogous to STM topic probabilities. Yearly averages of both metrics were regressed on publication year, and the slope coefficients provided quantitative estimates of temporal trends. Unlike STM probabilities, cosine similarities do not sum to one across clusters; however, their changes over time (Δ) are interpretable in parallel with STM prevalence.

For GPT-derived topics, we used two indicators: the mean document-centroid cosine similarity per year and document frequency. Regression-based slope estimates of these annual metrics allowed us to classify GPT topics as increasing, decreasing, or stable.

For the visualization and interpretation of thematic trends, we based the analysis on STM→GPT blended mapping.

For visualization, yearly proportions were smoothed using locally estimated scatterplot smoothing (LOESS) with a span of 0.75. LOESS provides a nonparametric regression fit that highlights overall temporal patterns while reducing noise from year-to-year fluctuations.

4. RESULTS

4.1. Structural topic modeling

We experimented with a range of topic numbers ($K = 10$ to 50) and selected $K = 30$ as optimal based on three criteria: 1) plateaus in held-out likelihood and residuals, both of which stabilize around $K = 30$; 2) stabilization

of semantic coherence and exclusivity, noting that semantic coherence declines sharply from $K = 10$ to 20 but remains relatively stable after $K = 30$, indicating no benefit in increasing K further; and 3) qualitative interpretability of topics (see Figure 1S). This model serves as the baseline for comparison

with newer methods.

All metrics expressing dynamic trends of STM topics were tabulated in a dynamic trend table (Table 2) and visualized (Figure 2S) to enable direct comparison between model-based prevalence and observed document-level dominance.

Table 2. Top 10 Most Dynamic STM Topics

Таблица 2. 10 наиболее динамичных тем (модель STM)

Rank	Topic Label	Δ Doc Count	Δ STM Prob	STM Prob Slope	Doc Count Slope	STM Prob Trend	Doc Count Trend	Emergent
1	AI Integration Challenges	92	0.21	0.066	34.1	↑	↑	TRUE
2	Dental Caries Detection	61	0.052	0.003	6.214	↑	↑	TRUE
3	Deep Learning Model Performance	56	0.093	-0.014	9.657	↓	↑	FALSE
4	Systematic AI Review	52	0.07	0.008	7.571	↑	↑	TRUE
5	Multi-View Segmentation	48	0.08	-0.009	8.5	↓	↑	FALSE
6	Panoramic Radiograph Analysis	46	0.104	-0.006	5.286	↓	↑	FALSE
7	Tooth Segmentation Techniques	45	0.172	0.025	7.929	↑	↑	TRUE
8	Chatbot Response Evaluation	42	0.223	0.028	8.786	↑	↑	TRUE
9	Dental AI Applications	40	0.086	-0.005	3.929	↓	↑	FALSE
10	Dental Image Detection	27	0.159	0.008	4.107	↑	↑	TRUE

The topics highlighted in red (Panel A) and green (Panel B) in Figure 2S can be classified as emergent, as they exhibit consistent upward trends in both STM topic prevalence and real-world document dominance. Notably, *AI Integration Challenges*, *Chatbot Response Evaluation*, *Dental Image Detection*, and *Tooth Segmentation Techniques* show clear concurrent increases in both dimensions, indicating rapidly growing attention in the

field. These topics, with synchronized growth in both measures, may represent emerging research frontiers in AI applications in dentistry, according to STM classification. Additionally, *Dental Caries Detection* and *Systematic AI Review* demonstrate an increasing number of documents and a slight rise in topic prevalence.

Some topics exhibit a mismatch between their model-based importance and actual publication activity, unlike clearly

emerging topics. For instance, *Deep Learning Model Performance* and *Panoramic Radiograph Analysis* demonstrate an increasing number of documents over time, yet their average importance is slightly declining. This indicates that while more papers mention these topics, they are often used in a supporting or technical capacity rather than as the primary focus of the study.

4.2. Embedding-based clustering

To identify the optimal number of semantic clusters within the UMAP-projected embedding space, we applied Leiden clustering across a range of resolution values from 0.5 to 1.5, resulting in 24 to 38 clusters. As illustrated in Figure 3S, lower resolution settings (e.g., 0.5) produced overly coarse groupings, with some clusters covering broad and visually heterogeneous regions. In contrast, higher resolutions (e.g., 1.5) caused excessive fragmentation, generating numerous small or marginal clusters with limited separation and semantic distinctiveness.

A resolution of 1.0, which generated 33 clusters, provided the optimal balance between granularity and interpretability. At this setting, the clusters appeared visually compact and well-separated in the UMAP space, exhibiting minimal overlap and strong internal cohesion. This configuration prevented over-partitioning while effectively capturing meaningful thematic distinctions throughout the corpus. Additionally, a resolution of 1.0 aligned with the natural semantic structure of the data, facilitating cluster-level labeling and trend analysis without introducing excessive noise or sparsity in cluster sizes.

Therefore, we chose the 33-cluster solution as the most stable and interpretable representation of the embedding-based semantic structure, which serves as a foundation for downstream topic labeling and comparison with STM-derived topics.

To analyze topic dynamics in the embedding-based model, we calculated the average cosine similarity between each

document and its assigned cluster centroid for each publication year. This metric measures the semantic closeness of a document to its cluster, with higher cosine similarity values indicating a stronger alignment with the core semantics of the topic.

This approach is conceptually similar to how STM-derived topic probabilities indicate the degree of association between a document and a topic. While STM probabilities are normalized outputs from a probabilistic generative model, embedding-based cosine similarities measure angular closeness in a high-dimensional semantic space. Both metrics capture topic-document relevance, and averaging them over time offers an effective proxy for tracking topic prevalence trends in the literature.

It is important to note that cosine similarities do not sum to 1 across all clusters, unlike STM topic probabilities. Therefore, although the absolute values of these metrics are not directly comparable, the direction and magnitude of their temporal changes (Δ) can be meaningfully interpreted in parallel.

This approach enables us to consistently evaluate and compare semantic topic trends across different modeling methods by analyzing the evolution over time of the mean document-level cosine similarity per cluster (similar to STM prevalence) and the number of dominant documents per cluster (see Table 3 and Figure 4S).

Dental LLM Evaluation and *Dental AI Chatbot Evaluation* have shown significant increases in both the average topic probability and the number of dominant documents from 2022 to 2023. This trend highlights the rapid emergence of new focus areas within LLMs and conversational AI applications in dentistry. The simultaneous rise observed in both evaluations suggests that these topics are gaining popularity as well as conceptual coherence.

AI-Enhanced Orthodontics and *CBCT Segmentation Accuracy* have shown a steady increase in the number of documents over time, accompanied by gradual but consistent improvements in cosine similarity. These

trends highlight a growing interest and a clearer conceptual understanding of imaging-based AI topics. The simultaneous rise suggests that these areas are becoming both more prevalent and better defined.

Caries Detection Models exhibits a significant rise in the number of documents, yet the average cosine similarity shows only modest or minimal change. This likely indicates a wider range of coverage within this topic across diverse subfields, resulting in less concentrated semantic clustering.

CBCT Artifact Reduction demonstrates relatively stable or even declining semantic

alignment, accompanied by a plateau in the number of documents. This trend may suggest that the research area is reaching maturation or saturation, resulting in fewer novel contributions.

Topics such as **Radiographic AI Diagnostics**, **Panoramic Tooth Segmentation**, and **Periodontal AI Diagnostics** demonstrate consistent, gradual growth across both panels. These areas likely represent well-established fields with steady progress rather than sudden, breakthrough trends.

Table 3. Top 10 Most Dynamic Embedding Topics

Таблица 3. 10 наиболее динамичных тем (модель эмбедингов)

Rank	Embedding Topic Label	Δ Doc Count	Δ Emb_cos	Δ Emb_cos (Slope)	Δ Doc Count Slope	Emb Trend	Cos Trend	Doc Count Trend	Emergent
1	Periodontal AI Diagnostics	66	0.028	0.004	11.68	↑	↑	↑	TRUE
2	AI-Enhanced Orthodontics	56	0.038	0.007	8.25	↑	↑	↑	TRUE
3	Caries Detection Models	55	0.029	0.003	6.29	↑	↑	↑	TRUE
4	CBCT Segmentation Accuracy	54	0.02	0.002	9.39	↑	↑	↑	TRUE
5	Dental LLM Evaluation	48	0.102	0.023	9.28	↑	↑	↑	TRUE
6	Dental AI Chatbot Evaluation	44	0.12	0.018	7.75	↑	↑	↑	TRUE
7	Dental Implant Diagnostics	42	0.06	0.007	5.79	↑	↑	↑	TRUE
8	Radiographic AI Diagnostics	40	0.014	0.001	6.36	↑	↑	↑	TRUE
9	Panoramic Tooth Segmentation	37	0.018	0.001	5.46	↑	↑	↑	TRUE
10	CBCT Artifact Reduction	30	0.032	0.003	-0.13	↑	—	—	FALSE

These trends indicate that embedding-based clustering effectively captures both emerging innovations and established subfields. Notably, topics related to large language models and interactive AI tools, such as chatbots, are exhibiting clear signs of rapid thematic consolidation and increasing publication volume. At the same time, imaging-related topics like cone-beam computed tomography (CBCT) and segmentation remain prominent and stable, although some may be approaching saturation.

As shown in Panel A of Figure 4S, several embedding clusters exhibit increasing average cosine similarity over time, reflecting patterns observed in STM topic probabilities. This suggests a growing thematic convergence in both modeling approaches. It is important to note that while the embedding model reveals some fine-grained clusters compared to STM, it also occasionally separates semantically similar topics into different clusters.

4.3. GPT-based topic extraction

To complement the STM and embedding-based topic models, we employed a zero-shot, LLM-based topic modeling approach using GPT-4o to generate topics

directly from scientific abstracts. Abstracts were organized into batches of 20 documents each, and GPT was prompted to produce three distinct topic descriptions per batch. Of the 159 total batches, 139 (88%) were successfully processed, yielding 417 batch-level topics that summarize 2,780 abstracts.

To minimize semantic redundancy among GPT-generated topics such as paraphrased labels representing similar themes, we embedded each topic's label and description using OpenAI's embedding model. Since graph-based Leiden clustering yielded only a few communities (6-7) at reasonable resolutions (0.5–1.5), we applied K-means clustering with $k=30$ to achieve a higher-resolution partition better suited for downstream comparison. Each global cluster was then labeled and summarized using GPT, resulting in unified topic labels, descriptions, and representative keywords for easier interpretation.

To assess thematic trends in GPT-derived topics, we analyzed changes over time in both average topic prevalence measured as the mean cosine similarity between each document and its assigned cluster centroid for each publication year and document frequency, defined as the number of documents assigned to each topic (see Table 4 and Figure 5S).

Table 4. Top 10 Most Dynamic Embedding Topics (GPT-based)

Таблица 4. 10 наиболее динамичных тем (модель GPT)

Rank	GPT Topic Label	Δ Doc Count	Δ Cos	Δ Document Count (Slope)	Δ Mean Cos (Slope)	GPT Cos Trend	Doc Count Trend	Emergent
1	Advanced Dental Image Segmentation Techniques	45	0.042	10.607	0.006	↑	↑	TRUE
2	AI-Enhanced Dental Imaging and Diagnostics	43	0.022	9.429	0.003	↑	↑	TRUE
3	AI in Dental Diagnostics and Treatment	43	-0.021	7.571	-0.001	↓	↑	FALSE

4	AI-Driven Dental Disease Diagnostics	40	0.04	10.214	0.006	↑	↑	TRUE
5	AI Integration in Dental Education and Practice	40	0.012	7.59	0,003	↑	↑	TRUE
6	Deep Learning in Dental Imaging and Diagnostics	35	0.004	6.214	0.002	↑	↑	TRUE
7	AI-Driven Dental Diagnostics and Treatment	32	0.089	6.893	0.015	↑	↑	TRUE
8	AI-Driven Oral Cancer Detection and Diagnosis	30	0.002	8.5	0.001	↑	↑	TRUE
9	AI in Orthodontics and Maxillofacial Analysis	25	0.029	5.393	0,009	↑	↑	TRUE
10	Automated Orthodontic and Forensic Image Analysis	21	-0.078	6.464	-0.003	↓	↑	FALSE

Several topics demonstrated a consistent increase in both prevalence and frequency, including *Advanced Dental Image Segmentation Techniques*, *AI in Dental Diagnostics and Treatment*, and *Deep Learning in Dental Imaging and Diagnostics*. These patterns suggest sustained and expanding interest. Other topics, such as *Automated Orthodontic and Forensic Image Analysis*, exhibited high average relevance despite fewer documents, indicating a more specialized yet focused research trajectory. *AI-Driven Dental Disease Diagnostics* showed strong topic probability with several document frequency peaks, reflecting a maturing area that retains scientific relevance.

Fluctuating trajectories were observed for topics such as *AI in Orthodontics and Maxillofacial Analysis*, which peaked around 2023–2024, and *AI-Driven Oral Cancer Detection and Diagnosis*, which showed recent increases in document count but less consistent prevalence trends. Notably, *AI Integration in Dental Education and Practice* emerged prominently post-2022 in both metrics, underscoring the growing interest in educational and clinical adoption of AI technologies.

Together, these results highlight both expanding thematic areas and focused research niches within AI applications in

dentistry, as captured by GPT-based zero-shot topic modeling.

4.4. Topic Model Quality

To evaluate the internal quality of topic assignments, we systematically assessed the compactness, distinctiveness, and redundancy of the clusters generated by three different topic modeling approaches.

Across all quantitative clustering metrics, STM outperformed the other methods in terms of topic compactness, coherence, and separation. It achieved the lowest mean Euclidean distance to centroids ($\text{mean_own_dist} = 0.225$) and the highest cosine similarity ($\text{mean_own_cos} = 0.775$), indicating the most internally consistent topics. Additionally, STM recorded the highest positive separation margin ($\text{mean_margin} = 0.013$) and the best silhouette score ($\text{mean_sil} = 0.056$), demonstrating an optimal balance between within-topic cohesion and between-topic distinctiveness. Its Davies–Bouldin (8.40) and Dunn (0.063) indices further confirmed the robustness of the cluster quality.

The embedding-based Leiden clustering method ranked second on most metrics, generating moderately compact topics ($\text{mean_own_cos} = 0.765$) while preserving positive separation ($\text{mean_margin} = 0.004$), albeit with slightly more overlap compared to STM.

In contrast, zero-shot GPT-based topic modeling demonstrated weaker internal cohesion ($\text{mean_own_cos} = 0.751$) and a negative separation margin ($\text{mean_margin} = -0.005$), indicating that, for some topics, documents were closer to a different centroid than their own. This was reflected in its lowest silhouette score (-0.014) and the poorest cluster validity indices ($\text{DB} = 13.75$;

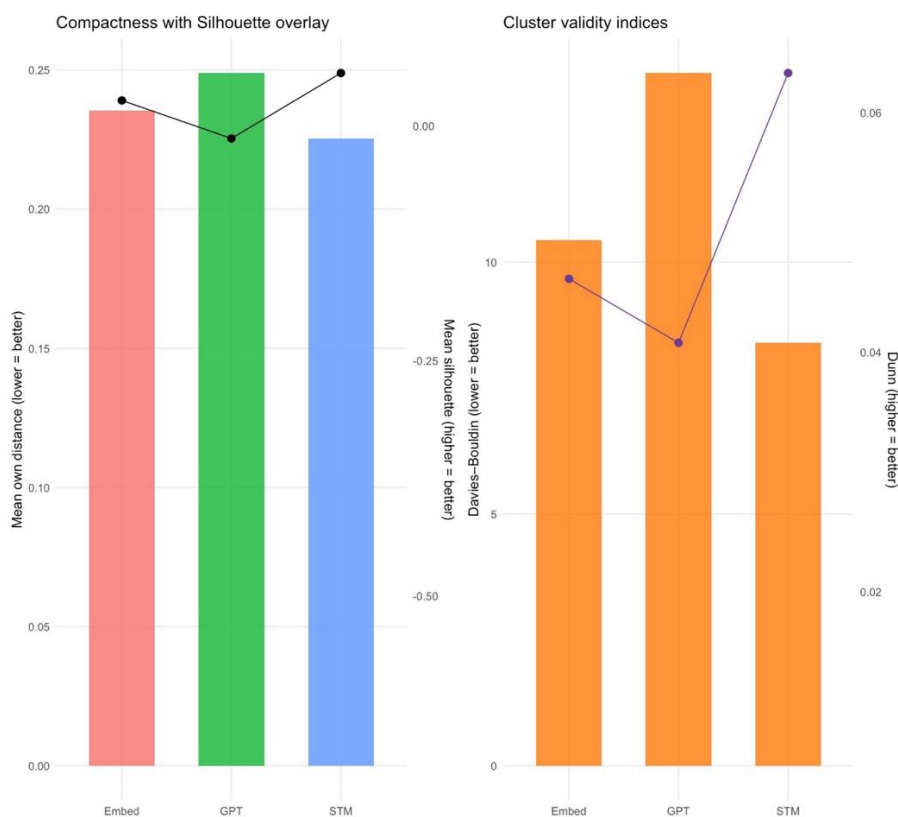
$\text{Dunn} = 0.041$), suggesting significant overlap among GPT-generated topics despite their interpretive clarity.

Overall, these metrics show that STM generated the most compact and well-separated topic structures, followed by the embedding-based approach. In contrast, GPT-based clustering was less structurally distinct and more susceptible to thematic redundancy. These findings are visually summarized in Figure 4, which compares compactness (mean document–centroid distance and silhouette), and distinctiveness (DB and Dunn indices) across models. Together, the results support STM as the most robust model for delineating clear, compact topics in this domain.

To statistically assess differences in clustering quality across approaches, we performed one-way ANOVAs for each evaluation metric (mean document–centroid cosine, Euclidean distance, margin, silhouette, Davies–Bouldin, and Dunn indices), treating modeling approach (STM, GPT, Embed) as the grouping factor. Post hoc pairwise comparisons were adjusted using the Benjamini–Hochberg false discovery rate (FDR) procedure. These tests confirmed that STM outperformed GPT on all core compactness and separation metrics (adjusted $p < 0.05$), with embedding clusters occupying an intermediate position.

Thus, STM generates the most compact and distinct topic structures, making it an excellent choice for applications that require non-overlapping and easily interpretable topics. The embedding model falls in the middle, producing coherent but somewhat broader themes. GPT, on the other hand, produces the loosest and most overlapping partitions, which may be more suitable for exploratory synthesis rather than precise topic delineation.

Figure 4. Compactness and distinctiveness of topic assignments across methods
Рисунок 4. Компактность и разделимость результатов моделирования по методам



4.5. Cross-Approach Topic Correspondence

Heatmap comparisons of topic label similarity (Figure 5) revealed both strong one-to-one matches and notable divergences among STM, embedding-based clustering, and GPT-labeled topics. TF-IDF similarity highlighted alignments driven by lexical overlap in topic keywords and descriptions. In STM-Embed comparisons, several topics exceeded 0.50 similarity, indicating consistent terminology use across methods. However, STM-GPT and Embed-GPT pairs showed low levels of lexical matches, reflecting GPT's tendency to generate more narrative-style labels.

Embedding-based similarity revealed a different pattern: many topic pairs had similarity scores exceeding 0.90 despite low

TF-IDF values. This was especially noticeable in STM-GPT matches, where semantically equivalent topics were described using different terms (e.g., “Dental Caries Detection” vs. “AI-Driven Dental Disease Diagnostics”).

The Blended Topic Similarity score, which combines TF-IDF and embedding similarity measures, provides a balanced perspective (see Table 5).

Overall, STM and embedding-based clustering achieved the highest levels of both lexical and semantic alignment. In contrast, GPT-labeled topics exhibited greater variability in expression but still demonstrated semantic alignment with several STM and embedding clusters. As clearly shown in Table 5, GPT topics tend to be more semantically broad.

Figure 5. Topic Label Similarity
Рисунок 5. Близость названий тем

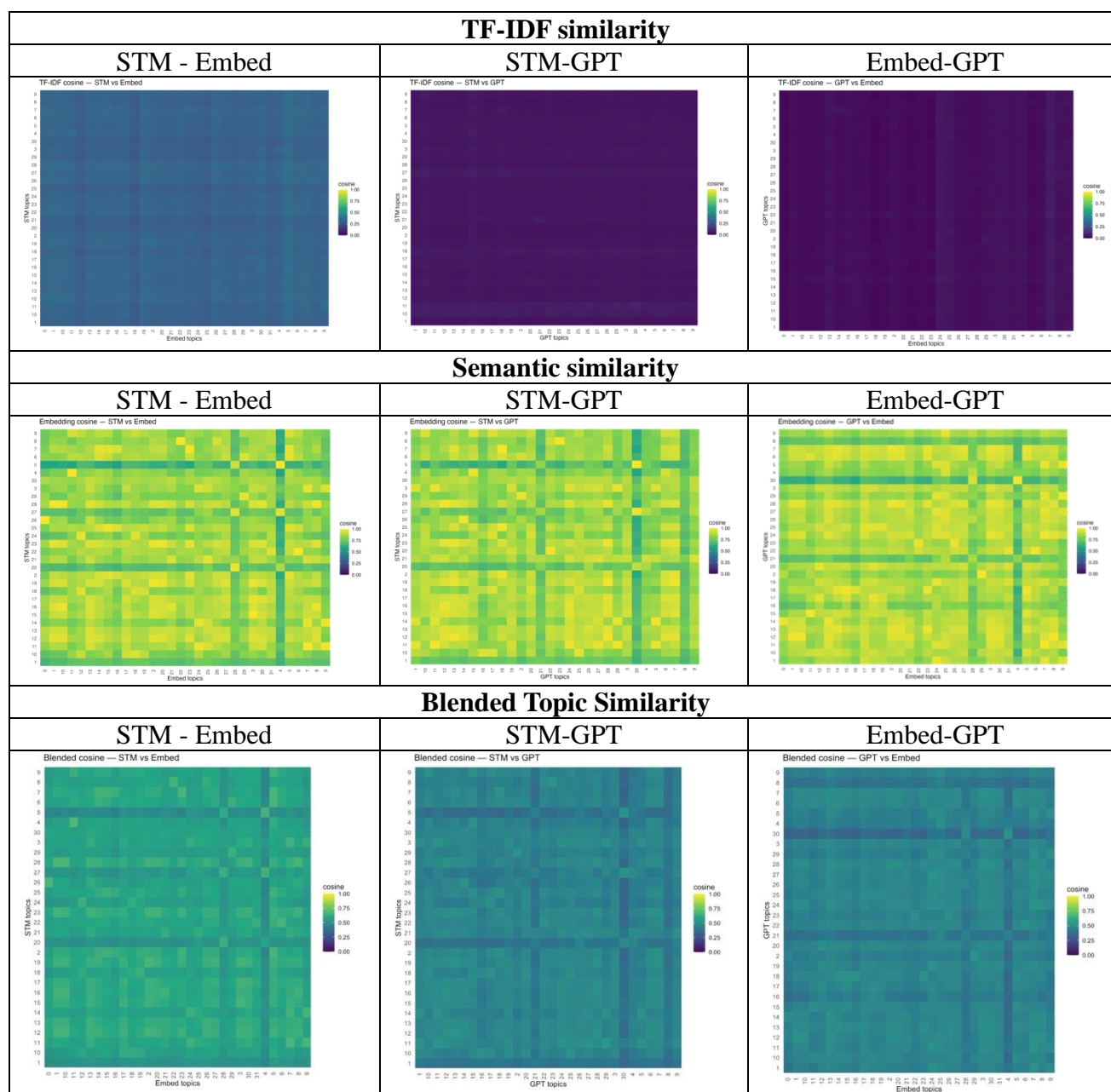


Table 5. Topic alignment via approaches

Таблица 5. Соответствие тем между подходами

Label_Embed	BlendSim (STM-embed)	Label_STM	BlendSim (STM_GPT)	Label_GPT	BlendSim (EMB_GPT)	Label_Embed
Periodontal AI Diagnostics	0.667	AI Integration Challenges	0.528	AI Integration and Ethics in Dental Care	0.507	AI-Enhanced Orthodontics
Panoramic Tooth Segmentation	0.662	Tooth Segmentation Techniques	0.524	Advanced Dental Image Segmentation Techniques	0.496	Dental Image Segmentation
Dental AI Chatbot Evaluation	0.655	Chatbot Response Evaluation	0.521	AI Integration in Dental Education and Practice	0.514	Dental AI Chatbot Evaluation
CBCT Artifact Reduction	0.677	Metal Artefact Reduction	0.517	Advanced Image Processing in Dental Imaging	0.51	CBCT Artifact Reduction
Dental Caries Detection	0.672	Dental Caries Detection	0.515	AI-Driven Dental Disease Diagnostics	0.504	Dental Caries Detection
Cephalometric Landmark Detection	0.661	Automated Cephalometric Landmark Detection	0.515	Automated Orthodontic and Forensic Image Analysis	0.506	Cephalometric Landmark Detection
AI-Driven Dental Innovations	0.654	Dental AI Applications	0.514	AI in Medical Imaging and Diagnostics	0.5	AI-Driven Dental Innovations
Panoramic Radiograph AI	0.669	Panoramic Radiograph Analysis	0.513	Deep Learning in Dental Imaging and Diagnostics	0.495	Panoramic Tooth Segmentation
Dental Age Estimation	0.652	Forensic Dental Age Estimation	0.512	AI in Forensic and Orthodontic Dentistry	0.499	Forensic Dental AI
Mandibular Third Molar AI	0.633	Deep Learning Model Performance	0.511	Deep Learning in Dental Imaging	0.495	Panoramic Tooth Segmentation
Oral Cancer Detection	0.668	Oral Cancer Detection	0.511	AI-Driven Oral Cancer Detection and Diagnosis	0.494	Oral Cancer Detection
Dental Implant Diagnostics	0.625	Digital Implant Surgery	0.5	3D Printing Innovations in Dentistry	0.47	Alveolar Ridge Techniques
Radiotherapy Auto-Segmentation	0.635	Head-Neck IMRT Automation	0.5	AI-Enhanced Cancer Diagnostics and Radiotherapy	0.52	Radiotherapy Auto-Segmentation
CBCT Segmentation Accuracy	0.652	CBCT Image Enhancement	0.5	Advanced Dental Imaging Technologies	0.482	CBCT Segmentation Accuracy
Caries Detection Models	0.656	Pediatric Plaque Detection	0.5	AI-Driven Dental Diagnostics and Treatment	0.483	Caries Detection Models
3D Tooth Segmentation	0.647	Tooth Segmentation	0.5	AI-Enhanced Dental Imaging and Diagnostics	0.502	Dental Image Segmentation
Dental Crown	0.631	AI Crown Design	0.495	Innovations in	0.486	Dental Crown

Automation				Digital Dental Technologies		Automation
AI-Enhanced Orthodontics	0.625	Systematic AI Review	0.493	AI-Driven Innovations in Dental Treatment Planning	0.478	Dental Implant Diagnostics
Dental Object Detection	0.651	Dental Image Detection	0.492	AI-Enhanced Dental Imaging and Analysis	0.485	Dental Object Detection
Plaque Detection Automation	0.618	Oral Health Disparities	0.491	AI Innovations in Dentistry and Oral Health	0.486	Caries Detection AI
Dental Anatomical Studies	0.641	Maxillofacial Reconstruction	0.491	Advanced 3D Dental Modeling and Reconstruction	0.489	3D Tooth Segmentation
Caries Detection AI	0.63	Dental Image Classification	0.49	AI and Machine Learning in Dentistry	0.475	Osteoporosis Detection AI
Radiographic AI Diagnostics	0.658	Radiographic Lesion Detection	0.488	AI in Dental Diagnostics and Treatment	0.527	Periodontal AI Diagnostics
Osteoporosis Detection AI	0.643	Deep Learning Diagnostics	0.487	Medical Innovations and Diagnostic Challenges	0.492	Radiographic AI Diagnostics
Dental Image Segmentation	0.622	Multi-View Segmentation	0.476	Advanced Dental Structure and Evolution Analysis	0.495	Dental Anatomical Studies
Dental AI Knowledge Extraction	0.574	AI Diagnostic Tools	0.471	Explainable AI and Sentiment Analysis in Healthcare	0.458	Plaque Detection Automation
Dental Image Segmentation	0.634	Deep Learning Segmentation	0.461	AI in Orthodontics and Maxillofacial Analysis	0.485	Cervical Maturation AI
Dental LLM Evaluation	0.636	GPT in Medical Exams	0.448	NLP and Text Mining in Dental Marketing	0.498	Dental AI Knowledge Extraction
Dental Lesion Deep Learning	0.606	Pathological Speech Analysis	0.418	AI-Driven Innovations in Sustainable Chemistry	0.419	Dental Lesion Deep Learning
Forensic Dental AI	0.533	Authorship Corrections	0.4	Social Media's Impact on Dental Health Information	0.44	Dental LLM Evaluation

Next, we identified conceptual overlaps between topics across different methods via the Document Consensus Score (DocCons) (Eq. 2). This involved evaluating both the overlap in document membership, using Jaccard similarity, and the semantic relatedness of topic labels and keywords. For each STM topic, we determined the most semantically aligned topic in the GPT and embedding models by combining the document-level and label-level similarities.

These pairwise mappings enabled us to trace how topics corresponded across methods. Using these mappings, we assigned each document a unified consensus label that reflected the closest matched topic across models (Eq. 2).

To demonstrate how model outputs converge through our alignment procedure, we present three examples in Table 6. In each instance, STM, embedding-based clustering, and GPT generated related yet differently

phrased topics. The STM→GPT blended alignment score which combines Jaccard document overlap and the mean per-document cosine similarity to the candidate topic was

used to identify the GPT topic closest to the STM reference topic. Subsequently, the document adopted that consensus GPT label for downstream analyses.

Table 6. Examples of topic alignment

Таблица 6. Примеры соответствия тем

Doc	STM assignment (prob)	Embed assignment (cos)	GPT assignment (cos)	STM→GPT consensus (DocCons)	Emb→GPT consensus (DocCons)
1	Panoramic Radiograph Analysis (0.264)	Tooth Segmentation Techniques (0.212)	AI-Driven Dental Disease Diagnostics (0.06)	Deep Learning in Dental Imaging and Diagnostics (0.513)	Medical Innovations and Diagnostic Challenges (0.492)
2	Dental Caries Detection (0.31)	Caries Detection AI (0.121)	AI-Driven Dental Diagnostics and Treatment (0.141)	AI-Driven Dental Disease Diagnostics (0.516)	AI Innovations in Dentistry and Oral Health (0.486)
3	Tooth Segmentation Techniques (0.212)	Panoramic Radiograph AI (0.076)	AI-Enhanced Dental Imaging and Diagnostics (0.154)	Advanced Dental Image Segmentation Techniques (0.525)	Deep Learning in Dental Imaging and Diagnostics (0.5)

For Document 1, all models highlight imaging-centric diagnostics but differ in scope ranging from a panoramic focus to a broader disease diagnostic. The STM→GPT Document Consensus Score of 0.513 identifies **Deep Learning in Dental Imaging and Diagnostics** as the closest semantic match, while EMB→GPT Document Consensus Score of 0.492 selects **Medical Innovations and Diagnostic Challenges**.

For Document 2, both the STM and Embedding models emphasize caries detection, whereas GPT adopts a broader diagnostics and treatment perspective. The Document Consensus Score STM→GPT match favors GPT's **AI-Driven Dental Disease Diagnostics** (0.516), producing a consensus label that encompasses caries within a wider diagnostic framework. Meanwhile, the EMB→GPT consensus label is even broader semantically, titled **AI**

Innovations in Dentistry and Oral Health (0.486).

For Document 3, STM focuses on tooth segmentation, GPT on imaging and diagnostics, and Embeddings on the panoramic modality. The Document Consensus Score STM→GPT (0.525) selects **Advanced Dental Image Segmentation Techniques**, aligning the document most directly with the segmentation theme, while the EMB→GPT blended label does not consider segmentation.

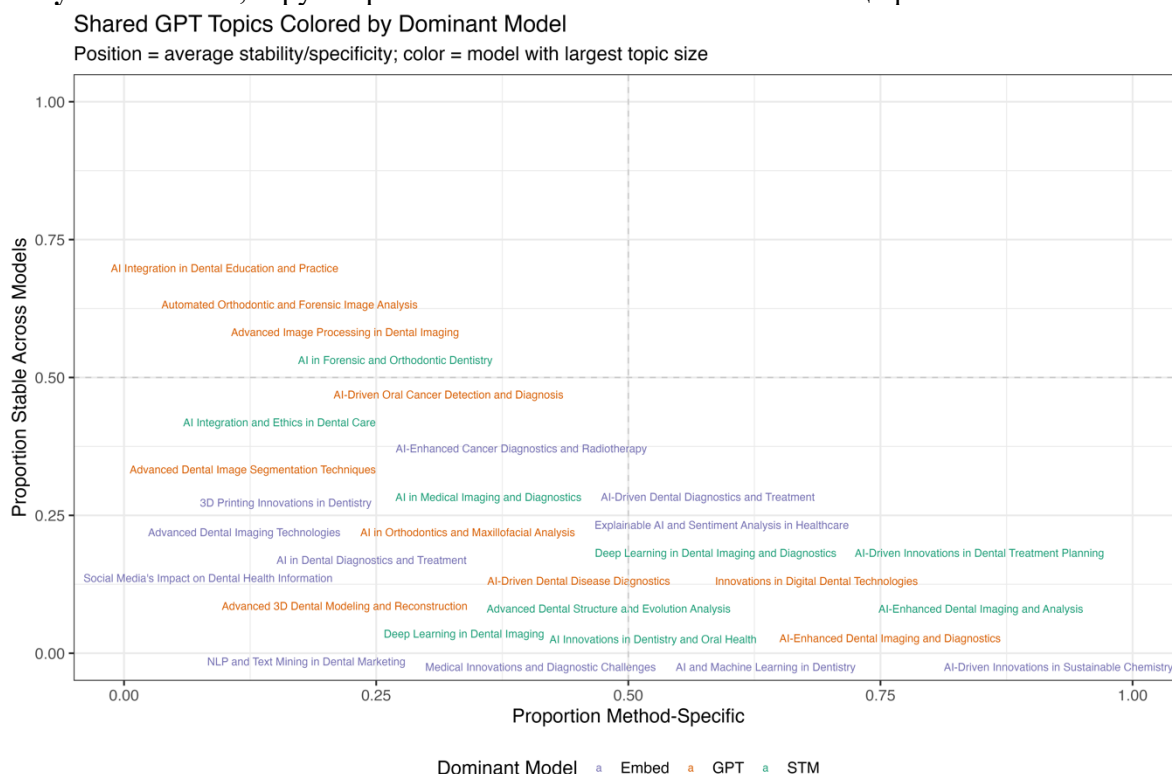
Figure 6 displays each consensus topic as a point within a two-dimensional space that illustrates both the level of agreement among the three models and which model primarily “owns” the topic. The y-axis, labeled stability, represents the proportion of documents that all three methods (STM, Embed, GPT – all mapped into GPT space, i.e., STM→GPT, EMB→GPT) consistently assign to the same

conceptual group. Higher values indicate strong agreement across models in terms of both topic identity and membership. The x-axis, labeled specificity, measures how exclusive a topic is to a single method: lower

values suggest that multiple models identify the topic, while higher values indicate it is mainly captured by one approach. The color of each point corresponds to the model that dominates the topic, determined by the largest share of mapped documents. Only topics with at least 20 documents are shown, and labels are positioned to avoid overlap.

Figure 6. Landscape of shared GPT-derived topic labels positioned by model agreement and specificity

Рисунок 6. Темы, сгруппированные по согласованности и специфичности



The upper-left quadrant (high stability, low specificity) includes the most robust, cross-method themes, i.e., topics that all three approaches identify using similar document sets. These topics are strong candidates for anchoring temporal analyses and domain summaries, as conclusions drawn from them are unlikely to be influenced by the choice of modeling method. The most stable topics in our analysis are *Metal Artefact Reduction* (STM)/ *CBCT Artifact Reduction* (Embed)/ *Advanced Image Processing in Dental Imaging* (GPT); *Chatbot Response Evaluation* (STM)/ *Dental AI Chatbot*

Evaluation(Embed)/ *AI Integration in Dental Education and Practice* (GPT); *AI Integration Challenges* (STM)/ *AI Integration and Ethics in Dental Care* (GPT); *Forensic Dental Age Estimation* (STM)/ *AI in Forensic and Orthodontic Dentistry* (GPT); *Automated Cephalometric Landmark Detection* (STM)/ *Cephalometric Landmark Detection* (Embed)/ *Automated Orthodontic and Forensic Image Analysis* (GPT); *Oral Cancer Detection* (STM)/ *Oral Cancer Detection and Diagnosis* (GPT); *AI Diagnostic Tools* (STM)/ *Dental*

AI Knowledge Extraction (Embed)/Explainable AI and Sentiment Analysis in Healthcare (GPT).

The upper-right quadrant (high stability, high specificity) includes topics that are widely agreed upon and most distinctly represented by a single model—for example, a method that defines a slightly sharper boundary around a shared theme. These topics are absent from our analysis. In contrast, the lower-left quadrant (low stability, low specificity) contains diffuse or boundary topics: multiple models identify related content, but there is weak agreement across documents. These topics often lie at conceptual intersections, such as between methods and applications. Examples include *Deep Learning Segmentation (STM)/ AI in Orthodontics and Maxillofacial Analysis (GPT)*; *Dental LLM Evaluation (Embed)/ NLP and Text Mining in Dental Marketing (GPT)*; *Tooth Segmentation Techniques (STM)/ Advanced Dental Image Segmentation Techniques (GPT)*; *Dental AI Applications (STM)/ AI in Medical Imaging and Diagnostics (GPT)*.

Finally, the lower-right quadrant (low stability, high specificity) highlights model-specific themes, where one approach clusters documents that others assign differently. These cases often result from differences in granularity such as broader GPT labels versus narrower STM or Embed clusters and modality-specific phrasing. Examples include: *Dental Image Classification, Tooth Segmentation, Deep Learning Model Performance, Dental Image Classification, Deep Learning Diagnostics, Dental Image Detection, Pathological Speech Analysis* by STM; *Dental Caries Detection, Caries Detection AI, Dental Lesion Deep Learning, Dental Image Segmentation, Dental Object Detection, Plaque Detection Automation, Panoramic Tooth Segmentation, Dental AI Knowledge Extraction, Osteoporosis Detection AI* by Embeddings, *AI and Machine Learning in Dentistry, AI-Driven Dental Diagnostics and Treatment* (GPT).

It should be noted that STM is most often associated with stable topics, while the output of embedding models is linked to low stability and high specificity, and GPT is characterized by low specificity and low stability.

4.6. Temporal Trends in Consensus GPT Space

To examine how research focus has evolved over time in the field of AI in dentistry, we performed a temporal analysis of GPT-labeled topic prevalence from 2019 to 2025. Figure 7 displays LOESS-smoothed curves showing the proportion of publications each year for each topic, normalized by the total annual output. This approach enables comparison of the relative importance of topics over time, independent of changes in overall publication volume.

Curves represent proportions based on the consensus GPT label obtained through the STM-to-GPT blended mapping.

We based temporal trend estimation on the STM→GPT mapping rather than on raw GPT or Embed→GPT assignments for three key reasons. First, STM offers the most reliable document–topic structure within our corpus, demonstrated by the highest compactness and separation metrics. In contrast, GPT topics showed less structural distinctness, while embedding clusters fell in between. Anchoring on STM reduces noise in the yearly counts. Second, STM explicitly models topic prevalence as a function of publication year, which minimizes label switching and enhances comparability over time. GPT labels are then applied to these stable STM topics to improve interpretability. Third, the STM→GPT mapping is a deterministic one-to-one function at the topic level, establishing a fixed label space for aggregation and preventing volatility caused by prompt sensitivity or run-to-run variability in zero-shot large language model outputs.

Each panel in the figure represents a different GPT-labeled topic, the *n* values indicate the overall size of each topic across 2019–2025, while the proportion curves

capture its changing relative importance within each publication year. The topics are color-coded based on their overall trend from 2019 to 2025, which is determined by the slope of their LOESS fit.

Across the study period, the annual proportions of most GPT-labeled topics remained below 0.1. This reflects the highly fragmented structure of the field: the corpus is distributed across roughly thirty distinct topics, each of which represents only a few percent of the total publication volume. When broken down by year, this dispersion means that individual topics contribute only a small share of the annual output, even if their absolute counts across the whole period are substantial. The consistently low proportion values therefore reflect the broad spread of research activity across many domains.

Several important patterns emerge. Notable growth is observed in areas that reflect recent shifts in both research focus and applied practice. *AI Integration in Dental Education and Practice (Chatbot Response Evaluation)* shows a steady and strong increase, indicating rising interest in incorporating AI/LLM into dental training and pedagogy. *AI in Dental Diagnostics and Treatment (Radiographic Lesion Detection)* and *AI-Enhanced Dental Imaging and Analysis (Dental Image Detection)* demonstrate a clear late-stage surge, especially post-2023, likely due to improvements in applied machine learning tools and clinical validation studies.

These trends suggest a growing emphasis on real-world integration of AI technologies, not only in technical domains but also in educational and operational contexts. Topic of *AI Integration and Ethics in Dental Care (AI Integration Challenges)* is also constantly growing since 2022.

In contrast, some previously prominent topics appear to be declining: *Advanced 3D Dental Modeling and Reconstruction (Maxillofacial Reconstruction)* and

Advanced Dental Imaging Technologies (CBCT Image Enhancement) have steadily decreased in relative prevalence. This may reflect either topic saturation or consolidation into broader AI-enhanced imaging themes. Similarly, *Advanced Image Processing in Dental Imaging (Metal Artefact Reduction)* show decreasing emphasis, suggesting that these methods are increasingly treated as technical subcomponents of broader diagnostic or clinical workflows.

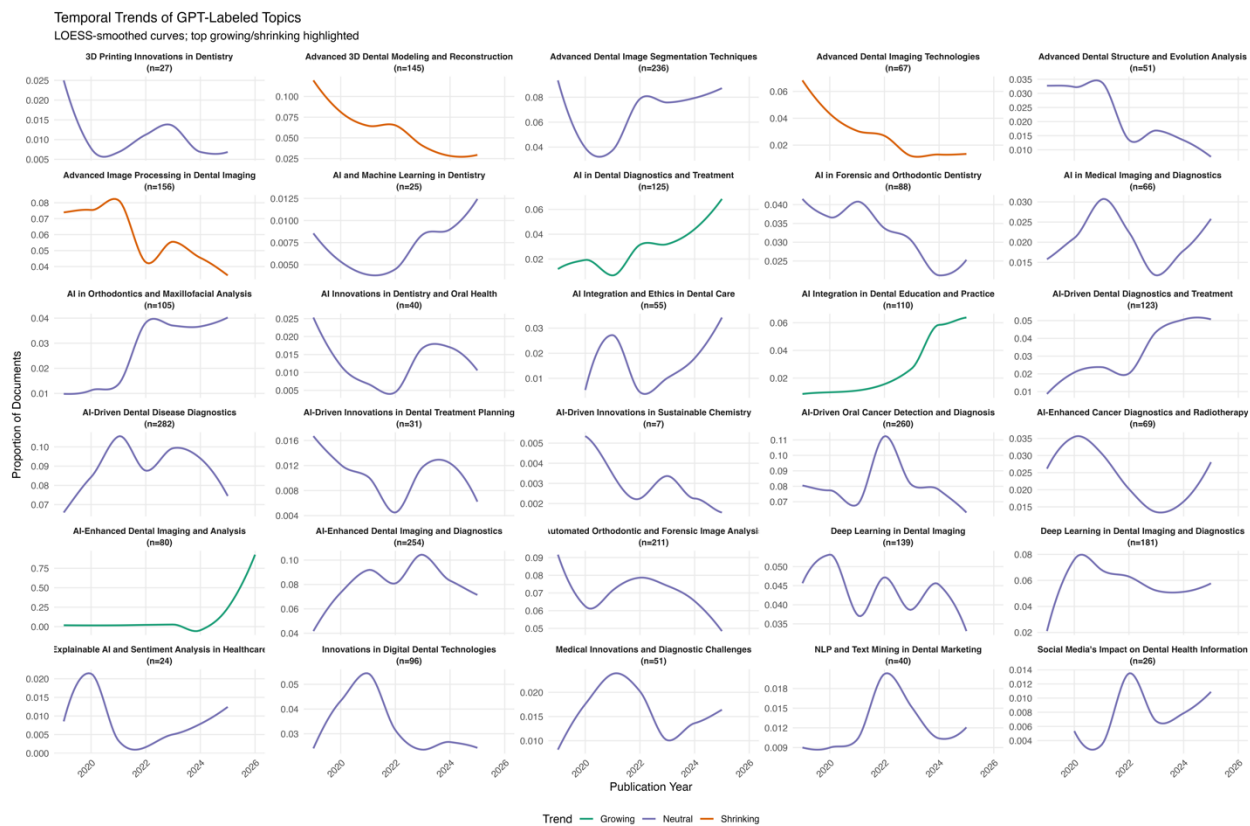
This should be noted that such topic as *AI and Machine Learning in Dentistry (Dental Image Classification)* shows stable growth after 2022.

Many topics display nonlinear patterns peaking in middle of period and then stabilizing or rebounding. *AI-Driven Oral Cancer Detection and Diagnosis (Oral Cancer Detection)* and *Deep Learning in Dental Imaging (Deep Learning Model Performance)* show early growth followed by a plateau, reflecting their maturation. *NLP and Text Mining in Dental Marketing (GPT in Medical Exams)*, while niche in size, shows cyclical attention, possibly tied to regulatory or ethical discourse spikes.

Interestingly, some smaller topics show recent resurgence, including *Explainable AI and Sentiment Analysis (AI Diagnostic Tools)*, hinting at emerging interest in communication-focused applications of AI.

This temporal mapping highlights the dynamic evolution of research priorities in AI-driven dentistry. While foundational techniques such as image processing and 3D modeling are receiving relatively less attention, application-focused topics – especially those related to clinical diagnostics, integration into practice, and patient communication – are rapidly gaining prominence. These insights provide both a retrospective understanding of the field's development and a forward-looking perspective on areas where innovation is currently accelerating.

Figure 7. Temporal Trends of STM→GPT Topics
Рисунок 7. Тренды тем с метками STM→GPT



5. DISCUSSION

This study offers a comprehensive comparative analysis of topic modeling approaches applied to the field of artificial intelligence in dentistry. We systematically examined three methodologies: Structural Topic Modeling, embedding-based clustering, and zero-shot GPT-based topic extraction to uncover latent research themes, temporal trends, and conceptual alignments within the literature corpus spanning 2019 to 2025.

The STM approach, which utilizes word co-occurrence and covariate modeling, generated interpretable and distinct topics. Its probabilistic framework enabled soft topic assignments, providing nuanced insights into thematic overlaps and ambiguities. By analyzing topic prevalence trends derived from posterior probabilities and dominant topic frequencies, emerging areas such as **AI Integration Challenges**, **Dental Image Detection**, and **Chatbot Response Evaluation**

were identified. The simultaneous increase in model-based topic probabilities and real-world document frequencies supports the conclusion that these topics are becoming key focal points in the field. Notably, STM outperformed other methods in terms of compactness and distinctiveness metrics, achieving the highest silhouette scores, tight document clustering, and the lowest redundancy in topic labeling.

The embedding-based approach, which utilized UMAP and Leiden clustering on OpenAI embeddings, provided an alternative perspective rooted in contextual semantic similarity rather than surface-level lexical patterns. The resulting clusters were generally coherent, and measuring cosine similarity to cluster centroids enabled quantification of topic-document relevance. This method highlighted emerging topics such as **Dental LLM Evaluation**, **Dental AI Chatbot Evaluation**, **CBCT Segmentation Accuracy**,

and AI-Enhanced Orthodontics, particularly after 2022. These findings align with observed technological advancements and the clinical adoption of AI-enhanced imaging tools in dental diagnostics, as well as a focus on NLP and communication within dentistry. However, the embedding model sometimes fragmented semantically related areas due to its sensitivity to local embedding structure and resolution settings.

The zero-shot GPT-based topic modeling offered a fresh perspective by generating topics and summaries without the need for training or predefined structures. While the GPT-derived topics were broader and less concise, they effectively captured thematic summaries consistent with trends observed in STM and embedding models. Temporal analysis of GPT-labeled topics highlighted the emergence of application-centered areas such as *Advanced Dental Image Segmentation Techniques*, *AI in Dental Diagnostics and Treatment*, and *Deep Learning in Dental Imaging and Diagnostics*.

A significant contribution of this study is the multi-dimensional evaluation of topic model quality. While STM produced the most distinct and compact clusters, the embedding and GPT models provided complementary insights, especially in capturing semantic nuances and summarization capabilities. Our redundancy analysis emphasized STM's advantage in generating less overlapping topic profiles, a valuable feature for downstream applications such as systematic reviews or curriculum design.

The consensus mapping across models revealed both stable and divergent conceptual structures. Topics such as *Advanced Image Processing in Dental Imaging*, *AI Integration in Dental Education and Practice*, *Automated Orthodontic and Forensic Image Analysis* were consistently represented across all models, underscoring their centrality in AI-dental research. In contrast, GPT occasionally merged or split topics in ways that deviated from STM and Embed clusters, illustrating the influence of

prompt-based generative modeling on topic formation.

Temporal analyses across all methods revealed a clear trajectory: a shift from foundational technical approaches (e.g., image processing, image segmentation) toward applied, interdisciplinary, and clinical implementations of AI. The steady increase in topics related to the use of large language models in dentistry, chatbot evaluation, and AI for education and treatment planning mirrors broader technological trends and policy changes in healthcare digitalization.

It is noteworthy that conclusions regarding the rise of NLP and hybrid models utilizing textual data, new modalities, and AI for dentistry education, treatment planning, image enhancement, and workflow optimization demonstrate considerable growth. These findings align with results obtained through manual coding of abstracts from WoS (Shirani, 2025; see also Büttner et al., 2025), thereby validating our automated approach.

Overall, our results indicate that no single modeling approach is sufficient on its own. STM excels in structural interpretability and compactness; embedding-based methods offer semantic grounding and clear cluster separation; and GPT-based models offer expressive labeling along with flexible, human-readable topic summaries. Together, these methods form a robust toolkit for analyzing and understanding domain-specific scientific literature.

This multi-method framework is adaptable across various disciplines and can support future research by facilitating the development of automated literature review tools, identifying research gaps, and monitoring the diffusion of technology in clinical fields. As AI advances, hybrid modeling approaches that integrate statistical inference, embedding spaces, and generative models will be essential for generating reliable, explainable, and comprehensive topic structures.

Like any study, ours has **limitations**. We used a single LLM with relatively simple

prompts, which may have limited model diversity and the range of topics identified. Our analyses were based on abstracts rather than full texts, potentially underrepresenting methodological details and affecting the accuracy of topic boundaries and trend estimates. Although topic labels were reviewed by the authors and showed high internal agreement, assessor bias may still be present. Additionally, the study primarily reports descriptive slopes and topic quality metrics without formal statistical testing. While these findings offer useful comparative insights, more rigorous methods such as ANOVA, bootstrapped confidence intervals, or permutation tests could provide stronger evidence of model differences and trend significance. We consider these extensions, along with broader model comparisons and external expert validation, to be important directions for future research.

6. CONCLUSIONS

This study introduces a cross-model GPT-labeled alignment framework that integrates probabilistic (STM), embedding-based (Leiden clustering), and generative (zero-shot GPT) approaches for topic modeling. To our knowledge, this is the first application of such a framework in dentistry. More broadly, it provides a replicable methodology for aligning and comparing heterogeneous topic modeling outputs. The key methodological innovation lies in combining topic-level and document-level similarity measures to construct a shared consensus topic space, enabling robust cross-model evaluation.

For method developers in NLP and bibliometrics, our findings demonstrate that no single approach dominates: STM produces compact and temporally interpretable topics, embeddings reveal fine-grained semantic clusters, and GPT generates highly readable thematic summaries. The blended alignment framework offer a generalizable tool for cross-model comparison that can be applied to other scientific domains beyond dentistry.

For domain experts in dentistry, this framework provides an automated, scalable method to map thematic landscapes and monitor research trends. When applied to AI in dentistry, it revealed a shift from foundational image-processing techniques toward applied and integrative themes, including LLM-driven applications, patient-facing AI tools, and educational initiatives. These insights are directly relevant to systematic reviews, guideline development, and research planning.

Overall, our study highlights that hybrid strategies, which combine probabilistic models with LLM-assisted labeling, produce the most comprehensive and interpretable topic analyses. While large language models enhance interpretability, they should not be used in isolation; careful integration with traditional NLP methods is essential to ensure reproducibility and explainability.

Future work will expand this approach by incorporating external human evaluations of topic quality, experimenting with diverse LLMs and prompting strategies, utilizing full-text corpora, and developing hybrid pipelines that more deeply integrate classic topic modeling with modern generative methods.

Declarations. We employed GPT-4o (OpenAI) to assist with topic extraction and topic labeling as a part of our methodology. The roles of these tools are described here and detailed in the Methods section; all analytical decisions and conclusions remain the authors' own. We used Wordservice.ai solely for English proofreading, including spelling, grammar, and stylistic edits. This service did not generate substantive content or perform any analysis. No generative tool made interpretive or methodological decisions without human oversight, and no confidential or personally identifiable data were shared to third-party services.

References

Allani, H., Santos, A. T. and Ribeiro-Vidal, H. (2024). Multidisciplinary Applications of AI in Dentistry: Bibliometric Review, *Applied*

Sciences, 14 (17), 7624.
<https://doi.org/10.3390/app14177624> (In English)

Benz, P., Pradier, C., Kozlowski, D., Shokida, N. S. and Larivière, V. (2025). Mapping the unseen in practice: comparing latent Dirichlet allocation and BERTopic for navigating topic spaces, *Scientometrics*, 10 June 2025. <https://doi.org/10.1007/s11192-025-05339-6> (In English)

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). Latent dirichlet allocation, *J. Mach. Learn. Res.*, 3, 993–1022. (In English)

Büttner, M., Leser, U., Schneider, L. and Schwendicke, F. (2024). Natural Language Processing: Chances and Challenges in Dentistry, *Journal of Dentistry*, 141, 104796. <https://doi.org/10.1016/j.jdent.2023.104796> (In English)

Cosola, Di M., Ballini, A., Prencipe F. A., De Tullio, F., Fabrocini, A., Cupelli, P., Rignani M., Cazzolla A. P., Bizzoca M. E., Musella G. (2025). Artificial intelligence in dentistry: a narrative review of applications, challenges, and future directions, *Minerva Dent Oral Sci*, Jun 18. <https://doi.org/10.23736/S2724-6329.25.05217-9> (In English)

de Magalhães, A. A. and Santos, A. T. (2025). Advancements in Diagnostic Methods and Imaging Technologies in Dentistry: A Literature Review of Emerging Approaches, *Journal of Clinical Medicine*, 14 (4), 1277. <https://doi.org/10.3390/jcm14041277> (In English)

Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure, *arXiv preprint arXiv:2203.05794* [Online], available at <https://arxiv.org/abs/2203.05794> (accessed 22.08.2025) (In English)

Hu, J., Miao, C., Wu, Yi and Su, J. (2025). Advances in hydrological research in China over the past two decades: Insights from advanced large language model and topic modeling, *Fundamental Research* [Online], available at <https://doi.org/10.1016/j.fmre.2025.05.002> (accessed 10.05.2025) (In English)

Islam, K. M. S. (2025). Contextual Embedding-based Clustering to Identify Topics for Healthcare Service Improvement, *arXiv:2504.14068*, <https://doi.org/10.48550/arXiv.2504.14068> (In English)

Jung, H. S., Lee, H., Woo, Y. S., Baek, S. Y. and Kim, J. H. (2024). Expansive data, extensive

model: Investigating discussion topics around LLM through unsupervised machine learning in academic papers and news, *Plos One*, 19 (5), e0304680.

<https://doi.org/10.1371/journal.pone.0304680> (In English)

Kozlowski, D. (2024). Generative AI for automatic topic labelling, *arXiv:2408.07003*. <https://doi.org/10.48550/arxiv.2408.07003> (In English)

Lee, V. V., van der Lubbe, S. C. C., Goh, L. H. and Valderas, J. M. (2024). Harnessing ChatGPT for Thematic Analysis: Are We Ready?. *Journal of medical Internet research*, 26, e54974. <https://doi.org/10.2196/54974> (In English)

Lee, Y., Oh, J. H., Lee, D., Kang, M. and Lee, S. (2025). Prompt engineering in ChatGPT for literature review: practical guide exemplified with studies on white phosphors, *Sci Rep*, 15, 15310. <https://doi.org/10.1038/s41598-025-99423-9> (In English)

Mahmoud, M., Mashaly, M. and Mervat, A.-E. (2025). Predicting Software Engineering Trends from Scientific Papers with a Combined Framework of Clustering and Topic Modeling, in 2025 15th International Conference on Electrical Engineering (ICEENG), 1–6. <https://doi.org/10.1109/ICEENG64546.2025.11031347> (In English)

Mathis, W. S., Zhao, S., Pratt, N., Weleff, J. and De Paoli, S. (2024). Mathis Inductive thematic analysis of healthcare qualitative interviews using open-source large language models: How does it compare to traditional methods?, *Computer Methods and Programs in Biomedicine*, 255. <https://doi.org/10.1016/j.cmpb.2024.108356> (In English)

Meng, F., Lu, Z., Li, X., Han, W., Peng, J., Liu, X. and Niu, Z. (2024). Demand-side energy management reimaged: A comprehensive literature analysis leveraging large language models, *Energy*, 291, 130303. <https://doi.org/10.1016/j.energy.2024.130303> (In English)

Mu, Y., Bai, P., Bontcheva, K. and Song, X. (2024a). Addressing topic granularity and hallucination in large language models for topic modelling. *arXiv preprint arXiv:2405.00611v1* [Online], available at <https://arxiv.org/html/2405.00611v1> (accessed 22.08.2025) (In English)

- Mu, Y., Dong, C., Bontcheva, K. and Song, X. (2024b). Large language models offer an alternative to the traditional approach of topic modelling. In LREC-COLING 2024, ELRA Language Resource Association, 2024, 10160–10171. (In English)
- Ogunleye, B., Lanchi Barrantes, B. S. and Zakariyyah, K. I. (2025). Topic modelling through the bibliometrics lens and its technique, *Artif Intell Rev*, 58, 74. <https://doi.org/10.1007/s10462-024-11011-x> (In English)
- Pham, C. M., Hoyle, A., Sun, S., Resnik, P. and Iyyer, M. 2024. 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)*, 2956–2984, Mexico City, Mexico. Association for Computational Linguistics. (In English)
- Reuter, A., Thielmann, A., Weisser, C., Fischer, S. and Säfken, B. (2024). Gptopic: Dynamic and interactive topic representations, *arXiv preprint arXiv:2403.03628* [Online], available at <https://arxiv.org/abs/2403.03628> (accessed 22.08.2025) (In English)
- Riaz, A., Abdulkader, O., Ikram, M. J. and Sadaqat, J. (2025). Exploring topic modelling: a comparative analysis of traditional and transformer-based approaches with emphasis on coherence and diversity, *International Journal of Electrical and Computer Engineering (IJECE)*, [S.l.], 15 (2), 1933–1948. <http://doi.org/10.11591/ijece.v15i2.pp1933-1948> (In English)
- Roberts, M. E., Stewart, B. M. and Tingley, D. (2019). Stm: An R package for structural topic models, *J. Stat. Softw.* 91 (2), 1–40. <https://doi.org/10.18637/jss.v091.i02> (In English)
- Şakar, S. and Tan, S. (2025). Research Topics and Trends in Gifted Education: A Structural Topic Model, *Gifted Child Quarterly*, 69 (1), 68–84. <https://doi.org/10.1177/0016986224128504> (In English)
- Sbalchiero, S. and Eder, M. (2020). Topic modeling, long texts and the best number of topics. Some Problems and solutions, *Qual Quant*, 54, 1095–1108. <https://doi.org/10.1007/s11135-020-00976-w> (In English)
- Shapurian, G. (2024). Large Language Models and Knowledge Graphs for Astronomical Entity Disambiguation, arXiv:2406.11400 [Online], available at: <https://arxiv.org/pdf/2406.11400> (accessed 22.08.2025) (In English)
- Sharma, A., Wallace, J. R. (2025). DeTAILS: Deep Thematic Analysis with Iterative LLM Support, in *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 28, 1–7. <https://doi.org/10.1145/3719160.3735657> (In English)
- Shirani, M. (2025). Trends and Classification of Artificial Intelligence Models Utilized in Dentistry: A Bibliometric Study, *Cureus*, 17 (4): e81836. <https://doi.org/10.7759/cureus.81836> (In English)
- Silveira, L. (2024). Cone Beam Computed Tomography and Artificial Intelligence. ¿Where We Are? *Rev Cient Odontol*, 12 (4), e214. <https://doi.org/10.21142/2523-2754-1204-2024-214> (In English)
- Tarek, A., Mahmoud, M., Afifi, B., Mashaly, M. and Abu-Elkheir, M. (2024). Query-Based Topic Modeling and Trend Analysis in Scientific Literature, in *2024 International Conference on Microelectronics (ICM)*, Doha, Qatar, 2024, 1–6. <https://doi.org/10.1109/ICM63406.2024.10815706> (In English)
- Torres, J., Mulligan, C., Jorge, J. and Moreira, C. (2025). PROMPTHEUS: A Human-Centered Pipeline to Streamline Systematic Literature Reviews with Large Language Models, *Information*, 16 (5), 420. <https://doi.org/10.3390/info16050420> (In English)
- Wu, X., Nguyen, T. and Luu, A. T. (2024). A survey on neural topic models: Methods, applications, and challenges, *Artif Intell Rev* 57(18): 1–30. <https://doi.org/10.1007/s10462-023-10661-7> (In English)
- Xie, B., Xu, D., Zou, X. Q., Lu, M. J., Peng, X. L. and Wen, X. J. (2024). Artificial intelligence in dentistry: a bibliometric analysis from 2000 to 2023, *J Dent Sci.*, 19, 1722–33. <https://doi.org/10.1016/j.jds.2023.10.025> (In English)
- Zatt, F. P., Rocha, A. O., Anjos, L. M., Caldas, R. A., Cardoso, M. and Rabelo, G. D. (2024). Artificial intelligence applications in dentistry: a bibliometric review with an emphasis

on computational research trends within the field, *J Am Dent Assoc.*, 155 (9), 755–764. <https://doi.org/10.1016/j.adaj.2024.05.013> (In English)

Список литературы

Allani H. Multidisciplinary Applications of AI in Dentistry: Bibliometric Review / H. Allani, A. T. Santos, H. Ribeiro-Vidal // *Applied Sciences*. 2024. Т. 14, № 17. Ст. 7624. <https://doi.org/10.3390/app14177624>

Benz P. Mapping the unseen in practice: comparing latent Dirichlet allocation and BERTopic for navigating topic spaces / P. Benz, C. Pradier, D. Kozłowski [и др.] // *Scientometrics*. 2025. Ранний онлайн-доступ (10.06.2025). <https://doi.org/10.1007/s1192-025-05339-6>

Blei D. M. Latent Dirichlet Allocation / D. M. Blei, A. Y. Ng, M. I. Jordan // *Journal of Machine Learning Research*. 2003. Т. 3. С. 993–1022.

Büttner M. Natural Language Processing: Chances and Challenges in Dentistry / M. Büttner, U. Leser, L. Schneider, F. Schwendicke // *Journal of Dentistry*. 2024. Т. 141. Ст. 104796. <https://doi.org/10.1016/j.jdent.2023.104796>

Cosola D. M. Artificial intelligence in dentistry: a narrative review of applications, challenges, and future directions / D. M. Cosola, A. Ballini, F. A. Prencipe [и др.] // *Minerva Dent Oral Sci*. 2025. Ранний онлайн-доступ (18.06.2025). <https://doi.org/10.23736/S2724-6329.25.05217-9> (In English)

de Magalhães A. A. Advancements in Diagnostic Methods and Imaging Technologies in Dentistry: A Literature Review of Emerging Approaches / A. A. de Magalhães, A. T. Santos // *Journal of Clinical Medicine*. 2025. Т. 14, № 4. Ст. 1277. <https://doi.org/10.3390/jcm14041277>

Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure / M. Grootendorst // *arXiv preprint arXiv:2203.05794*. URL: <https://arxiv.org/abs/2203.05794> (дата обращения: 22.08.2025).

Hu J. Advances in hydrological research in China over the past two decades: Insights from advanced large language model and topic modeling / J. Hu, C. Miao, Y. Wu, J. Su // *Fundamental Research*. 2025. Ранний онлайн-доступ (10.05.2025). <https://doi.org/10.1016/j.fmre.2025.05.002>

Islam K. M. S. Contextual Embedding-based Clustering to Identify Topics for Healthcare Service Improvement / K. M. S. Islam // *arXiv:2504.14068*. URL:

<https://arxiv.org/abs/2504.14068> (дата обращения: 22.08.2025). <https://doi.org/10.48550/arXiv.2504.14068>

Jung H. S. Expansive data, extensive model: Investigating discussion topics around LLM through unsupervised machine learning in academic papers and news / H. S. Jung, H. Lee, Y. S. Woo [и др.] // *PLOS ONE*. 2024. Т. 19, № 5. Ст. e0304680. <https://doi.org/10.1371/journal.pone.0304680>

Kozłowski D. Generative AI for automatic topic labelling / D. Kozłowski // *arXiv:2408.07003*. URL:

<https://arxiv.org/abs/2408.07003> (дата обращения: 22.08.2025). <https://doi.org/10.48550/arxiv.2408.07003>

Lee V. V. Harnessing ChatGPT for Thematic Analysis: Are We Ready? / V. V. Lee, S. C. C. van der Lubbe, L. H. Goh, J. M. Valderas // *Journal of Medical Internet Research*. 2024. Т. 26. Ст. e54974. <https://doi.org/10.2196/54974>

Lee Y. Prompt engineering in ChatGPT for literature review: practical guide exemplified with studies on white phosphors / Y. Lee, J. H. Oh, D. Lee [и др.] // *Sci Rep*. 2025. Т. 15. Ст. 15310. <https://doi.org/10.1038/s41598-025-99423-9>

Mahmoud M. Predicting Software Engineering Trends from Scientific Papers with a Combined Framework of Clustering and Topic Modeling / M. Mahmoud, M. Mashaly, A.-E. Mervat // *Proc. 2025 15th International Conference on Electrical Engineering (ICEENG)*. 2025. С. 1–6. <https://doi.org/10.1109/ICEENG64546.2025.11031347>

Mathis W. S. Inductive thematic analysis of healthcare qualitative interviews using open-source large language models: How does it compare to traditional methods? / W. S. Mathis, S. Zhao, N. Pratt [и др.] // *Computer Methods and Programs in Biomedicine*. 2024. Т. 255. Ст. 108356. <https://doi.org/10.1016/j.cmpb.2024.108356>

Meng F. Demand-side energy management reimaged: A comprehensive literature analysis leveraging large language models / F. Meng, Z. Lu, X. Li [и др.] // *Energy*. 2024. Т. 291. Ст. 130303. <https://doi.org/10.1016/j.energy.2024.130303>

Mu Y. Addressing topic granularity and hallucination in large language models for topic modelling / Y. Mu, P. Bai, K. Bontcheva, X. Song // arXiv:2405.00611v1. URL: <https://arxiv.org/html/2405.00611v1> (дата обращения: 22.08.2025).

Mu Y. Large language models offer an alternative to the traditional approach of topic modelling / Y. Mu, C. Dong, K. Bontcheva, X. Song // Proc. LREC-COLING 2024. 2024. C. 10160–10171.

Ogunleye B. Topic modelling through the bibliometrics lens and its technique / B. Ogunleye, B. S. Lancho Barrantes, K. I. Zakariyyah [и др.] // Artificial Intelligence Review. 2025. Т. 58. Ст. 74. <https://doi.org/10.1007/s10462-024-11011-x>

Pham C. M. TopicGPT: A Prompt-based Topic Modeling Framework / C. M. Pham, A. Hoyle, S. Sun // Proc. of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2024). 2024. C. 2956–2984. Mexico City. Association for Computational Linguistics.

Reuter A. Gptopic: Dynamic and interactive topic representations / A. Reuter, A. Thielmann, C. Weisser [и др.] // arXiv:2403.03628. URL: <https://arxiv.org/abs/2403.03628> (дата обращения: 22.08.2025).

Riaz A. Exploring topic modelling: a comparative analysis of traditional and transformer-based approaches with emphasis on coherence and diversity / A. Riaz, O. Abdulkader, M. J. Ikram, J. Sadaqat // International Journal of Electrical and Computer Engineering (IJECE). 2025. Т. 15, № 2. C. 1933–1948. <http://doi.org/10.11591/ijece.v15i2.pp1933-1948>

Roberts M. E. STM: An R package for structural topic models / M. E. Roberts, B. M. Stewart, D. Tingley // Journal of Statistical Software. 2019. Т. 91, № 2. C. 1–40. <https://doi.org/10.18637/jss.v091.i02>

Şakar S. Research Topics and Trends in Gifted Education: A Structural Topic Model / S. Şakar, S. Tan // Gifted Child Quarterly. 2025. Т. 69, № 1. C. 68–84. <https://doi.org/10.1177/0016986224128504>

Sbalchiero S. Topic modeling, long texts and the best number of topics. Some problems and solutions / S. Sbalchiero, M. Eder // Quality & Quantity. 2020. Т. 54. C. 1095–1108. <https://doi.org/10.1007/s11135-020-00976-w>

Shapurian G. Large Language Models and Knowledge Graphs for Astronomical Entity Disambiguation / G. Shapurian // arXiv:2406.11400. URL: <https://arxiv.org/pdf/2406.11400> (дата обращения: 22.08.2025).

Sharma A. DeTAILS: Deep Thematic Analysis with Iterative LLM Support / A. Sharma, J. R. Wallace // Proc. of the 7th ACM Conference on Conversational User Interfaces (CUI '25). 2025. Article 28. C. 1–7. <https://doi.org/10.1145/3719160.3735657>

Shirani M. Trends and Classification of Artificial Intelligence Models Utilized in Dentistry: A Bibliometric Study / M. Shirani // Cureus. 2025. Т. 17, № 4. Ст. e81836. <https://doi.org/10.7759/cureus.81836>

Silveira L. Cone Beam Computed Tomography and Artificial Intelligence. Where We Are? / L. Silveira // Rev Cient Odontol. 2024. Т. 12, № 4. Ст. e214. <https://doi.org/10.21142/2523-2754-1204-2024-214>

Tarek A. Query-Based Topic Modeling and Trend Analysis in Scientific Literature / A. Tarek, M. Mahmoud, B. Afifi [и др.] // 2024 International Conference on Microelectronics (ICM). Doha, Qatar. 2024. C. 1–6. <https://doi.org/10.1109/ICM63406.2024.10815706>

Torres J. PROMPTHEUS: A Human-Centered Pipeline to Streamline Systematic Literature Reviews with Large Language Models / J. Torres, C. Mulligan, J. Jorge, C. Moreira // Information. 2025. Т. 16, № 5. Ст. 420. <https://doi.org/10.3390/info16050420>

Wu X. A survey on neural topic models: Methods, applications, and challenges / X. Wu, T. Nguyen, A. Luu // Artificial Intelligence Review. 2024. Т. 57, № 18. C. 1–30. <https://doi.org/10.1007/s10462-023-10661-7>

Xie B. Artificial intelligence in dentistry: a bibliometric analysis from 2000 to 2023 / B. Xie, D. Xu, X. Q. Zou [и др.] // Journal of Dental Sciences. 2024. Т. 19. C. 1722–1733. <https://doi.org/10.1016/j.jds.2023.10.025>

Zatt F. P. Artificial intelligence applications in dentistry: a bibliometric review with an emphasis on computational research trends within the field / F. P. Zatt, A. O. Rocha, L. M. Anjos [и др.] // Journal of the American Dental Association. 2024. Т. 155, № 9. C. 755–764. <https://doi.org/10.1016/j.adaj.2024.05.013>

The authors have read and approved the final manuscript.

Авторы прочитали и одобрили окончательный вариант рукописи.

Conflicts of interests: the authors have no conflicts of interest to declare.

Конфликты интересов: у авторов нет конфликтов интересов для декларации.

Литвинова Татьяна Александровна, доктор филологических наук, профессор кафедры русского языка, современной русской и зарубежной литературы Воронежского государственного педагогического университета, Воронеж, Россия; ведущий научный сотрудник в области компьютерных наук, Воронежский государственный университет, Воронеж, Россия

Tatiana A. Litvinova, Doctor of Philological Sciences, Professor of the Department of Russian Language, Modern

Russian and Foreign Literature, Voronezh State Pedagogical University, Voronezh, Russia; leading researcher in computer science, Voronezh State University, Voronezh, Russia

Ипполитов Юрий Алексеевич, доктор медицинских наук, заведующий кафедрой детской стоматологии с ортодонтией, Воронеж, Россия

Yury A. Ippolitov, Doctor of Medical Sciences, Head of the Department of Pediatric Dentistry with Orthodontics, Voronezh State Medical University, Voronezh, Russia

Середин Павел Владимирович, доктор физико-математических наук, заведующий кафедрой физики твердого тела и наноструктур, Воронежский государственный университет, Воронеж, Россия

Pavel V. Seredin, Doctor of Physical and Mathematical Sciences, Head of the Department of Solid State Physics and Nanostructures, Voronezh State University, Voronezh, Russia