EPiC Series in Technology
Volume 3, 2026, Pages 208–217
Proceedings of AI for Accelerated Research Symposium

Unveiling Fibromyalgia Research Frontiers: Transformer-Based Topic and Sentiment Modeling for Biomedical Meta-Analysis

Yetunde O. Longe-Folajimi[1]*, Leonidas Deligiannidis[1]†, Salem Othman[1]‡, and Ibukun Folajimi[1]§

Wentworth Institute of Technology, Boston, Massachusetts, USA
folajimiy@wit.edu
deligiannidisl@wit.edu
othmans1@wit.edu
folajimii@wit.edu

Abstract

The exponential growth of biomedical literature poses challenges for synthesizing thematic and emotional insights, particularly in underexplored conditions like fibromyalgia. We present a reproducible and modular pipeline that integrates BERTopic—an explainable topic modeling framework—with sentiment analysis to map 5,861 PubMed abstracts on fibromyalgia. It primarily spans publications from 1990 to 2020, with a small number of records predating 1990; this coverage enables longitudinal analysis of research themes and sentiment. Our approach combines Sentence-BERT embeddings, density-based clustering, and TF-IDF topic representation to extract 111 interpretable topics and one noise cluster. Key themes include sleep dysfunction, multimodal treatment, genetic biomarkers, and patient experience—an emergent area increasingly emphasized in chronic illness research. We benchmark BERTopic against Latent Dirichlet Allocation (LDA) and Contextual Topic Modeling (CTM) using four coherence metrics (C_V, UMass, NPMI, and C_UCI). While CTM achieved the highest coherence score (C_V = 0.6748), BERTopic (C_V = 0.6331) offered superior visualization, adaptability, and usability. Sentiment analysis, conducted using a DistilBERT classifier trained on the SST-2 dataset, revealed domain-specific polarity patterns—e.g., overwhelmingly negative tone in sleep-related studies and balanced sentiment in patient-centered topics. Although the sentiment model was not fine-tuned on biomedical text, it provided meaningful first-order approximations. This work contributes a scalable framework for scientific landscape mapping in low-data medical domains.

1 Introduction

Fibromyalgia remains difficult to diagnose, treat, and study despite a large and growing research literature. More than 5,000 publications now address the condition, yet the field still lacks

*Design, implementation, and presentation
†Testing, proofreading, and suggestions
‡Testing, proofreading, and suggestions
§Testing and proofreading

coherent, evidence-based pathways for synthesizing its clinical, biological, and psychosocial dimensions. This challenge reflects a broader problem in biomedical research: the volume of literature has outpaced our ability to interpret and integrate it.

This paper addresses that challenge by developing and evaluating an interpretable, automated pipeline for literature synthesis in a complex, clinically heterogeneous domain. We combine BERTopic with transformer-based sentiment analysis to identify major themes and emotional patterns in fibromyalgia abstracts, while benchmarking BERTopic against LDA and CTM. Our goal is to support scalable, explainable synthesis in low-data medical settings where semantic nuance and clinical interpretability matter.

2 Related Work

Our study lies at the intersection of biomedical NLP, topic modeling, and AI-assisted research synthesis.

2.1 Biomedical NLP and Scientific Language Models

Domain-adapted language models such as BioBERT [13] and PubMedBERT [10] demonstrated the value of transfer learning for biomedical text, while larger models such as GatorTron [21] and BioMedLM [11] extended these gains to broader clinical and scientific reasoning tasks. These developments have helped establish transformer-based methods as core tools for biomedical text analysis.

2.2 Topic Modeling in Healthcare Literature

Topic modeling has long been used to uncover latent structure in healthcare corpora. LDA [4] remains widely used, but its bag-of-words assumption ignores semantic context and requires manual topic selection, which can reduce coherence in sparse or heterogeneous datasets. Neural and contextual topic models, including CTM [2], address some of these weaknesses through embeddings and variational inference. BERTopic [9] extends this trend by combining Sentence-BERT embeddings, HDBSCAN clustering, and class-based TF-IDF to produce interpretable, context-aware topics.

2.3 Automated Research Synthesis

AI-assisted review tools such as TLDR [5], ASReview, and Elicit [15] automate parts of scientific synthesis, including summarization and relevance ranking. However, few studies integrate transformer-based topic modeling and sentiment analysis into a reproducible workflow for chronic illness literature. In prior work [7], we combined LDA with lexicon-based sentiment analysis to map fibromyalgia research. The present study advances that framework through BERTopic, transformer-based sentiment classification, and comparative benchmarking against LDA and CTM.

3 Methodology

We utilized the publicly available dataset curated by Phaterpekar [16]. The corpus contains 5,861 PubMed abstracts with publication years spanning 1982–2020 and no missing year values. Although the analysis primarily emphasizes 1990–2020, 121 pre-1990 abstracts were retained to preserve full coverage of the source dataset. Abstracts had a mean length of 211.2 words (median 217; IQR 160–251; range 6–949) and were drawn from approximately 1,117 journals, with the highest concentrations in rheumatology and pain venues such as *J Rheumatol*, *Clin Exp Rheumatol*, and *Pain*. Table 1 summarizes the corpus characteristics, and Table 2 reports publication counts by time bin.

Table 1: Dataset summary statistics.

Characteristic	Value
Number of abstracts	5,861
Year range (from pub_date)	1982–2020
Missing year values	0%
Pre-1990 abstracts	121
Mean abstract length (words)	211.2
Median abstract length (words)	217
IQR abstract length (words)	160–251
Min/Max abstract length (words)	6 / 949
Approx. unique journals	1,117

Table 2: Abstract counts by publication bins (5-year bins up to 2014; final bin grouped as 2015–2020).

Years	Count
1980–1984	5
1985–1989	116
1990–1994	260
1995–1999	435
2000–2004	658
2005–2009	1,015
2010–2014	1,528
2015–2020	1,844

3.1 Text Preprocessing

All abstracts were preprocessed using `spaCy` [1] and `NLTK` [3]. The workflow included lowercasing, tokenization, lemmatization, and removal of stopwords and punctuation. The cleaned text was used as input for both topic modeling and sentiment analysis.

3.2 Topic Modeling and Benchmarking

To capture semantic similarity among abstracts, we generated dense document embeddings using Sentence-BERT (SBERT) [17] through the `sentence-transformers` library. These embeddings were then used within BERTopic [9], which combines transformer-based representations with clustering and interpretable topic representations.

Specifically, BERTopic applied Uniform Manifold Approximation and Projection (UMAP) [14] for dimensionality reduction, Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [6] for clustering, and class-based TF-IDF (c-TF-IDF) to derive topic descriptors. The number of topics was inferred automatically from the data rather than specified in advance.

To assess the quality of BERTopic, we benchmarked it against two established baselines:

1. **Latent Dirichlet Allocation (LDA)** [4], implemented with `gensim`, with the number of topics selected through coherence-based grid search.

2. **Contextual Topic Model (CTM)** [2], implemented using the `contextualized-topic-models` library.

All three models were evaluated using four topic coherence metrics:

- **C_V**, which combines normalized pointwise mutual information and cosine similarity;
- **C_UCI**, based on word co-occurrence counts and pointwise mutual information;
- **C_NPMI**, which measures normalized pointwise mutual information; and
- **U_Mass**, which uses document co-occurrence counts from the reference corpus.

These metrics were computed using the `gensim` and `octis` packages. To improve robustness, each model was trained and evaluated across five random seeds.

3.3 Sentiment Analysis and Visualization

To estimate affective tone across the literature, we applied a DistilBERT-based sentiment classifier [18] using the HuggingFace `transformers` library [20]. Each abstract was assigned a binary sentiment label (`POSITIVE` or `NEGATIVE`), and predictions were aggregated within topic clusters to examine topic-level polarity patterns.

For temporal analysis, publication years were extracted from the `pub_date` field and grouped into time bins. Topic prevalence trends were then visualized across these periods. Figures were generated using `matplotlib` [12], `pgfplots`, and BERTopic's built-in plotting functions, including topic timelines and inter-topic visualizations.

All experiments were conducted in Google Colab using a T4 GPU and 13GB RAM, supporting a reproducible and accessible workflow.

4 Results

The BERTopic model identified **112 distinct topics** across the 5,861 fibromyalgia-related abstracts. Of these, **Topic -1** accounted for 1,942 abstracts (33.1%). In BERTopic, Topic -1 represents documents not confidently assigned to a coherent cluster by HDBSCAN, a common outcome in heterogeneous biomedical corpora. The remaining 111 topics were interpretable and aligned with major clinical and biomedical themes in fibromyalgia research.

4.1 Top Interpretable Topics

Table 3 presents the five largest interpretable topics (excluding Topic -1). Labels were assigned through review of high-frequency terms and representative abstracts.

Table 3: Top 5 Interpretable Topics Identified via BERTopic

Topic ID	Label and Description
0	**Sleep Dysfunction and EEG Analysis**: Studies on alpha-delta sleep abnormalities, polysomnographic markers, and qEEG diagnostics.
1	**Multimodal Treatment and Diagnosis**: Clinical management strategies, diagnostic criteria, pharmacologic and non-pharmacologic interventions.
2	**Patient Experiences and Expectations**: Qualitative insights into patient interactions with healthcare systems, expectations, and social reception.
3	**Genetic and Biomarker Studies**: Associations between genetic polymorphisms (e.g., *COMT*, β_2-AR) and symptom severity or neuroimmune responses.
4	**Physical Activity and Mental Health**: Effects of exercise interventions on depression, fatigue, and physical functioning in fibromyalgia patients.

These topics reflect central strands of fibromyalgia research. Topic 0 emphasized disrupted sleep architecture and alpha-wave intrusion during non-REM sleep. Topic 1 captured multimodal treatment and diagnosis, including pharmacologic, behavioral, and exercise-based interventions. Topic 2 focused on patient experiences, including stigma, diagnostic delay, and healthcare access. Topic 3 centered on genetic and biomarker studies, especially adrenergic and pain-related polymorphisms. Topic 4 linked physical activity to mood, fatigue, and functional outcomes while also reflecting patient reluctance driven by fear of pain exacerbation.

4.2 Topic Volume Distribution

Figure 1 shows the relative frequency of abstracts across the five largest interpretable topics.

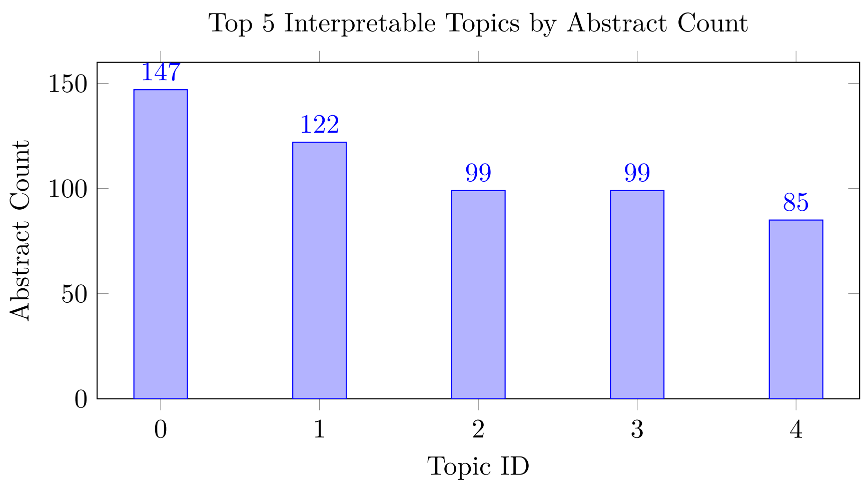

Figure 1: Distribution of abstracts across top 5 interpretable BERTopic topics.

4.3 Temporal Trends in Topic Prevalence

Before examining topic-level dynamics, we summarize overall publication growth. Figure 2 shows a steady increase in fibromyalgia-related abstracts over time, with the highest concentration in 2015–2020. The 5-year publication bins reported earlier in Table 2 provide the context for interpreting topic prevalence trends.

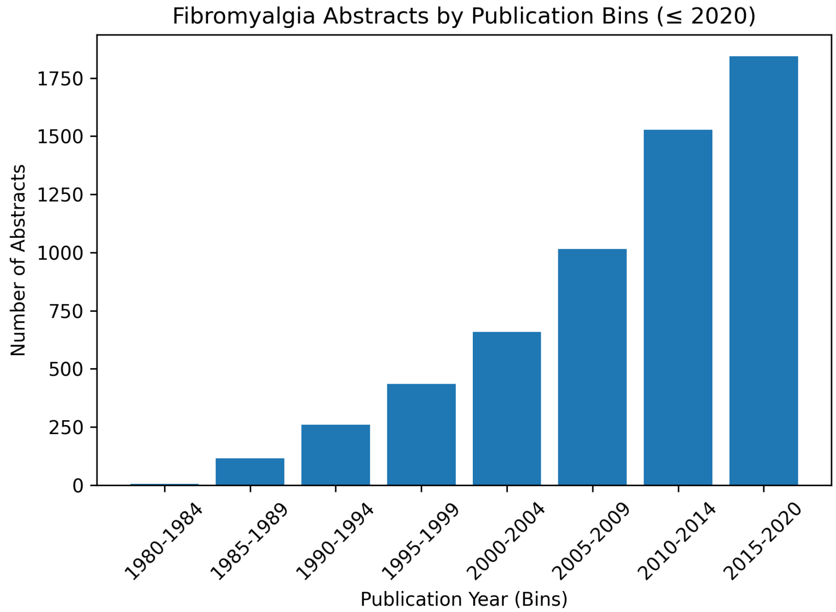

Figure 2: Distribution of fibromyalgia-related PubMed abstracts in the dataset by publication bins (5-year bins through 2014; 2015–2020 grouped).

Figure 3 shows how the top interpretable topics evolved over time. Topic 1 (Multimodal Treatment and Diagnosis) increased steadily across the corpus, suggesting sustained interest in integrative care strategies. Topic 3 (Genetic and Biomarker Studies) became more prominent in the 2010s, reflecting growing interest in mechanistic and personalized approaches. Topic 2 (Patient Experiences and Expectations) gained visibility in later years, consistent with broader shifts toward patient-centered care and qualitative inquiry in chronic illness research.

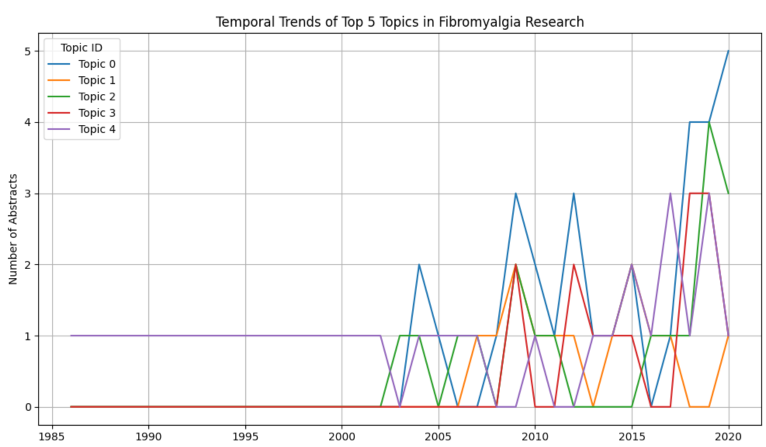

Figure 3: Temporal distribution of abstracts for the top 5 interpretable topics (excluding noise Topic -1).

4.4 Sentiment Analysis of Abstracts

We performed binary sentiment analysis using a DistilBERT classifier trained on SST-2 [19]. Although the model was not fine-tuned on biomedical text, it provided a useful first-order approximation of affective tone across topics. Figure 4 summarizes the proportion of positive and negative abstracts within the top five topics.

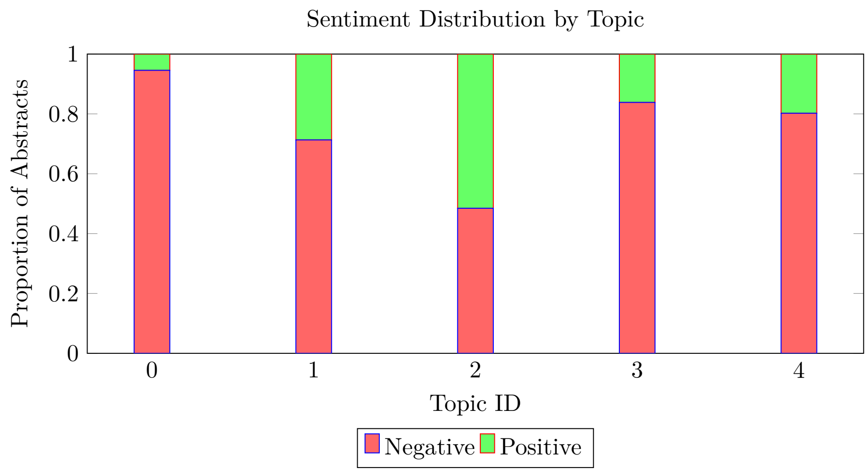

Figure 4: Proportion of negative and positive sentiment labels across top 5 BERTopic clusters.

Topic 0 (Sleep Dysfunction) showed overwhelmingly negative sentiment (94.6%), consistent with studies emphasizing persistent pain and sleep disruption. Topic 1 (Treatment and Diagnosis) was more balanced but still predominantly negative. Topic 2 (Patient Experiences) was the only cluster with slightly more positive than negative abstracts (51.5%), possibly reflecting greater attention to advocacy and patient-centered care. Topics 3 and 4 remained predominantly negative, suggesting continuing uncertainty in biomarker and exercise-related research.

4.5 Quantitative Benchmarking of Topic Models

To evaluate topic model quality, we compared BERTopic with LDA and CTM using four coherence metrics: C_V, U_Mass, C_UCI, and C_NPMI. Figure 5 summarizes the C_V scores.

CTM achieved the highest coherence (C_V = 0.6748), followed by BERTopic (0.6331), while LDA performed substantially lower (0.4504). These results indicate that contextual embedding methods produced more semantically coherent topics than the traditional bag-of-words baseline. Although CTM yielded the best coherence score, BERTopic provided a stronger balance of coherence, interpretability, visualization, and usability for exploratory analysis.

4.6 Interpretation and Implications

Several broader patterns emerge from these findings. First, BERTopic isolated patient experience as a distinct thematic area rather than subsuming it within treatment or quality-of-life categories, suggesting a growing emphasis on patient-centered research in fibromyalgia. Second, the prominence of biomarker and genetic topics reflects continued interest in individualized and mechanistic explanations of chronic pain. Third, the sentiment results suggest that emotionally negative framing remains especially common in symptom-focused and exploratory biomedical studies, while patient-centered work shows a more balanced tone.

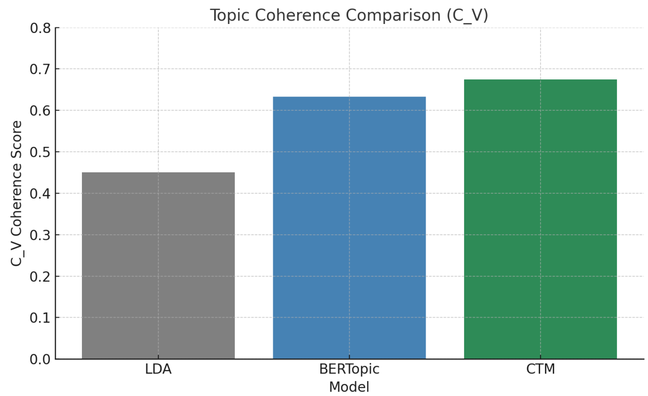

Figure 5: Comparison of topic coherence scores (C_V) across LDA, BERTopic, and CTM.

Taken together, the results support the use of BERTopic as a practical and interpretable framework for literature synthesis in heterogeneous biomedical domains. Beyond fibromyalgia, the same pipeline could be adapted to other chronic and medically contested conditions, including chronic fatigue syndrome, and irritable bowel syndrome.

5 Conclusions

We presented a transformer-based, explainable pipeline for analyzing the thematic and emotional structure of fibromyalgia research. Using 5,861 PubMed abstracts, we combined BERTopic with transformer-based sentiment analysis to identify major research themes and compare topic-model quality across BERTopic, LDA, and CTM. The results showed that contextual embedding approaches produced more coherent and interpretable topics than traditional bag-of-words modeling, while BERTopic offered a strong practical balance of semantic quality, visualization, and usability.

Our findings also highlighted several clinically meaningful patterns in the literature, including the prominence of sleep dysfunction, multimodal treatment, patient experience, biomarker discovery, and physical activity. In particular, the emergence of patient experience as a distinct topic suggests a growing shift toward patient-centered perspectives in fibromyalgia research.

This study has several limitations. First, the analysis is based on PubMed abstracts only and therefore does not capture the richer detail available in full-text articles. Second, the DistilBERT sentiment classifier was not fine-tuned on biomedical text, which may limit domain-specific affective precision. Third, although BERTopic performed well overall, topic interpretation remains partly subjective and clustering results are sensitive to parameter settings.

Despite these limitations, the use of open-access data and publicly available models supports reproducibility and reuse. Code, preprocessed data, and visualizations are available in our companion repository, https://github.com/folajimiy/fibroai [8]. Future work will extend this pipeline through biomedical sentiment fine-tuning, full-text integration, and dynamic knowledge representations for longitudinal scientific synthesis.

References

[1] Explosion AI. spacy: Industrial-strength natural language processing in python, 2023.

[2] Federico Bianchi, Silvia Terragni, and Dirk Hovy. Pre-training is a hot topic: Contextualized

document embeddings improve topic coherence. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Short Papers*. Association for Computational Linguistics, 2020.

[3] Steven Bird, Ewan Klein, and Edward Loper. Natural language toolkit, 2023.

[4] David Blei, Andrew Ng, and Michael Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[5] Ivan Cachola, Kyle Lo, Arman Cohan, and Noah A. Smith. Tldr: Extreme summarization of scientific documents. In *EMNLP*, 2020.

[6] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. Density-based clustering based on hierarchical density estimates. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 160–172, 2013.

[7] Yetunde O. Folajimi. Navigating the complexities of fibromyalgia research: an artificial intelligence-driven explorat. *Journal of Medical Artificial Intelligence*, 8(29), 2025.

[8] Yetunde O. Folajimi, Leonidas Deligiannidis, Salem Othman, and Ibukun Folajimi. Fibroai: Transformer topic and sentiment modeling for fibromyalgia literature, 2025. GitHub repository.

[9] Maarten Grootendorst. Bertopic: Neural topic modeling with class-based tf-idf. *arXiv preprint arXiv:2203.05794*, 2022.

[10] Yu Gu, Robert Tinn, Hao Cheng, and et al. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare*, 3(1):1–23, 2021.

[11] Vivek Gupta, Baolin Peng, and Emily Alsentzer. Biomedlm: A foundation model for biomedical natural language processing. https://huggingface.co/stanford-crfm/BioMedLM, 2024.

[12] John D Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[13] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, and et al. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

[14] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[15] Harsha Nori, Ayesha Malik, and et al. Large language models in scientific discovery and literature review. *Nature Reviews Methods Primers*, 2023.

[16] Tejas Phaterpekar. Pubmed fibromyalgia article abstracts data set. https://www.kaggle.com/datasets/tphaterp/pubmed-fibromyalgiaarticle-abstracts/data, 2019. Accessed August 2025.

[17] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

[18] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.

[19] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1631–1642. Association for Computational Linguistics, 2013.

[20] Thomas Wolf, Lysandre Debut, Victor Sanh, et al. Transformers: State-of-the-art natural language processing. *EMNLP*, 2020.

[21] Xi Yang, Selcuk Bozkurt, and et al. Gatortron: A large language model for clinical natural language processing. *NPJ Digital Medicine*, 5(1), 2022.