2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-8622-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/BIBM62325.2024.10822835

On the Role of LLM to Forecast the Next Pandemic

Tommaso Ruga
DIMES, University of Calabria
CNR-NANOTEC
Rende (CS), Italy
tommaso.ruga@dimes.unical.it

Eugenio Vocaturo
CNR-NANOTEC
DIMES, University of Calabria
Rende (CS), Italy
eugenio.vocaturo@cnr.it

Ester Zumpano
DIMES, University of Calabria
CNR-NANOTEC
Rende (CS), Italy
e.zumpano@dimes.unical.it

***Abstract*—Large Language Models (LLMs) have emerged as powerful tools in medical applications, demonstrating significant potential for analyzing vast clinical datasets, identifying patterns, and facilitating early detection of health anomalies. These capabilities are particularly relevant for detecting both individual disease onset and potential pandemic scenarios at a population level. The integration of LLMs into medical workflows could revolutionize how data-driven insights are harnessed for pandemic prediction and management, although implementing these technologies requires careful consideration of critical issues, particularly data privacy and security. This paper presents a comprehensive examination of current research at the intersection of LLMs and pandemic response, analyzing the literature from both bibliometric and medical perspectives. Through the selection of 849 publications across major databases, we provide an overview of the current state of research in this domain, identify emerging patterns from a clinical standpoint, and evaluate potential implications for future pandemic prediction. Our findings reveal significant trends in the application of LLMs to pandemic-related challenges, highlighting both opportunities and critical areas requiring further investigation, particularly in mental health impacts and early warning systems. The analysis reveals that while LLMs show promise in early detection and pattern recognition, challenges remain in data privacy, model interpretability, and the integration of diverse data sources. This research serves as a foundation for understanding how LLMs can be effectively deployed in pandemic prediction and management, while acknowledging the complexities and ethical considerations inherent in such applications.**

***Index Terms*—Large Language Models, LLMs, Healthcare, Pandemic, Literature Review**

I. Introduction

Pandemics, particularly the COVID-19 pandemic, have had profound impacts globally. According to the World Health Organization (WHO), as of mid-2024, there have been over 770 million confirmed cases of COVID-19, see Fig. 1, and nearly 7 million deaths globally [1]. Global vaccination efforts have been a central response. By 2024, billions of vaccine doses were administered worldwide, with initiatives like COVAX aiming to ensure equitable access for lower-income countries [2]. Artificial Intelligence (AI) has transformed the medical field in recent years, introducing innovative solutions designed to optimize and streamline processes integral to healthcare. Rather than replacing medical professionals, AI tools are used to provide solutions that assist clinicians in making more accurate and timely decisions. These tools provide support during critical decision-making phases, offering insights derived from extensive datasets and advanced analytical capabilities. In fact,

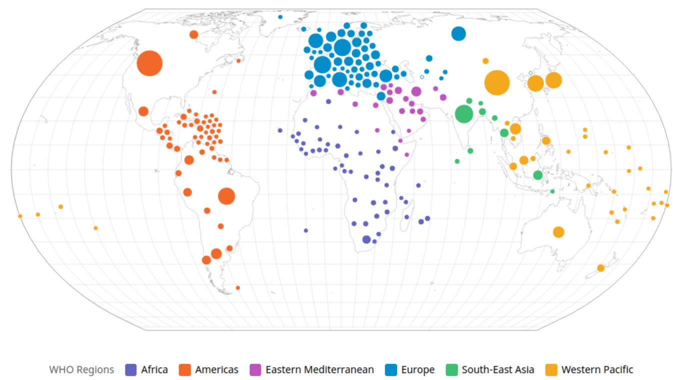

Fig. 1. Map of COVID-19 cases reported to WHO (*Source: World Health Organization*)

the strength of AI lies in its ability to analyze enormous datasets rapidly with the aim of identifying patterns and correlations that are challenging to detect manually. In this context, Large Language Models (LLMs) can play a significant role in the field of medicine and more specifically can make a difference in the analysis of large masses of clinical data especially in a pandemic context, identifying patterns that may suggest medical problems in a large population sample. Early detection of anomalies in patients in a specific world region could suggest the presence of an evolving and expanding virus, just as it could suggest particular health problems due to lowered immune defenses. In such a context, it would be necessary to ensure the right level of user data protection and consequently ensure a high level of privacy in accordance with the General Data Protection Regulation (GDPR). Again, AI could provide federated solutions capable of analyzing data only within the medical institutions involved, without harming the security of citizens [3]. It is in this context that the proposed work is set. The paper aims to review the literature and the solutions currently proposed to understand, at a high level of granularity, what has been done in the context of LLMs to serve the prediction of a possible future pandemic.

The aim of the work is not to propose a comprehensive systematic literature review but rather proposing a preliminary report on the domain, offering a quantitative statistical overview of the existing literature. This study should therefore be understood as a starting point for a more low-

979-8-3503-8622-6/24/$31.00 ©2024 IEEE

level future analysis, potentially leading to alternative more in depth approaches, although it follows a rigorous and objective methodology of analysis.

The work is organized as follows. After a brief introduction, the research question on which the entire paper is based is proposed:

- **RQ**: What is the state of research regarding Large Language Models and Pandemic scenario?

To comprehensively address the main research question, we structured our investigation through a series of targeted sub-queries that explore different aspects of LLMs in pandemic prediction. Our analysis follows a methodical approach, beginning with a detailed examination of research methodology and selection criteria for relevant works. We then present an in-depth analysis of our findings, offering insights that span both editorial trends and medical applications. This investigation not only shows the significant advances made in applying LLMs to healthcare challenges but also reveals emerging patterns and potential future directions in the field. To establish a proper foundation for our analysis, we first present an overview of LLMs and their current applications in healthcare, highlighting the unique capabilities and challenges these technologies bring to medical applications.

A. Large Language Models (LLMs)

Large Language Models (LLMs) represent a breakthrough in artificial intelligence, offering sophisticated capabilities in natural language understanding and generation. These models leverage advanced machine learning techniques and are trained on vast corpora of textual data, enabling them to perform complex language tasks with remarkable accuracy. In the medical domain, LLMs demonstrate particular promise by efficiently processing and analyzing clinical documentation, extracting crucial information from medical records [4], and providing rapid, evidence-based responses to clinical queries [5]. Their ability to understand context and identify subtle patterns makes them especially valuable for processing unstructured medical data, potentially accelerating diagnostic processes and improving patient care outcomes. Unlike traditional rule-based systems, LLMs can adapt to the nuanced and evolving nature of medical knowledge, making them particularly suitable for healthcare applications. Most LLMs are built on the *Transformer* architecture [6], which use a mechanism called self-attention. This mechanism allows the model to weigh the importance of each word in a sequence relative to others, capturing long-range dependencies in text effectively. The original *Transformer* inspired two main branches of development in natural language processing (NLP): *BERT* (Bidirectional Encoder Representations from Transformers) and *GPT* (Generative Pre-trained Transformer). These branches reflect distinct focus areas in model architecture and applications. BERT uses only the encoder portion of the original architecture. It focuses on understanding text by capturing contextual relationships in both directions (left-to-right and right-to-left) simultaneously [7]. Instead, GPT uses only the decoder portion of the Transformer, making it optimized for text generation. GPT processes text left-to-right in a unidirectional manner, predicting the next word based on previous ones [8]. In the medical field, the significance of this architectural differentiation becomes even more apparent since depending on the task to be performed, one architecture may be preferred over the other. These models could not only enhance efficiency for healthcare providers but also improve accessibility to reliable information for patients.f On one hand, we have an architecture capable of supporting the creation of solutions to assist domain experts by identifying significant patterns within patients' medical records. On the other hand, we have an architecture designed to facilitate the development of chat-bots that can assist in self-diagnosis, aiming to optimize diagnostic processes.

II. Research Process

This section provides an introduction to the implementation process of literature review. To highlight the development process, a review plan is proposed, following the procedural steps outlined in the works [9], [10].

Our research process adheres to the following steps: Formulation of research, questions, development and execution of a search strategy, development of selection criteria for literature, execution of literature selection procedures, formulation and implementation of literature extraction strategies and, finally, data extraction and comprehensive analysis of extracted data. The sub-sections below provide a detailed account of how each of these steps has been implemented, during the development of this work.

A. Research Questions

- **RQ**: What is the state of research regarding Large Language Models and Pandemic scenario?
  - **Sub-RQ1** What is the trend in publications over the years?
  - **Sub-RQ2.1** Which publishers have most published works in this domain?
  - **Sub-RQ2.2** What types of works have been published?
  - **Sub-RQ3** What are the key research trends from a keyword perspective?

B. Search Strategy

1) Used Databases: Results from three well-known databases were used in this work:

- Scopus;
- Web of Science;
- PubMed;

2) Time Range: In literature, it is possible to find articles related to LLMs and the pandemic field starting from 2009. Considering the already small interval and relatively small production, the entire production interval from 2009 to September 2024 was considered, date that coincides with the beginning of the drafting of this work.

3) Search String: The research question was formalized in the query outlined in Table I.

TABLE I
RESEARCH STRING USED TO IMPLEMENT RESEARCH QUESTION

Research String
(LLM OR (Large AND Language AND Model*)) AND (Pandemic*)*

4) Search Results: The search results are shown in the Table II.

TABLE II
SEARCH RESULTS IN USED DATABASES

Scopus	Web of Science	PubMed	Total
433	607	212	1252

C. Selection Criteria

In line with the intended nature of the work, the following selection criteria were established:

- Include only works written in English;
- Include only papers published in international conference proceedings and articles published in journals, also considering reviews and systematic literature reviews, while excluding other types of works;
- Consider only works in their final version.

D. Selection Procedure

The paper selection process followed a rigorous two-stage filtering approach to ensure comprehensive yet focused coverage of the research domain. The initial corpus of 1.252 publications, identified within our specified time range, underwent systematic refinement through two sequential filtering stages.

- **Step 1**: In the first stage, we applied a language criterion, retaining only English-language publications, which reduced the corpus to 1.238 papers.
- **Step 2**: The second stage focused on publication type and completion status: we included peer-reviewed articles, conference papers, and literature reviews (both traditional and systematic), while excluding works in pre-print or non-final states. This second filter aligned with our objective of analyzing verified, peer-reviewed research contributions, resulting in a final dataset of 849 publications.

This reduction process ensured that our analysis remained focused on relevant, verified research while maintaining sufficient breadth to capture significant trends in the field Table III, reports the total amount of analyzed papers , distinguishing the step phase and the used database.

TABLE III
SELECTION PROCEDURE STEPS AND RESULTS

	Scopus	Web of Science	PubMed	Total
Step 1	428	598	212	1238
Step 2	405	421	23	849

E. Extraction Strategies

Our extraction methodology prioritized bibliometric and statistical analysis over detailed content examination, aligning with our research objectives of identifying broad patterns and trends in LLM applications for pandemic prediction. We developed a unified database structure by consolidating data from Scopus, Web of Science, and PubMed into a single dataset. The integration process required careful consideration of unique identification methods across sources. While Digital Object Identifiers (DOIs) typically serve as standard unique identifiers in academic literature, their inconsistent availability and occasional discrepancies across databases necessitated the use of publication titles as primary keys for our consolidated database. We structured our database according to the attributes detailed in Table IV, ensuring systematic capture of relevant metadata while maintaining data integrity across all consolidated sources. This approach enabled robust statistical analysis while minimizing data redundancy and inconsistencies in our dataset.

TABLE IV
SEARCH RESULTS IN USED DATABASES

Attribute	**Description**
Title	Title of work
DOI	Digital Object Identifier
Author Keywords	The keywords chosen by the authors
Index Keywords	The keywords used by the databases
Authors	Authors of work
Document Type	The type of work
Publisher	The publisher of work
Publication Year	Year of publication
Open Access	The degree of accessibility of work

III. RESULTS

The following results address the research questions outlined in Section 2.2, focusing respectively on trends in research output over time, editorial patterns, and thematic developments.

A. Annual Trend

The temporal analysis of research publications reveals distinct phases in the evolution of LLM applications in pandemic research. Prior to 2019, publication rates maintained a relatively consistent but modest baseline, reflecting the nascent state of LLM applications in epidemiology (Fig. 2). The emergence of COVID-19 in 2020 catalyzed a significant surge in research activity, marking a pivotal shift in both volume and focus of publications. This momentum culminated in 2022, when research output reached its peak, driven by widespread recognition of AI's potential in pandemic management and increased availability of pandemic-related datasets. Recent publication trends suggest a gradual stabilization rather than decline, with research focus shifting from immediate pandemic response to more sophisticated applications of LLMs in predictive epidemiology and healthcare systems. This evolution reflects the field's maturation from crisis-driven research to the investigation of LLMs' role in pandemic preparedness and prediction.

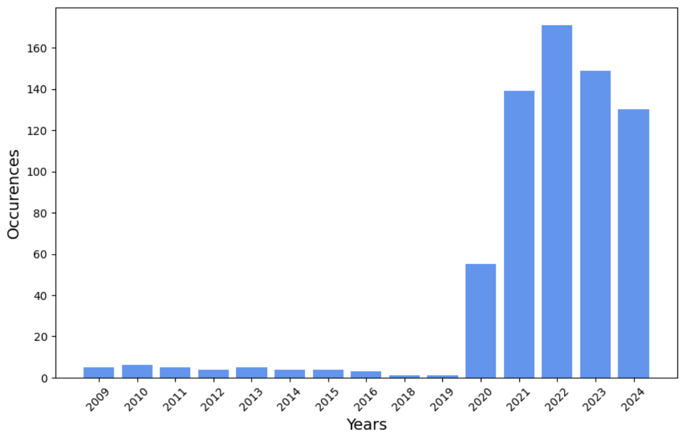

Fig. 2. Distribution of publications per year (2009 - 2024)

B. Editorial Trends

From an editorial perspective, in most cases within the analyzed datasets, the publisher was not specified, minimizing the available information on the subject. However, excluding works where the publisher is not indicated and those publishers appearing in the database with only one occurrence (grouped in the graph under the label "others"), it can be observed that one of the major publishers is Springer, with a total of 13 works, followed by Routledge, with a total of 7 works (Fig. 3). The works for which publisher information is unavailable amount to a significant 376.

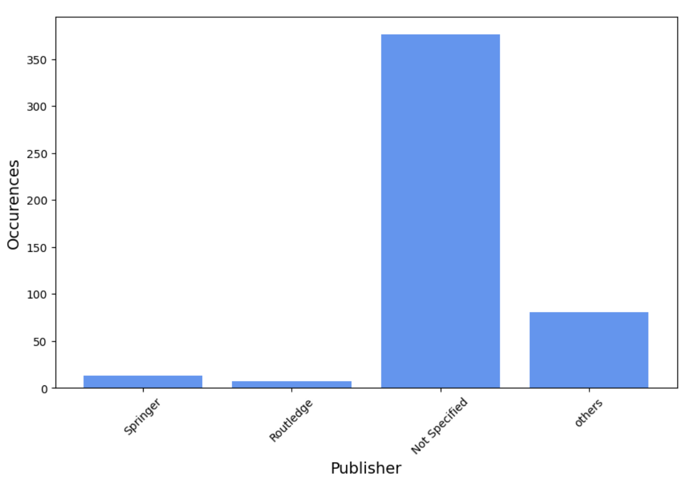

Fig. 3. Main publishers (2009 - 2024)

Regarding the types of published works, the majority of the output consists of journal articles, with a total of 497 occurrences (Fig. 4). Another significant portion is made up of conference papers, totaling 122 works, followed by reviews, which account for 37 works. Excluded from the chart are the types "proceedings paper" and "short survey," each with a single occurrence, as their low numbers made them unreadable in the visualization.

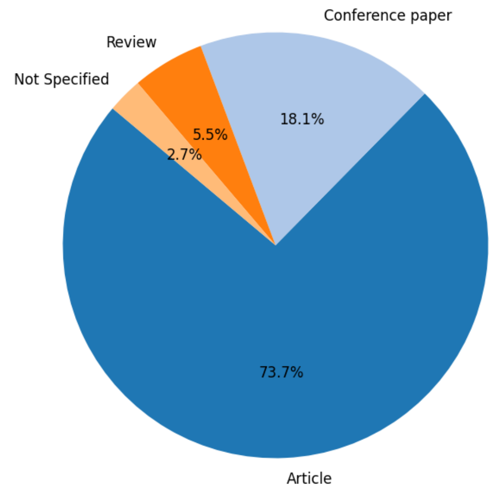

Fig. 4. Type of published works (2009-2024)

Of the 37 literature reviews identified, none deal with the topic specifically and for this reason no study of related work was presented in the proposed work.

C. Topics Trends

Our thematic analysis employed a three-phase approach to keyword examination, designed to capture the multifaceted nature of LLM applications in pandemic research. The first phase analyzed author-selected keywords, providing insight into researchers' own categorization of their work and revealing emerging trends in research focus. The second phase examined database-indexed keywords, offering a standardized perspective on research categorization and enabling comparison across studies. The final phase concentrated specifically on medical terminology within author-provided keywords, examining the intersection between pandemic research and broader medical domains. This layered analytical approach enabled us to identify not only explicit research trends but also subtle patterns in the relationship between pandemic studies and various medical subdisciplines. By triangulating these different keyword perspectives, we could map the evolution of research priorities and uncover previously unrecognized connections between pandemic dynamics and specific medical conditions. This methodological approach proved particularly valuable in identifying potential early indicators of pandemic emergence across different medical contexts.

The main keyword used by the authors is *"Natural Language Processing"* with a total of 81 occurrences, followed by *"Sentiment Analysis"* with 58 occurrences, as shown in the Fig. 5.

Given the domain, it is unsurprising to find a large portion of works using keywords such as *"COVID-19"*, *"SARS-CoV-2"*, or *"coronavirus"* from a medical perspective. Similarly,

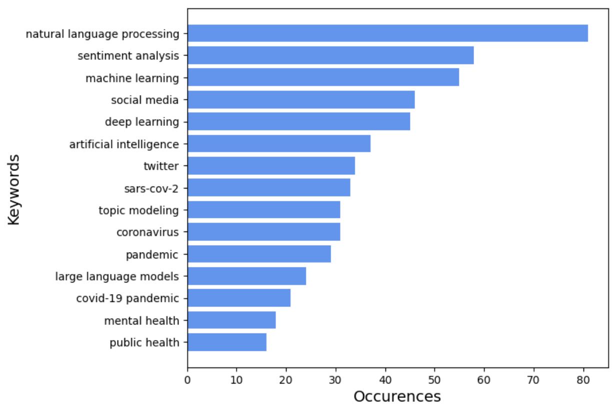

Fig. 5. Main keywords used by the authors (2009-2024)

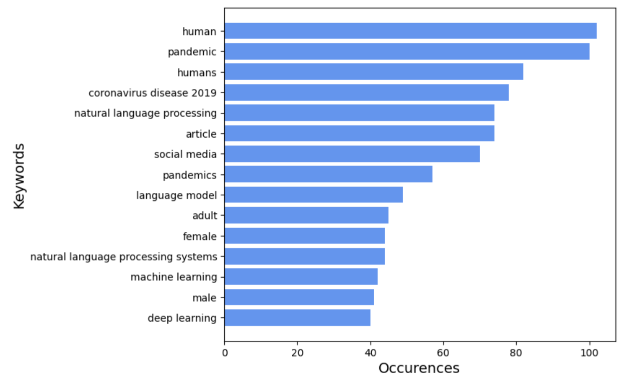

Fig. 7. Main keywords indexed in databases (2009-2024)

from the standpoint of AI solutions provided, it is equally expected to observe works employing keywords like *"large language model"* or *"artificial intelligence"*. However, two important observations can be made, both from a medical and a computational perspective, as seen in the graph. From the technological perspectove, it is evident that works utilizing machine learning techniques slightly outnumber those employing deep learning techniques. One possible explanation is the intent to offer more transparent solutions in terms of results, given the medical domain and the less predictable nature of deep learning models. From the medical perspective, it is noteworthy to see keywords like *"mental health"* frequently recurring in the works. This highlights the significant impact of a pandemic—whether past or potential future—on people's mental health. This observation will be confirmed by the analyses in the following sections. An overview of the keywords analyzed in this phase is presented in Fig. 6.

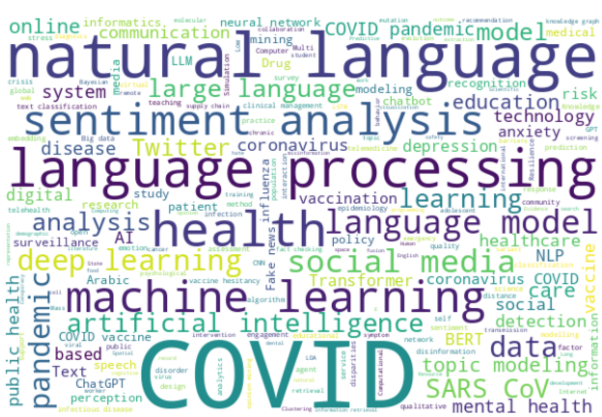

Fig. 6. Word cloud of main author keywords (2009-2024)

Moving on indexed keywords analysis, the main one is attributable to *"Human"*, with 102 occurrences, which, if we also consider occurrences of the same word in its plural form, amount to 184 (Fig. 7). This is followed by *"Pandemic*"* with 157 occurrences, supported by the keyword *"Coronavirus disease 2019"* with 78 occurrences.

The analysis of indexed keywords corroborates and extends our earlier findings while revealing additional nuances in research methodologies. Particularly noteworthy is the persistent preference for traditional machine learning approaches over deep learning techniques, though this distinction appears more subtle in the indexed keywords than in author-selected terms. This consistent pattern across different keyword sources suggests a deliberate methodological choice in the field rather than a mere artifact of keyword selection. An interesting observation can be made regarding works using the keywords *"female"* and *"male"*. Works using the keyword *"female"* are slightly more numerous than those using *"male"*, which could suggest that the analyses conducted in the reviewed works focused slightly more on data involving female patients, or that clinical and mental health disorders caused by the pandemic affected women more. This difference is generally minimal and, when compared to the occurrences of the keyword *"human*"*, it loses much of its significance. More interesting is the presence of the keyword *"adult*"*, which indicates that some of the most affected patients were those from the adult age group. Finally, the presence of the keyword "Social Media" highlights the importance that social media platforms have had, or could have, in a pandemic context. This observation is reinforced by the fact that, in the author-provided keywords analysis, the same keyword or others like *"twitter"* were identified. Again, an overview of the keywords analyzed in this phase is presented in Fig. 8.

D. Medical Topics

The third phase of our keyword analysis focused specifically on medical terminology, aiming to uncover potential relationships between pandemic patterns and other medical conditions. We developed a targeted approach to examine medical keywords, with particular attention to terms related to specific diseases and cancers. This methodological refinement sought to identify potential correlations between pandemic emergence and existing medical conditions, exploring whether certain health conditions might serve as early indicators of pandemic risk. Our investigation specifically examined immune system disorders, responses to pathogens, and disease

Fig. 8. Word cloud of main database indexed keywords (2009-2024)

progression patterns that could signal emerging pandemic threats. While this analysis revealed complex interconnections within medical terminology, it did not identify statistically significant correlations between specific medical conditions and pandemic emergence. This finding, while not supporting our initial hypothesis about disease-specific pandemic indicators, suggests that pandemic prediction may require a more multifaceted approach incorporating broader health metrics beyond individual disease categories. By maintaining a broader scope on the medical keywords, the occurrences of the top 15 most represented keywords in the database were analyzed. Excluding the keywords that are obviously directly related to the pandemic, such as *"COVID-19 pandemic"*, *"SARS-CoV-2"*, or *"coronavirus"*, an interesting observation can be made regarding those related to mental disorders (Fig. 9).

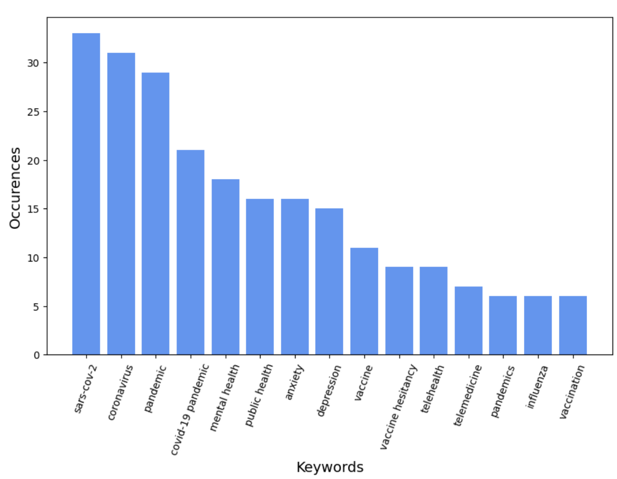

Fig. 9. Main medical topics (2009-2024)

The observations made during the analysis of the keywords chosen by the authors are confirmed by the presence of keywords with several occurrences related to mental health. The keyword *"mental health"*, with a total of 18 occurrences, is followed by keywords such as *"anxiety"* and *"depression"* with 16 and 15 occurrences respectively in the database. This demonstrates how a pandemic context, beyond tangible effects like influenza disorders, has a significant impact on people's mental health, especially those forced to stay in the same places for safety reasons. Another widely discussed topic is vaccines, represented by keywords like *"vaccine"*, *"vaccination"*, but also *"vaccine hesitancy"*, highlighting skepticism towards vaccination methods or vaccines themselves during the pandemic period. Finally, a topic closely related to the pandemic is telemedicine, represented by keywords like *"telehealth"* and *"telemedicine"*, which could, as it has in past years, represent the only feasible solution in situations where movement is highly restricted or impossible, especially for elderly individuals.

IV. Conclusion, Discussion and Future Works

This literature review has examined the intersection of Large Language Models and pandemic scenarios through the analysis of 849 publications spanning from 2009 to 2024. Our investigation reveals several significant patterns and insights that contribute to understanding both the current state and future potential of LLMs in pandemic prediction and management.

The temporal analysis of publications demonstrates a clear correlation between research interest and real-world pandemic events, with a marked increase in publications following the COVID-19 outbreak in 2020 and reaching its peak in 2022. This surge in research activity has generated valuable insights into both the technical capabilities and limitations of LLM applications in healthcare settings. A particularly noteworthy finding is the preference for traditional machine learning approaches over deep learning methods in medical applications, suggesting a prioritization of model interpretability over raw performance metrics. The research has identified several promising opportunities for LLM applications in pandemic prediction. The ability to process and analyze vast amounts of clinical data, combined with natural language processing capabilities, positions LLMs as powerful tools for early warning systems. The analysis also revealed gaps in current research that warrant further investigation. While many studies focus on text analysis and patient record processing, there is limited exploration of multimodal approaches that could integrate clinical data with imaging and real-time monitoring. Additionally, the relative scarcity of studies specifically addressing pandemic prediction, as opposed to pandemic response, suggests an important direction for future research.

Looking forward, several key areas emerge as critical for advancing the field. First, the development of interpretable LLM models specifically designed for medical applications could address current limitations in model transparency. To this aim the integration of logical reasoning frameworks [11], [13] with LLMs emerges as a promising direction for enhancing model explainability by complementing LLM capabilities. Second, the integration of privacy-preserving techniques into LLM architectures would facilitate broader adoption in healthcare settings. Third, a critical direction for future research lies in developing integrated multimodal approaches that combine LLMs' natural language processing capabilities with other data modalities [12]. In conclusion, future research should focus on developing solutions that balance analytical power with privacy protection, while maintaining the interpretability necessary for medical applications. The field would benefit from more targeted studies on predictive capabilities and the integration of diverse data sources, moving beyond current limitations to create more effective pandemic prediction systems. This conclusion synthesizes the paper's findings while providing critical analysis and future directions, maintaining a balanced perspective on both the potential and limitations of LLM applications in pandemic prediction.

ACKNOWLEDGMENT

This work was supported by the POS RADIOAMICA project funded by the Italian Minister of Health (CUP: H53C2200065000).

REFERENCES

[1] COVID-19 World Health Organization Statistics, https://data.who.int/dashboards/covid19/cases, [Accessed 16-10-2024]

[2] Provisional COVID-19 Mortality Surveillance, https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm, [Accessed 18-10-2024]

[3] Caroprese, L., Ruga, T., Vocaturo, E., & Zumpano, E. (2023, December). Lung Cancer Detection via Federated Learning. In 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 3862-3867). IEEE.

[4] Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J. B., Aali, A., Bluethgen, C., ... & Chaudhari, A. S. (2023). Clinical text summarization: adapting large language models can outperform human experts. Research Square.

[5] Chakraborty, S., Paul, H., Ghatak, S., Pandey, S. K., Kumar, A., Singh, K. U., & Shah, M. A. (2022). An AI-based medical chatbot model for infectious disease prediction. Ieee Access, 10, 128469-128483.

[6] Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems.

[7] Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[8] Radford, A. (2018). Improving language understanding by generative pre-training.

[9] B.A Kitchenham. Guidelines for performing systematic literature reviews in software engineering (version 2.3).Technical Report EBSE-2007-01, Keele University and Durham University, UK, Jan 2007.

[10] Kraus Sascha, Matthias Breier, and Sonia Dasí-Rodríguez. "The art of crafting a systematic literature review in entrepreneurship research." International Entrepreneurship and Management Journal 16 (2020): 1023-1042.

[11] Luciano Caroprese, Eugenio Vocaturo, Ester Zumpano: Argumentation approaches for explanaible AI in medical informatics. Intell. Syst. Appl. 16: 200109 (2022).

[12] Ruga, T. (2025). Classifying chest X-ray images with deep learning techniques: Challenges and explainable analysis. In Communications in Computer and Information Science. Communications in Computer and Information Science (pp. 343–350). doi:10.1007/978-3-031-70421-5_31

[13] Luciano Caroprese, Ester Zumpano, Bart Bogaerts: Computing Abductive Explanations. IEEE Intell. Syst. 37(6): 18-26 (2022)