A Survey of Time-Aware Language Models and Temporal Reasoning in AI

www.surveyx.cn

Abstract

This survey paper explores the transformative impact of temporal reasoning in artificial intelligence, emphasizing its critical role in enhancing language models' capabilities to process, interpret, and predict temporal information across diverse applications. The integration of temporal context in time-aware language models significantly improves their ability to retain and predict facts over time, underscoring the necessity of incorporating temporal knowledge for achieving more accurate and contextually relevant AI systems. Recent advancements, such as the TimeR 4 framework, demonstrate substantial improvements in the temporal reasoning abilities of large language models, achieving state-of-the-art performance on temporal datasets. Despite these advancements, challenges persist in effectively capturing temporal nuances and dependencies. The survey identifies the strengths and weaknesses of existing models and suggests potential areas for future research to enhance temporal reasoning in AI. It highlights the importance of developing robust benchmarks and sophisticated methodologies to address these challenges and improve the generalizability of AI systems. As AI continues to evolve, integrating sophisticated temporal reasoning mechanisms will be crucial for advancing their capabilities and ensuring their relevance in dynamic and complex environments.

1 Introduction

1.1 Significance of Temporal Reasoning in AI

Temporal reasoning serves as a critical pillar in artificial intelligence, significantly augmenting language models' ability to process and interpret time-sensitive information. This capability is vital for accurately understanding event durations and sequences, particularly in applications such as video content analysis, where discerning the order and subtleties of events is essential for generating coherent narratives and predictions. The integration of temporal reasoning into AI systems is especially crucial in multimodal contexts, where visual data is combined with other information types to enhance reasoning capacities [1].

In Large Language Models (LLMs), temporal reasoning is an underexplored yet essential area, enabling models to retain and reason about temporal information across diverse applications [2]. By facilitating adaptation to evolving word meanings over time, temporal reasoning enhances linguistic capabilities and maintains relevance in dynamic contexts [3]. Furthermore, it is indispensable for processing irregularly sampled time series data, a common challenge in real-world scenarios such as those involving IoT and wearable devices [4].

Accurate future event prediction exemplifies another critical dimension of temporal reasoning in AI. Its application in traffic flow prediction necessitates effective integration of both temporal and spatial factors [5]. Moreover, human mobility prediction, which is integral to various domains such as epidemic modeling, transport planning, and emergency responses, relies heavily on the accuracy afforded by temporal reasoning [6]. As AI technologies progress, the incorporation of advanced

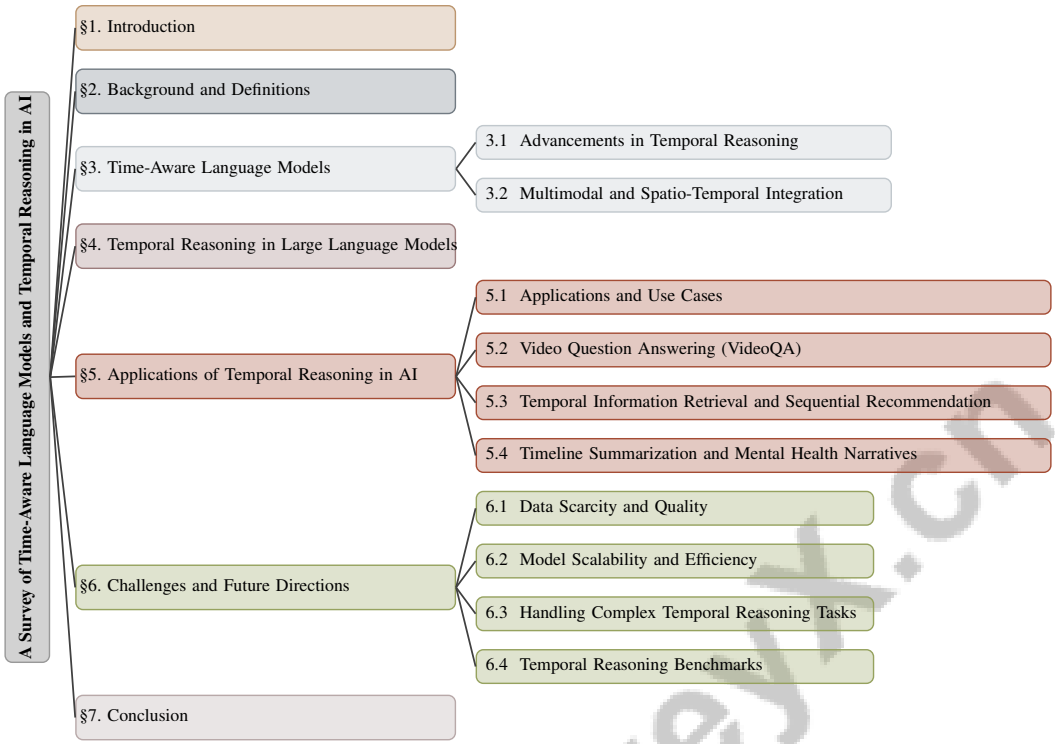


Figure 1: chapter structure

temporal reasoning mechanisms will be vital for enhancing language models and their applications across various fields, ensuring their effectiveness in an ever-evolving landscape.

1.2 Scope of the Survey

This survey delineates the landscape of time-aware language models and their temporal reasoning capabilities, emphasizing both theoretical advancements and practical applications. It investigates innovative methodologies for integrating temporal information into language understanding, thereby enhancing models' ability to manage temporal relevance effectively [3]. The survey also evaluates LLMs' performance in temporal reasoning, particularly in forecasting future events and formulating coherent explanations for these predictions [2].

A significant focus is placed on misinformation detection, where the temporal dynamics of knowledge are crucial. The survey explores frameworks that utilize structured representations to enhance temporal reasoning. It underscores the potential of time-aware language models to accurately retain and predict temporally-scoped knowledge, highlighting their diverse applications [3].

Additionally, the survey covers LSTM-based approaches for anomaly detection across sectors such as manufacturing, healthcare, and communication systems, concentrating on neural network methodologies [7]. It assesses temporal knowledge retention and reasoning capabilities through multiple-choice questions (MCQs) and categorizes Knowledge Graph Representation (KGR) models into static, temporal, and multimodal types, along with various reasoning techniques and scenarios.

The survey acknowledges the importance of chronological context in translation tasks, particularly concerning ancient texts and their historical evolution. It addresses existing limitations in time-aware language models that tend to be overly task-specific, advocating for more versatile approaches. Furthermore, it discusses the application of deep learning in spatio-temporal data modeling, enhancing understanding and predictive capabilities in fields such as transportation, climate science, and neuroscience [7].

Benchmarks are highlighted for their role in evaluating temporal reasoning capabilities in LLMs, providing a systematic approach to measure knowledge alignment with the correct temporal context and fostering further research in this domain [2]. The survey also emphasizes time-aware language

models' applications in understanding temporal context, including benchmarks for temporal question-answering.

Finally, the survey addresses challenges in dealing with historical texts lacking explicit temporal markers, underscoring the necessity for robust temporal reasoning capabilities. It covers the evolution of MultiModal Large Language Models (MM-LLMs) from image understanding to long video comprehension, highlighting the unique challenges associated with long videos [8].

1.3 Structure of the Survey

This survey is meticulously organized to provide a comprehensive analysis of time-aware language models and their temporal reasoning capabilities within artificial intelligence, addressing critical gaps in understanding how these models comprehend and process temporal information across diverse contexts and datasets [2, 9, 10]. The survey begins with an **Introduction**, establishing the significance of temporal reasoning in AI and outlining the paper's scope and structure.

In **Section 2**, *Background and Definitions*, core concepts and definitions are clarified, including time-aware language models and temporal reasoning LLMs. This section reviews the evolution of temporal reasoning in AI, highlighting historical developments and advancements.

Section 3, *Time-Aware Language Models*, investigates the development and capabilities of these models, focusing on advancements in temporal reasoning and methodologies. The section further explores the integration of multimodal and spatio-temporal data, which is essential for enhancing model performance in dynamic contexts.

Section 4, *Temporal Reasoning in Large Language Models*, examines LLMs' ability to perform temporal reasoning, discussing various approaches and techniques. This includes a detailed exploration of the integration of temporal knowledge graphs, as exemplified by frameworks like TG-LLM, which utilize temporal graphs for improved reasoning [11]. Additionally, the section discusses methods such as TempoBERT, which incorporates time as an additional context through time masking [12].

, *Applications of Temporal Reasoning in AI*, explores a range of practical applications across multiple domains, such as social media timeline summarization and natural language inference. It highlights how advancements in temporal reasoning can enhance decision-making processes and improve the understanding of complex narratives, particularly in contexts lacking explicit timestamps or requiring contextual inference to ascertain chronological order. The section emphasizes the transformative potential of integrating temporal reasoning into AI systems, evidenced by improved performance in summarizing lengthy event sequences and capturing emotional fluctuations in social media posts [13, 14]. This encompasses applications such as video question answering, temporal information retrieval, and timeline summarization.

, titled *Challenges and Future Directions*, highlights significant obstacles in developing time-aware models, including data scarcity that hampers the availability of temporal information for training and challenges in model scalability that complicate the adaptation of these models to handle extensive diachronic documents and diverse temporal contexts effectively [15, 16]. It suggests future research directions, emphasizing the need for improved benchmarks to evaluate temporal reasoning capabilities.

Finally, **Section 7**, *Conclusion*, summarizes key findings, reflecting on progress and potential future developments in the field while encouraging continued research. Throughout the survey, the integration of innovative mechanisms like the Temporal Attention mechanism, which extends the self-attention process of transformers to include temporal information, is highlighted [3].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Core Concepts and Definitions

Time-aware language models (TALMs) and temporal reasoning large language models (LLMs) are pivotal in AI for capturing temporal nuances. TALMs incorporate temporal data into text, facilitating the adaptation to temporal changes, which is crucial for tasks like dating historical documents and understanding event contexts [3]. Temporal reasoning LLMs excel in managing complex temporal

contexts, enhancing natural language inference and performing tasks such as temporal relation classification, predicting event relationships like 'before' or 'after' [17].

Benchmarks play a key role in evaluating LLMs' temporal reasoning capabilities, especially in time-sensitive question answering, which involves understanding temporal relationships across different periods [4]. However, many benchmarks fall short in complexity and scope, inadequately assessing LLMs' ability to tackle intricate temporal reasoning problems [2].

In multivariate time series (MTS) data, which often presents complex structures and label shortages, advanced models are needed to manage both short-term and long-term dependencies. This is particularly relevant in urban computing, where spatio-temporal dependencies are critical for predictive learning in domains such as transportation and public safety. Spatio-temporal representation learning, capturing temporal resolution, defines temporal reasoning in AI, with self-supervised methods addressing limitations in capturing both temporal resolution and long-short term characteristics in video data [18].

Temporal reasoning extends to video-based applications, essential for evaluating models' ability to reason over both textual and visual information. The development of benchmarks to enhance video LLMs' temporal reasoning capabilities, integrating spatial and temporal information, underscores the importance of leveraging textual temporal reasoning tasks from existing datasets [8]. Temporal Sentence Grounding (TSG), employing multimodal information, is crucial for contextualizing temporal reasoning in AI [19].

These core concepts and definitions establish the foundation for understanding TALMs and temporal reasoning LLMs, facilitating AI systems' dynamic adaptation to temporal information and enhancing decision-making across various domains. Capturing temporal-spectral relations further emphasizes understanding the global context in AI [20].

2.2 Evolution of Temporal Reasoning in AI

The evolution of temporal reasoning in AI has advanced significantly, adapting to complex and dynamic environments. Early efforts using traditional statistical methods struggled with spatio-temporal data complexities, particularly in urban settings where predicting future trends is crucial [21]. These methods often failed to capture dependencies within spatio-temporal data due to assumptions of independence among data points [7].

The emergence of deep learning frameworks marked a transformative shift, enabling sophisticated handling of time series data. Models like LSTNet have been pivotal in addressing temporal dependencies, enhancing feature engineering for spatio-temporal data [7]. However, effective benchmark development remains challenging, as many existing ones inadequately assess the full spectrum of temporal reasoning capabilities, often focusing on limited aspects such as single modalities or failing to capture the complexity of multi-answer and multi-hop reasoning.

Self-supervised learning methods have improved the ability to learn long-short term representations, yet challenges persist in managing temporal resolution in video data [18]. The transition from image understanding to long video understanding in multimodal large language models (MM-LLMs) highlights the ongoing evolution of temporal reasoning, as these models adapt to processing extended visual sequences [8].

Integrating temporal dynamics with multimodal data remains critical. Current benchmarks often do not address the complexities of temporal reasoning in real-world scenarios where multimodal understanding is essential [2]. The need for robust temporal reasoning mechanisms was underscored by the rapid dissemination of misinformation during the COVID-19 pandemic, necessitating AI systems to adapt swiftly to evolving narratives [4].

The evolution of temporal reasoning in AI reflects a trend towards dynamic models capable of handling complex temporal interactions. As AI technology advances, developing comprehensive benchmarks and integrating advanced temporal reasoning mechanisms is essential for addressing real-world intricacies, where understanding time-dependent information is critical. Recent studies indicate a substantial portion of knowledge relies on temporal context, with approximately 48% of qualifiers in widely-used knowledge bases being time-related. To enhance LLMs' temporal reasoning capabilities, innovative datasets and learning frameworks focus on multi-answer and multi-hop reasoning, vital for improving performance in temporal question answering tasks [10].

In recent years, advancements in time-aware language models have garnered significant attention within the field of artificial intelligence. These models are particularly notable for their ability to integrate temporal reasoning with multimodal inputs, which enhances their applicability across various domains. Figure 2 illustrates the hierarchical categorization of these advancements, focusing on key areas such as video understanding, traffic flow prediction, and video-based question answering. The figure highlights significant methodologies, including SViTT, VCOT, TimeChat, ATFM, and TCGL, which have contributed to the development of more sophisticated and contextually aware systems. This visual representation not only encapsulates the complexity of the advancements but also serves as a valuable reference for understanding the interconnections among different methodologies and their respective applications.

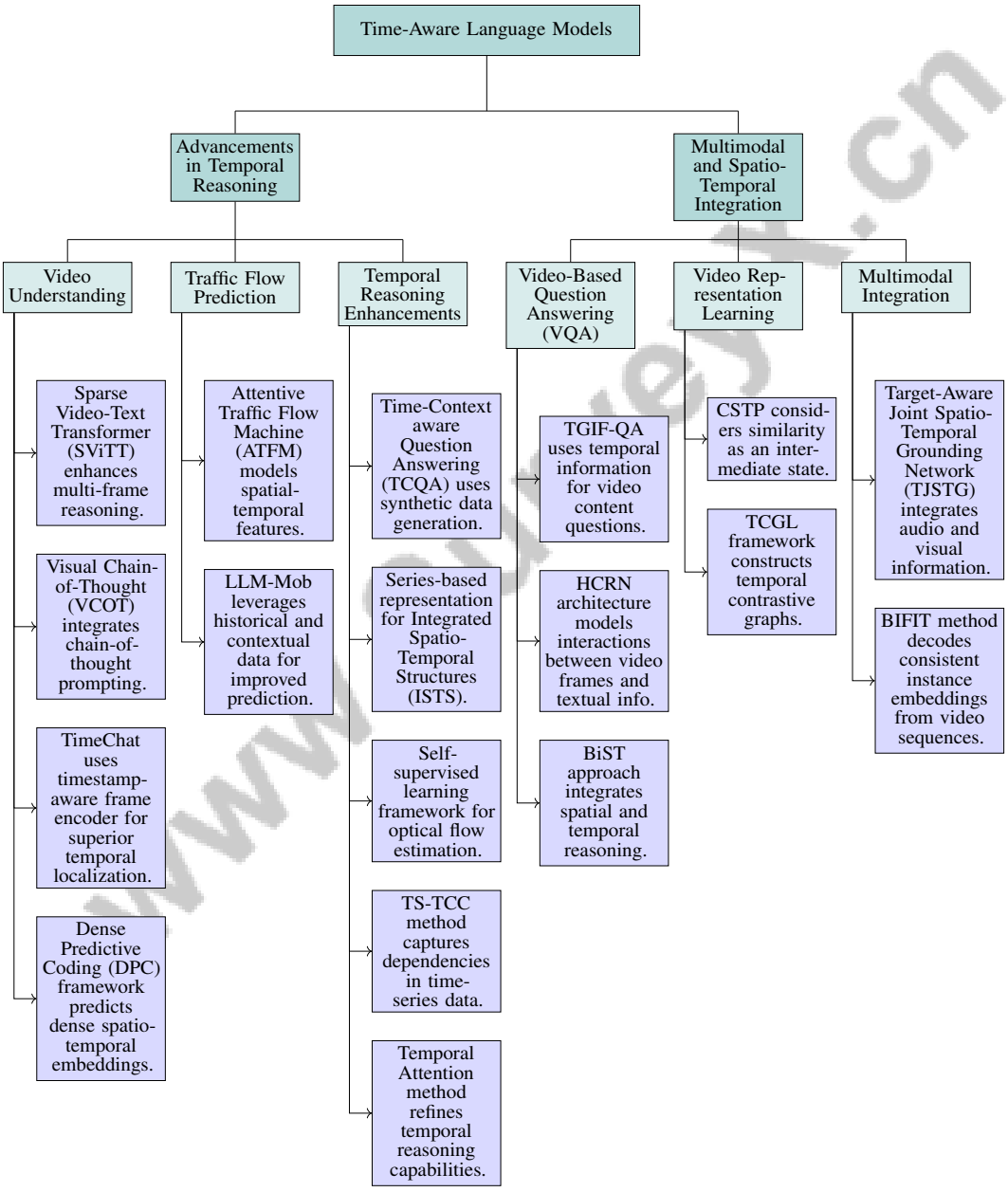


Figure 2: This figure illustrates the hierarchical categorization of advancements in time-aware language models, focusing on temporal reasoning and multimodal integration. Key areas include video understanding, traffic flow prediction, and video-based question answering, highlighting significant methodologies such as SViTT, VCOT, TimeChat, ATFM, and TCGL.

3 Time-Aware Language Models

3.1 Advancements in Temporal Reasoning

Recent progress in temporal reasoning has significantly improved the capabilities of time-aware language models. Models designed for long video understanding, such as those highlighted in recent research, have outperformed traditional short video approaches, indicating substantial advancements in temporal reasoning [8]. The Sparse Video-Text Transformer (SViTT) exemplifies this progress by employing edge and node sparsity to enhance multi-frame reasoning while reducing computational demands [22].

In video understanding, methodologies like the Visual Chain-of-Thought (VCOT) integrate chain-of-thought prompting with visual language grounding, advancing temporal reasoning by synthesizing multimodal infillings [1]. TimeChat further improves video comprehension with a timestamp-aware frame encoder and a sliding video Q-Former, offering superior temporal localization compared to previous video large language models (VidLLMs) [23]. The Dense Predictive Coding (DPC) framework enhances this domain by predicting dense spatio-temporal embeddings through a curriculum learning strategy that incrementally increases task complexity [24].

In traffic flow prediction, the Attentive Traffic Flow Machine (ATFM) uses an attention mechanism to dynamically model spatial-temporal features, advancing temporal reasoning [5]. The integration of large language models into mobility prediction frameworks, as demonstrated by LLM-Mob, leverages historical and contextual data to improve prediction accuracy and interpretability, surpassing traditional models [6].

The Time-Context aware Question Answering (TCQA) framework has pioneered synthetic data generation to enhance model performance in time-sensitive tasks, showcasing innovative approaches to boosting temporal reasoning [15]. Additionally, a series-based representation method for Integrated Spatio-Temporal Structures (ISTS) contrasts with traditional set-based and vector-based methods, optimizing the use of pre-trained language models (PLMs) [20].

A proposed self-supervised learning framework for optical flow estimation emphasizes the integration of temporal information from small event slices, underscoring key advancements in time-aware language models [25]. The TS-TCC method introduces temporal reasoning enhancements through a temporal contrasting module that captures dependencies in time-series data [4]. Furthermore, the Temporal Attention method modifies standard self-attention processes by incorporating temporal information directly into attention score computations, refining temporal reasoning capabilities [3].

These advancements underscore the critical role of temporal reasoning in augmenting the capabilities of time-aware language models, enabling more accurate processing, interpretation, and prediction of temporal information across diverse applications. Figure 3 illustrates recent advancements in temporal reasoning, categorizing them into three primary areas: video understanding, traffic flow prediction, and question answering and time series analysis. Each category highlights key methodologies that demonstrate significant progress in integrating temporal reasoning with AI models. The continued integration of advanced methodologies in temporal reasoning significantly enhances AI systems, particularly in information retrieval and natural language understanding. For instance, the development of TempRALM, a temporally-aware Retriever Augmented Language Model, demonstrates notable improvements in processing temporal queries, achieving up to a 74

3.2 Multimodal and Spatio-Temporal Integration

The integration of multimodal and spatio-temporal data in time-aware models represents a pivotal advancement in enhancing AI systems' interpretative and predictive capabilities. This approach leverages diverse data sources—textual, visual, auditory, and more—to foster a comprehensive understanding of temporal contexts. In video-based question answering (VQA) tasks, spatio-temporal reasoning enables models to learn from both spatial and temporal dimensions of video data. The TGIF-QA method exemplifies this by effectively utilizing temporal information to answer questions based on video content [26].

The HCRN architecture enhances this capability by employing a stack of Conditional Relation Networks to model interactions between video frames and associated textual information, thus improving the model's reasoning about temporal sequences [27]. Similarly, the BiST approach

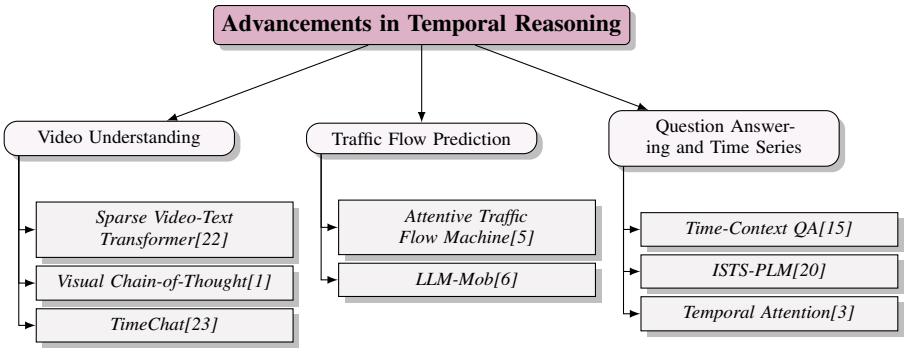


Figure 3: This figure illustrates recent advancements in temporal reasoning, categorizing them into three primary areas: video understanding, traffic flow prediction, and question answering and time series analysis. Each category highlights key methodologies that demonstrate significant progress in integrating temporal reasoning with AI models.

integrates spatial and temporal reasoning through a bidirectional reasoning strategy that facilitates dynamic information diffusion between spatial and temporal feature spaces, enhancing the model's processing of complex temporal interactions [28].

The CSTP method advances video representation learning by considering similarity as an intermediate state, aiding in capturing both spatial and temporal nuances in video data [29]. Moreover, the TCGL framework improves spatial-temporal representations by sampling non-overlapping video snippets and constructing intra-snippet and inter-snippet temporal contrastive graphs, thus deepening the understanding of temporal dynamics within video sequences [30].

The Target-Aware Joint Spatio-Temporal Grounding Network (TJSTG) integrates audio and visual information within a single-stream framework, enhancing the temporal association between audio and video, thereby demonstrating the effectiveness of multimodal integration in temporal reasoning tasks [31]. Additionally, the BIFIT method efficiently decodes consistent instance embeddings from video sequences by leveraging inter-frame interactions and bidirectional vision-language correlations, further enriching the spatio-temporal reasoning capabilities of AI models [32].

Surveys by Liang et al. and Jin et al. have organized research into static, temporal, and multi-modal knowledge graph representations, as well as spatial learning, temporal learning, and spatio-temporal fusion. These surveys emphasize the necessity of integrating various data types and reasoning techniques, including temporal awareness and multimodal processing, to effectively address predictive learning tasks, particularly in dynamic environments where information evolves over time and is presented in diverse formats [22, 20, 33].

The integration of multimodal and spatio-temporal data in time-aware models is crucial for advancing AI's ability to process and understand complex temporal information, thereby enhancing decision-making and predictive accuracy across various domains. The GEMINI dataset, with its multimodal and multilingual design, exemplifies the potential of such integration by providing a rich resource for testing and developing multimodal understanding [19].

4 Temporal Reasoning in Large Language Models

4.1 Techniques for Enhancing Temporal Reasoning

Advancing temporal reasoning in large language models (LLMs) requires sophisticated strategies for handling time-sensitive data. Temporal knowledge graphs are crucial, providing structured frameworks that enhance models' ability to extract and navigate intricate temporal relationships, thus improving task accuracy [3]. The Visual Chain-of-Thought (VCOT) method addresses sequential data gaps through visual guidance, enhancing reasoning in LLMs [1]. TimeChat tackles spatial-temporal semantic degradation in video LLMs, preserving temporal information in video content [23].

In traffic flow prediction, the Attentive Traffic Flow Machine (ATFM) integrates convolutional and LSTM layers to boost temporal reasoning, resulting in more precise traffic predictions [5].

LLM-Mob showcases effective use of LLMs' reasoning capabilities to improve temporal reasoning in mobility predictions from structured data [6]. The Time-Context aware Question Answering (TCQA) framework employs contrastive learning and Time-Context dependent Span Extraction (TCSE) to enhance QA models' temporal processing [15]. Meanwhile, the Temporal Attention mechanism allows dynamic adjustment of word importance based on temporal context, refining reasoning processes [3].

In video analysis, the Dense Predictive Coding (DPC) framework fosters semantic understanding by predicting future spatio-temporal embeddings from recent frames [24]. The Target-Aware Joint Spatio-Temporal Grounding Network (TJSTG) improves scene comprehension through target-aware spatial grounding and joint audio-visual temporal grounding, enhancing LLMs' temporal reasoning [31].

These methods underscore the importance of structured temporal data in enhancing LLMs' temporal reasoning, enabling effective interpretation and use of temporal information across applications like information retrieval and temporal relation classification. This integration not only boosts model performance but also addresses challenges of dynamic information, ensuring contextually accurate outputs in real-time scenarios [34, 33, 35]. As research progresses, these approaches will be vital in developing AI systems capable of navigating complex temporal landscapes.

As depicted in Figure 4, this figure illustrates the key techniques for enhancing temporal reasoning in large language models, categorized into three main areas: Temporal Knowledge Graphs, Visual and Video Methods, and Traffic and Mobility Prediction. Each category highlights specific methods or frameworks that contribute to the advancement of temporal reasoning capabilities in AI systems. The "Fact Retrieval and Answering System" illustrates a structured approach involving fact retrieval, rewriting, temporal attribute retrieval, and reasoning to deliver accurate responses, emphasizing structured information retrieval's role in temporal reasoning. The "Prompt and Model Response for a Question about Head Coaches" example shows the model's capability to deduce temporal sequences through simple interaction. The "Text-to-Temporal Graph Translation and Temporal Graph Reasoning" example demonstrates translating textual information into a temporal graph for advanced reasoning via graph-based methods. These examples collectively highlight diverse methodologies enhancing LLMs' capabilities in handling time-based queries with increased accuracy and depth [9, 36, 11].

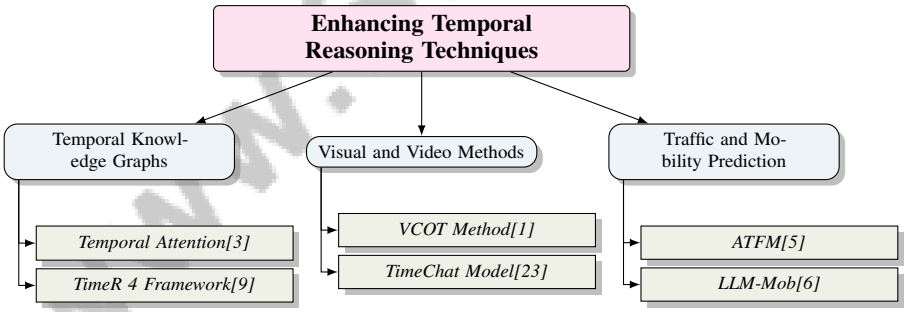


Figure 4: This figure illustrates the key techniques for enhancing temporal reasoning in large language models, categorized into three main areas: Temporal Knowledge Graphs, Visual and Video Methods, and Traffic and Mobility Prediction. Each category highlights specific methods or frameworks that contribute to the advancement of temporal reasoning capabilities in AI systems.

4.2 Integration of Temporal Knowledge Graphs

Integrating temporal knowledge graphs into LLMs significantly enhances their temporal reasoning abilities. These graphs offer structured representations of temporal information, enabling LLMs to efficiently navigate complex temporal relationships. The TimeR 4 framework exemplifies this by employing a sequence of retrieving, rewriting, and reranking processes to strengthen LLMs' temporal information processing [9].

In sequential recommendation tasks, the Tempura framework enhances LLMs' temporal awareness without requiring fine-tuning, providing a training-free, domain-agnostic solution that empowers

Method Name	Temporal Integration	Framework Examples	Evaluation and Validation
T4[9]	Timer 4	Timer 4	Multitq
T[34]	Temporal Structure Analysis	Tempura	Ndcg@n Metrics
INSTA[37]	-	Videoinsta	Accuracy Metric
CREMA[38]	-	-	Crema

Table 1: Comparison of various frameworks integrating temporal knowledge graphs into large language models (LLMs), highlighting their methods of temporal integration, example frameworks, and evaluation metrics. The table provides a concise overview of the distinct approaches and validation techniques employed to enhance temporal reasoning in LLMs.

LLMs to effectively utilize temporal information [34]. This is crucial for applications where temporal dynamics are key to decision-making and predictions.

The VideoINSTA framework further illustrates the benefits of temporal knowledge integration, enhancing reasoning capabilities over long video content. Evaluations against various baselines underscore the importance of temporal reasoning in video-based applications, demonstrating improved performance in understanding lengthy video sequences [37].

CREMA serves as a comprehensive evaluation framework, validating the integration of temporal knowledge in LLMs across seven video-language reasoning tasks, including VideoQA and multimodal QA involving diverse data types such as audio and 3D point clouds. This validation highlights the robustness of temporal knowledge graphs in supporting various multimodal reasoning challenges [38].

Benchmark evaluations of models like GPT-3.5 and FLAN-T5 emphasize the critical role of temporal knowledge graphs in enhancing LLMs' temporal reasoning capabilities, providing insights into current models' strengths and limitations and guiding future research directions [2].

The integration of temporal knowledge graphs represents a significant advancement in AI, enabling LLMs to process and interpret temporal information with greater accuracy and relevance. This integration facilitates the development of advanced AI systems, such as TempRALM and TALM, designed to navigate complex temporal landscapes by incorporating both semantic and temporal dimensions in information retrieval and text analysis, thereby enhancing performance across diverse applications, including question answering, text dating, and event detection [33, 16]. Table 1 presents a comparative analysis of different frameworks that integrate temporal knowledge graphs into large language models, illustrating their methodologies, practical applications, and evaluation metrics.

5 Applications of Temporal Reasoning in AI

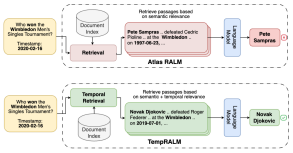
5.1 Applications and Use Cases

Temporal reasoning significantly enhances AI's ability to process time-sensitive information across various domains. In intelligent transportation systems, it is crucial for traffic flow prediction by leveraging historical mobility data to optimize traffic management [5]. In video question answering (VideoQA), temporal reasoning integrates spatio-temporal models with multimodal structures like subtitles and audio, improving visual content interpretation [27]. The VCGBench-Diverse benchmark, with 4,354 question-answer pairs from 877 videos, evaluates video understanding models' temporal reasoning capabilities, showcasing LLMs' potential in complex temporal challenges [39, 40].

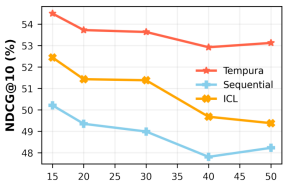
The Time-Context aware Question Answering (TCQA) framework demonstrates temporal reasoning's role in addressing deficiencies in models handling temporal expressions, enhancing accuracy for time-sensitive queries [15]. In misinformation detection, it is vital for analyzing knowledge dissemination dynamics, particularly during rapidly evolving situations like the COVID-19 pandemic, where timely intervention is crucial [41].

Frameworks like VideoINSTA and TCGL highlight temporal reasoning's practical applications in VideoQA, achieving improvements in multi-choice and open-question answering tasks and action recognition [37, 30]. The Target-Aware Joint Spatio-Temporal Grounding Network (TJSTG) further underscores its impact by enhancing audio-visual question answering accuracy compared to state-of-the-art methods [31].

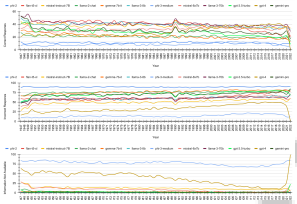
Temporal reasoning is also pivotal in timeline summarization on social media, where inferring chronological order from posts without explicit timestamps enhances information clarity. Recent LLM advancements suggest robust temporal reasoning can significantly improve summarizing lengthy narratives and multi-hop question-answering tasks, refining decision-making in dynamic contexts [13, 10]. As research progresses, integrating temporal reasoning is essential for developing AI systems capable of effectively handling time-sensitive information.



(a) Temporal Retrieval for Semantic and Temporal Relevance in Document Retrieval[33]



(b) The image shows a line graph with three lines representing different datasets: Tempura, Sequential, and ICL.[34]



(c) The image shows a line graph with multiple lines representing different models and datasets over time.[42]

Figure 5: Examples of Applications and Use Cases

As depicted in Figure 5, temporal reasoning is integral to AI, enabling effective analysis of time-varying data. In document retrieval, understanding semantic and temporal relevance enhances information retrieval processes. Comparisons between Atlas RALM and TempRALM illustrate how temporal reasoning improves retrieval performance by considering semantic content and temporal context. The consistent superiority of temporally-aware datasets like Tempura, Sequential, and ICL is highlighted, while another graph shows the evolution of models and datasets over time, emphasizing temporal reasoning's critical role in tracking trends. These examples collectively demonstrate temporal reasoning's diverse applications and significant impact across various AI use cases [33, 34, 42].

5.2 Video Question Answering (VideoQA)

The integration of temporal reasoning has substantially advanced VideoQA systems, essential for accurately interpreting and responding to video content queries. Traditional approaches often face challenges like hallucinations or ungrounded guesses due to over-reliance on linguistic cues [43]. Frameworks such as Flipped-VQA address these limitations by predicting all combinations of the $\langle$V, Q, A$\rangle$ triplet, enhancing answer grounding [43].

In Audio-Visual Scene-Aware Dialog (AVSD) tasks, temporal reasoning is crucial for evidence-based answering, highlighting the necessity of temporal cues for improved performance [44]. The TGIF-QA dataset exemplifies spatio-temporal reasoning's importance in VideoQA through tasks requiring models to interpret animated GIFs based on temporal sequences [26].

Advanced methodologies like Temporal Transformation Learning (TTL) have improved action recognition by accurately capturing motion dynamics, vital for effective temporal reasoning [45]. TemporalVLM demonstrates how capturing both local fine-grained and global temporal information enhances VideoQA task performance, enabling models to process complex temporal contexts more effectively [46].

The Multi-Modal Temporal Fusion (MMTF) method highlights capturing fine-grained temporal contexts for understanding temporally specific information, thereby enhancing VideoQA accuracy [47]. The BIFIT approach further improves performance by capturing temporal coherence and enhancing cross-modal interactions [32].

Benchmarks like ViteVQA and models such as ATP illustrate temporal reasoning's role in VideoQA, with ViteVQA emphasizing spatiotemporal reasoning of texts and visual information [48], while ATP demonstrates strong performance in video-language tasks [49]. Playback Rate Perception (PRP) in action recognition tasks underscores temporal reasoning's role in improving VideoQA systems [50]. The HCRN method has shown significant improvements in VideoQA tasks, validating its

effectiveness in handling various video formats [27], while the CSTP framework achieves state-of-the-art performance in video representation learning [29].

The CREMA framework efficiently integrates multiple modalities, enhancing reasoning capabilities in VideoQA systems [38]. The Bi-directional Spatio-Temporal Learning (BiST) method improves dialogue responses based on user queries referencing specific video segments and objects [28]. ST-LLM establishes a new state-of-the-art performance across multiple video benchmarks, demonstrating its effectiveness in understanding spatial-temporal sequences while maintaining efficiency [51].

Temporal reasoning is pivotal in advancing VideoQA systems, enabling them to process complex video content with greater accuracy. As AI research progresses, incorporating advanced temporal reasoning mechanisms will be crucial for developing systems capable of navigating intricate temporal dynamics in video data. This is particularly important as studies show that while Video LLMs exhibit promising video comprehension capabilities, they often struggle with tracking temporal changes and relationships. Recent advancements, such as TemporalVLM and VideoINSTA, highlight the necessity of integrating time-aware features to enhance performance in dense video captioning and temporal action segmentation tasks [52, 37, 53, 46].

5.3 Temporal Information Retrieval and Sequential Recommendation

Temporal reasoning is crucial for enhancing information retrieval and recommendation systems by considering the evolving nature of user interactions and content relevance over time. By integrating temporal dynamics like recency and chronological sequences, systems can improve accuracy and relevance, addressing content update challenges. Models like TempRALM demonstrate significant performance improvements by incorporating semantic and temporal relevance, while recent approaches to sequential recommendations leverage temporal awareness to better understand user behaviors in historical contexts [13, 34, 33]. In information retrieval, temporal reasoning allows systems to prioritize information based on its temporal context, improving search result accuracy, especially in dynamic environments where timeliness impacts decision-making.

Advancements in temporal reasoning have enhanced sequential recommendation systems by accurately modeling user behavior patterns over time, particularly through specialized prompting strategies that improve LLMs' ability to utilize temporal information from historical interactions. Evaluations on datasets like MovieLens-1M and Amazon Reviews validate the marked improvement in LLM performance in sequential recommendation tasks [34]. These systems leverage temporal information to predict user preferences and recommend items that align with evolving interests, facilitating personalized and contextually relevant recommendations.

Recent developments in temporal recommendation frameworks, such as Tempura, effectively employ temporal prompts to enhance LLMs' temporal awareness without extensive fine-tuning, showcasing the potential for training-free and domain-agnostic solutions [34]. The integration of temporal knowledge graphs in recommendation systems allows for structured capturing of temporal relationships, further enhancing recommendation precision by enabling models to navigate complex temporal interactions based on historical and contextual data [9].

The application of temporal reasoning in information retrieval and recommendation systems represents a significant advancement in AI, enabling more accurate, relevant, and timely results. As AI research progresses, incorporating advanced temporal reasoning mechanisms will be crucial for developing systems that adeptly manage user interactions and ensure content relevance across various timeframes, particularly in dynamic environments where information is continually updated, and user queries may involve multiple temporal contexts. This is especially important given the challenges faced by existing models, such as Retriever Augmented Language Models (RALMs), which struggle to differentiate between temporally relevant content and often fail to provide the most current information in response to user queries [2, 13, 14, 33].

5.4 Timeline Summarization and Mental Health Narratives

Temporal reasoning is essential in timeline summarization and understanding mental health narratives, enabling AI systems to generate contextually appropriate inferences from temporal inputs. In mental health, it facilitates the extraction of meaningful insights from patient narratives, enhancing the

understanding of healthcare journeys through improved temporal relation extraction in medical documents, which aids in tracking emotional and psychological changes over time [54].

In timeline summarization, especially on social media, temporal reasoning is crucial for accurately capturing the chronological sequence of events and associated emotional states. Current challenges in summarization arise from the absence of explicit timestamps, complicating the construction of coherent narratives from fragmented data [13]. Temporal reasoning provides a framework for understanding the dynamics of events and their emotional implications, thereby enhancing summary quality and relevance.

Moreover, temporal reasoning aids in generating contextually appropriate inferences within mental health narratives, enabling the identification of patterns and trends that may not be immediately apparent, which is vital for developing personalized interventions responsive to individuals' evolving needs [55]. As research advances, integrating sophisticated temporal reasoning mechanisms will be instrumental in improving the interpretative and predictive capabilities of AI systems in both timeline summarization and mental health applications.

6 Challenges and Future Directions

Addressing the challenges of time-aware language models requires a focus on data scarcity and quality issues. The lack of high-quality labeled time-series data is a significant barrier to advancing temporal reasoning capabilities. This section examines the obstacles posed by data scarcity, including the absence of comprehensive evaluation benchmarks and their impact on model training and performance, setting the stage for discussions on scalability, efficiency, and complex temporal reasoning tasks.

6.1 Data Scarcity and Quality

The development of time-aware language models is impeded by data scarcity and quality issues, primarily due to the limited availability of high-quality labeled time-series data necessary for training effective models [4]. This scarcity is compounded by the lack of comprehensive evaluation benchmarks, especially for complex tasks like long video understanding, which require extensive datasets for accurate training and evaluation [8]. Computational constraints in processing large datasets further complicate the development of robust models, potentially affecting their generalizability and reliability [17]. Existing benchmarks often focus on single-hop reasoning, neglecting the complexity of multi-hop reasoning tasks crucial for understanding intricate temporal dependencies, highlighting the need for more sophisticated benchmarks that simulate real-world temporal reasoning challenges effectively.

Furthermore, integrating diverse data types into time-aware models is hindered by the computational costs of updating numerous parameters for each modality, complicating the effective use of multimodal data sources [38]. The requirement for input texts to be accompanied by corresponding time points presents another significant obstacle, as such annotations are not always readily available [3]. This issue is particularly pronounced in dynamic environments, such as urban settings, where capturing complex spatio-temporal dependencies is essential for accurate modeling [5]. Moreover, the interpretability of current models remains a challenge, as many function as black boxes, obscuring their decision-making processes and hindering the understanding of temporal reasoning mechanisms [7]. The reliance on proprietary large language models (LLMs) and associated costs pose efficiency issues, particularly when API calls are required for each prediction, leading to potential hallucination problems in generated outputs [6].

To address these challenges, it is essential to create comprehensive and varied datasets encompassing a wide range of temporal contexts and question types. Establishing robust benchmarks that accurately reflect the complexities of temporal reasoning is crucial for evaluating large language models' capabilities in understanding time-dependent information. Such advancements will enhance the effectiveness and reliability of time-aware language models across various applications [36, 33, 10].

6.2 Model Scalability and Efficiency

The scalability and efficiency of time-aware language models are critical for their performance, particularly in complex temporal reasoning tasks. The computational complexity associated with

processing extensive video sequences poses significant challenges, as these models often require substantial resources to handle large volumes of data effectively. This complexity is compounded by the reliance on high-quality language expressions to ensure accurate temporal reasoning, which can constrain model performance [32].

As models evolve to handle larger datasets and a broader range of applications, the escalating demand for computational resources can lead to inefficiencies and bottlenecks, especially in real-time processing scenarios where timely information delivery is crucial. Traditional language models often struggle with temporal queries, failing to differentiate between document versions from various time points, underscoring the necessity of developing sophisticated models like temporally-aware Retriever Augmented Language Models (RALMs) that enhance performance by considering both semantic and temporal relevance without imposing additional computational burdens [56, 33]. This is especially pertinent in applications requiring rapid decision-making based on temporal data, such as traffic flow prediction and video question answering, where delays can significantly impact utility.

Integrating multimodal data into temporal models significantly increases complexity, as each modality—such as textual information and knowledge graphs—demands tailored processing techniques and distinct computational resources. This arises from the need to harmonize diverse data types to effectively extract temporal relationships and enhance understanding in applications like clinical narratives and long video analysis [33, 8, 54]. Such integration can lead to increased computational costs and challenges in maintaining model efficiency, particularly when updates are necessary for multiple parameters across different data types. Thus, scalable solutions that efficiently process and integrate diverse data sources are paramount for advancing time-aware language models.

Current research focuses on developing advanced algorithms and architectures, such as the TempRALM model, which enhance resource efficiency while achieving high accuracy in temporal reasoning tasks. This includes addressing limitations in existing LLMs that struggle with time-sensitive queries and ensuring effective differentiation of information based on temporal context, thereby improving performance in applications like question answering and timeline summarization [13, 33, 10]. Innovations such as sparse attention mechanisms and optimized data processing pipelines are being explored to enhance model scalability and efficiency, enabling more effective handling of complex temporal interactions across various applications.

6.3 Handling Complex Temporal Reasoning Tasks

Handling complex temporal reasoning tasks in AI involves addressing challenges stemming from the intricate nature of temporal information. A primary challenge is the inherent non-deterministic nature of future frames in videos, complicating the prediction of future representations and necessitating methods capable of effectively managing uncertainty [24]. This issue is particularly evident in traffic flow prediction, where dynamic variations in traffic patterns pose significant difficulties in accurately forecasting conditions [5].

The irregularity and asynchrony in Integrated Spatio-Temporal Structures (ISTS) complicate the identification of temporal semantics and dependencies, especially in scenarios characterized by extreme data sparsity [20]. In question answering, existing models often struggle to recognize and process temporal expressions, hindering performance in time-sensitive tasks [15]. This challenge is compounded by traditional augmentation techniques, which do not effectively translate to time-series data, further complicating the management of complex temporal reasoning tasks [4].

The integration of multimodal data introduces additional complexities, as some modalities may not contribute beneficial information, leading to increased computational costs without significant reasoning performance gains. For instance, inadequate modeling of dynamic audio-visual scenes may hinder performance, particularly when unrelated information could assist in answering questions [31]. Furthermore, challenges associated with semantic changes over time present significant obstacles in effectively managing complex temporal reasoning tasks [3].

Addressing these challenges necessitates ongoing research and development focused on enhancing model training for temporal reasoning and expanding datasets to include diverse modalities. Future research should explore the implications of multimodal reasoning in real-world applications and address current model limitations, such as efficiency and hallucination issues identified in mobility prediction frameworks. These initiatives are essential for enhancing AI systems' capabilities to navigate complex temporal dynamics, addressing challenges such as retrieving relevant and up-

to-date information from multiple versions of web content, accurately dating historical texts, and analyzing irregularly sampled time series data. By developing temporally aware models that integrate both semantic and temporal relevance, these efforts aim to improve AI performance in tasks like question answering, event detection, and long-term planning, ultimately leading to more sophisticated and reliable AI applications [20, 33, 57, 16].

6.4 Temporal Reasoning Benchmarks

Benchmark	Size	Domain	Task Format	Metric
DSTC10-AVSD[44]	1,450,754	Dialog Systems	Answer Generation	BLEU4, METEOR
MPS[58]	1,521,909	Instructional Manuals	Task Sequencing	Perfect Match Ratio, Accuracy
TeCFaP[59]	10,144	Temporal Reasoning	Sentence Completion	Temporal Factuality, Temporal Consistency
ToT[36]	49,280	Temporal Reasoning	Question Answering	Accuracy, F1-score
ExpTime[60]	26,000	Temporal Reasoning	Event Forecasting	F1-score, BLEU
PPNL[57]	160,000	Path Planning	Path Planning	Success Rate, Optimal Rate
TEMPLAMA[61]	50,310	Temporal Knowledge Representation	Fill-in-the-blank Queries	F1-score
TRC[35]	1,000	Temporal Relation Classification	Classification	micro-F1

Table 2: This table presents a comprehensive summary of representative benchmarks utilized for evaluating temporal reasoning capabilities in AI systems. It includes key details such as benchmark size, domain, task format, and performance metrics, highlighting the diversity and complexity necessary for effective assessment across various applications.

Benchmark	Size	Domain	Task Format	Metric
DSTC10-AVSD[44]	1,450,754	Dialog Systems	Answer Generation	BLEU4, METEOR
MPS[58]	1,521,909	Instructional Manuals	Task Sequencing	Perfect Match Ratio, Accuracy
TeCFaP[59]	10,144	Temporal Reasoning	Sentence Completion	Temporal Factuality, Temporal Consistency
ToT[36]	49,280	Temporal Reasoning	Question Answering	Accuracy, F1-score
ExpTime[60]	26,000	Temporal Reasoning	Event Forecasting	F1-score, BLEU
PPNL[57]	160,000	Path Planning	Path Planning	Success Rate, Optimal Rate
TEMPLAMA[61]	50,310	Temporal Knowledge Representation	Fill-in-the-blank Queries	F1-score
TRC[35]	1,000	Temporal Relation Classification	Classification	micro-F1

Table 3: This table presents a comprehensive summary of representative benchmarks utilized for evaluating temporal reasoning capabilities in AI systems. It includes key details such as benchmark size, domain, task format, and performance metrics, highlighting the diversity and complexity necessary for effective assessment across various applications.

The development of robust benchmarks is crucial for advancing the evaluation of temporal reasoning capabilities in AI systems. Existing benchmarks often lack the complexity required to thoroughly assess diverse temporal reasoning tasks, particularly those involving the retrieval and processing of temporally relevant information. Improved benchmarks are essential for providing a comprehensive framework for effective evaluation [3]. Table 3 provides a detailed summary of representative benchmarks that are pivotal for assessing temporal reasoning capabilities in AI systems across multiple domains.

Future research should focus on enhancing representation methods and integrating additional contextual information to improve modeling of Integrated Spatio-Temporal Structures (ISTS), indicating a need for improved benchmarks in this area [20]. More sophisticated benchmarks are also necessary to evaluate models like TS-TCC in learning representations from unlabeled time-series data, as highlighted by Eldele et al. [4].

In VideoQA, the need for enhanced benchmarks that effectively evaluate temporal reasoning capabilities is underscored by Maaz et al., who stress the necessity of robust assessments for video understanding models [25]. Furthermore, performance metrics in existing studies demonstrate the critical role of benchmarks in assessing temporal reasoning capabilities, suggesting that future research should enhance Spiking Neural Network (SNN) architectures and explore more powerful recurrent units to contribute to better benchmarks in this domain [25].

The development of enhanced benchmarks for temporal reasoning is vital for advancing research in this field, as it provides essential resources for evaluating the temporal reasoning capabilities of large language models (LLMs) while addressing limitations in current datasets. By introducing novel synthetic datasets and comprehensive evaluation frameworks, researchers can systematically investigate various factors affecting LLM performance in temporal reasoning tasks. This progress will ensure that AI systems are better equipped to accurately process and interpret temporal information across a wide range of applications, ultimately bridging the performance gap between LLMs and human understanding of time-dependent knowledge [62, 36, 63, 10]. These benchmarks will play a pivotal role in guiding the development of more sophisticated and contextually aware AI systems as research continues to evolve.

7 Conclusion

Temporal reasoning has emerged as a pivotal element in advancing artificial intelligence, particularly in enhancing the capabilities of language models to effectively process and predict time-dependent information. Time-aware language models have demonstrated their potential in significantly improving the memorization and prediction of temporally scoped knowledge, underscoring the necessity of incorporating temporal dimensions into AI systems for achieving enhanced contextual accuracy.

Recent frameworks like TimeR 4 have marked a substantial leap in the temporal reasoning capabilities of large language models, achieving impressive results on temporal datasets. These advancements highlight the progression of AI systems towards more nuanced understanding and reasoning about temporal data, as evidenced by superior performance in tasks that require such capabilities. Techniques such as TimeLMs and TaCOMET have further validated the integration of temporal knowledge, enhancing natural language processing tasks by embedding temporal reasoning.

Despite these strides, challenges remain in capturing the intricate nuances of temporal dependencies, with existing models often falling short. The introduction of the Temporal Attention mechanism has shown promise in surpassing traditional methods by effectively detecting semantic shifts over time, emphasizing the critical role of temporal reasoning in AI. This highlights the ongoing need for research to overcome current limitations and explore innovative ways to integrate temporal aspects into AI.

Furthermore, the survey identifies both strengths and areas for improvement in current knowledge graph representation models, pointing to potential research avenues for enhancing temporal reasoning in AI. The establishment of new benchmarks provides a framework for evaluating temporal reasoning tasks, demonstrating significant improvements and applicability across diverse fields.

The survey reaffirms the integral role of temporal reasoning in AI and calls for continued exploration of novel approaches and methodologies. As AI technologies evolve, incorporating advanced temporal reasoning mechanisms will be essential for enhancing their capabilities and maintaining their relevance in dynamic and complex environments.

References

[1] Daniel Rose, Vaishnavi Himakunthala, Andy Ouyang, Ryan He, Alex Mei, Yujie Lu, Michael Saxon, Chinmay Sonar, Diba Mirza, and William Yang Wang. Visual chain of thought: bridging logical gaps with multimodal infillings. *arXiv preprint arXiv:2305.02317*, 2023.

[2] Raghav Jain, Daivik Sojitra, Arkadeep Acharya, Sriparna Saha, Adam Jatowt, and Sandipan Dandapat. Do language models have a common sense regarding time? revisiting temporal commonsense reasoning in the era of large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6750–6774, 2023.

[3] Guy D Rosin and Kira Radinsky. Temporal attention for language models. *arXiv preprint arXiv:2202.02093*, 2022.

[4] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. *arXiv preprint arXiv:2106.14112*, 2021.

[5] Ieee transactions on intelligent.

[6] Xinglei Wang, Meng Fang, Zichao Zeng, and Tao Cheng. Where would i go next? large language models as human mobility predictors. *arXiv preprint arXiv:2308.15197*, 2023.

[7] Senzhang Wang, Jiannong Cao, and S Yu Philip. Deep learning for spatio-temporal data mining: A survey. *IEEE transactions on knowledge and data engineering*, 34(8):3681–3700, 2020.

[8] Heqing Zou, Tianze Luo, Guiyang Xie, Fengmao Lv, Guangcong Wang, Juanyang Chen, Zhuochen Wang, Hansheng Zhang, Huaijian Zhang, et al. From seconds to hours: Reviewing multimodal large language models on comprehensive long video understanding. *arXiv preprint arXiv:2409.18938*, 2024.

[9] Xinying Qian, Ying Zhang, Yu Zhao, Baohang Zhou, Xuhui Sui, Li Zhang, and Kehui Song. Timer4: Time-aware retrieval-augmented large language models for temporal knowledge graph question answering. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6942–6952, 2024.

[10] Qingyu Tan, Hwee Tou Ng, and Lidong Bing. Towards robust temporal reasoning of large language models via a multi-hop qa dataset and pseudo-instruction tuning. *arXiv preprint arXiv:2311.09821*, 2023.

[11] Siheng Xiong, Ali Payani, Ramana Kompella, and Faramarz Fekri. Large language models can learn temporal reasoning. *arXiv preprint arXiv:2401.06853*, 2024.

[12] Time masking for temporal langua.

[13] Jiayu Song, Mahmud Akhter, Dana Atzil Slonim, and Maria Liakata. Temporal reasoning for timeline summarisation in social media. *arXiv preprint arXiv:2501.00152*, 2024.

[14] Siddharth Vashishtha, Adam Poliak, Yash Kumar Lal, Benjamin Van Durme, and Aaron Steven White. Temporal reasoning in natural language inference. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4070–4078, 2020.

[15] Jungbin Son and Alice Oh. Time-aware representation learning for time-sensitive question answering. *arXiv preprint arXiv:2310.12585*, 2023.

[16] Han Ren, Hai Wang, Yajie Zhao, and Yafeng Ren. Time-aware language modeling for historical text dating. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 13646–13656, 2023.

[17] Himanshu Beniwal, Dishant Patel, D Kowsik, Hritik Ladia, Ankit Yadav, and Mayank Singh. Remember this event that year? assessing temporal information and understanding in large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 16239–16348, 2024.

[18] Yuan Yao, Chang Liu, Dezhao Luo, Yu Zhou, and Qixiang Ye. Video playback rate perception for self-supervised spatio-temporal representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6548–6557, 2020.

[19] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.

[20] Weijia Zhang, Chenlong Yin, Hao Liu, and Hui Xiong. Unleash the power of pre-trained language models for irregularly sampled time series. *arXiv preprint arXiv:2408.08328*, 2024.

[21] Guangyin Jin, Yuxuan Liang, Yuchen Fang, Zezhi Shao, Jincai Huang, Junbo Zhang, and Yu Zheng. Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 36(10):5388–5408, 2023.

[22] Yi Li, Kyle Min, Subarna Tripathi, and Nuno Vasconcelos. Svitt: Temporal learning of sparse video-text transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18919–18929, 2023.

[23] Shuhuai Ren, Linli Yao, Shicheng Li, Xu Sun, and Lu Hou. Timechat: A time-sensitive multimodal large language model for long video understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14313–14323, 2024.

[24] Tengda Han, Weidi Xie, and Andrew Zisserman. Video representation learning by dense predictive coding. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pages 0–0, 2019.

[25] Jesse Hagenaars, Federico Paredes-Vallés, and Guido De Croon. Self-supervised learning of event-based optical flow with spiking neural networks. *Advances in Neural Information Processing Systems*, 34:7167–7179, 2021.

[26] Yunseok Jang, Yale Song, Youngjae Yu, Youngjin Kim, and Gunhee Kim. Tgif-qa: Toward spatio-temporal reasoning in visual question answering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2758–2766, 2017.

[27] Hierarchical conditional relatio.

[28] Hung Le, Doyen Sahoo, Nancy F Chen, and Steven CH Hoi. Bist: Bi-directional spatio-temporal reasoning for video-grounded dialogues. *arXiv preprint arXiv:2010.10095*, 2020.

[29] Yujia Zhang, Lai-Man Po, Xuyuan Xu, Mengyang Liu, Yexin Wang, Weifeng Ou, Yuzhi Zhao, and Wing-Yin Yu. Contrastive spatio-temporal pretext learning for self-supervised video representation, 2021.

[30] Yang Liu, Keze Wang, Lingbo Liu, Haoyuan Lan, and Liang Lin. Tcgl: Temporal contrastive graph for self-supervised video representation learning. *IEEE Transactions on Image Processing*, 31:1978–1993, 2022.

[31] Yuanyuan Jiang and Jianqin Yin. Target-aware spatio-temporal reasoning via answering questions in dynamics audio-visual scenarios. *arXiv preprint arXiv:2305.12397*, 2023.

[32] Meng Lan, Fu Rong, Zuchao Li, Wei Yu, and Lefei Zhang. Bidirectional correlation-driven inter-frame interaction transformer for referring video object segmentation, 2023.

[33] Anoushka Gade and Jorjeta Jetcheva. It's about time: Incorporating temporality in retrieval augmented language models. *arXiv preprint arXiv:2401.13222*, 2024.

[34] Zhendong Chu, Zichao Wang, Ruiyi Zhang, Yangfeng Ji, Hongning Wang, and Tong Sun. Improve temporal awareness of llms for domain-general sequential recommendation. In *ICML 2024 Workshop on In-Context Learning*, 2024.

[35] Gabriel Roccabruna, Massimo Rizzoli, and Giuseppe Riccardi. Will llms replace the encoder-only models in temporal relation classification? *arXiv preprint arXiv:2410.10476*, 2024.

[36] Bahare Fatemi, Mehran Kazemi, Anton Tsitsulin, Karishma Malkan, Jinyeong Yim, John Palowitch, Sungyong Seo, Jonathan Halcrow, and Bryan Perozzi. Test of time: A benchmark for evaluating llms on temporal reasoning. *arXiv preprint arXiv:2406.09170*, 2024.

[37] Ruotong Liao, Max Erler, Huiyu Wang, Guangyao Zhai, Gengyuan Zhang, Yunpu Ma, and Volker Tresp. Videoinsta: Zero-shot long video understanding via informative spatial-temporal reasoning with llms. *arXiv preprint arXiv:2409.20365*, 2024.

[38] Shoubin Yu, Jaehong Yoon, and Mohit Bansal. Generalizable and efficient video-language reasoning via multimodal modular fusion. In *The Thirteenth International Conference on Learning Representations*.

[39] Muhammad Maaz, Hanoona Rasheed, Salman Khan, and Fahad Khan. Videogpt+: Integrating image and video encoders for enhanced video understanding, 2024.

[40] Houlun Chen, Xin Wang, Hong Chen, Zihan Song, Jia Jia, and Wenwu Zhu. Grounding-prompter: Prompting llm with multimodal information for temporal sentence grounding in long videos. *arXiv preprint arXiv:2312.17117*, 2023.

[41] Abhijit Suprem, Sanjyot Vaidya, Joao Eduardo Ferreira, and Calton Pu. Time-aware datasets are adaptive knowledgebases for the new normal. *arXiv preprint arXiv:2211.12508*, 2022.

[42] Remember this event that year? a.

[43] Dohwan Ko, Ji Soo Lee, Wooyoung Kang, Byungseok Roh, and Hyunwoo J Kim. Large language models are temporal and causal reasoners for video question answering. *arXiv preprint arXiv:2310.15747*, 2023.

[44] Ankit Shah, Shijie Geng, Peng Gao, Anoop Cherian, Takaaki Hori, Tim K Marks, Jonathan Le Roux, and Chiori Hori. Audio-visual scene-aware dialog and reasoning using audio-visual transformers with joint student-teacher learning. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7732–7736. IEEE, 2022.

[45] Video representation learning by.

[46] Fawad Javed Fateh, Umer Ahmed, Hamza Khan, M Zeeshan Zia, and Quoc-Huy Tran. Video llms for temporal reasoning in long videos. *arXiv preprint arXiv:2412.02930*, 2024.

[47] Mobeen Ahmad, Geonwoo Park, Dongchan Park, and Sanguk Park. Mmtf: Multi-modal temporal fusion for commonsense video question answering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4657–4662, 2023.

[48] Minyi Zhao, Bingjia Li, Jie Wang, Wanqing Li, Wenjing Zhou, Lan Zhang, Shijie Xuyang, Zhihang Yu, Xinkun Yu, Guangze Li, et al. Towards video text visual question answering: Benchmark and baseline. *Advances in Neural Information Processing Systems*, 35:35549–35562, 2022.

[49] Shyamal Buch, Cristóbal Eyzaguirre, Adrien Gaidon, Jiajun Wu, Li Fei-Fei, and Juan Carlos Niebles. Revisiting the" video" in video-language understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2917–2927, 2022.

[50] Yuan Yao, Chang Liu, Dezhao Luo, Yu Zhou, and Qixiang Ye. Video playback rate perception for self-supervisedspatio-temporal representation learning, 2020.

[51] Ruyang Liu, Chen Li, Haoran Tang, Yixiao Ge, Ying Shan, and Ge Li. St-llm: Large language models are effective temporal learners. In *European Conference on Computer Vision*, pages 1–18. Springer, 2024.

[52] Lei Li, Yuanxin Liu, Linli Yao, Peiyuan Zhang, Chenxin An, Lean Wang, Xu Sun, Lingpeng Kong, and Qi Liu. Temporal reasoning transfer from text to video. *arXiv preprint arXiv:2410.06166*, 2024.

[53] Published at the workshop on und.

[54] Timotej Knez and Slavko Žitnik. Multimodal learning for temporal relation extraction in clinical texts. *Journal of the American Medical Informatics Association*, 31(6):1380–1387, 2024.

[55] Eiki Murata and Daisuke Kawahara. Time-aware comet: a commonsense knowledge model with temporal knowledge. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16162–16174, 2024.

[56] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and F Prabhat. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019.

[57] C an l arge l anguage m odels be.

[58] Te-Lin Wu, Alex Spangher, Pegah Alipoormolabashi, Marjorie Freedman, Ralph Weischedel, and Nanyun Peng. Understanding multimodal procedural knowledge by sequencing multimodal instructional manuals. *arXiv preprint arXiv:2110.08486*, 2021.

[59] Ashutosh Bajpai, Aaryan Goyal, Atif Anwer, and Tanmoy Chakraborty. Temporally consistent factuality probing for large language models. *arXiv preprint arXiv:2409.14065*, 2024.

[60] Chenhan Yuan, Qianqian Xie, Jimin Huang, and Sophia Ananiadou. Back to the future: Towards explainable temporal reasoning with large language models. In *Proceedings of the ACM on Web Conference 2024*, pages 1963–1974, 2024.

[61] Bhuwan Dhingra, Jeremy R. Cole, Julian Martin Eisenschlos, Daniel Gillick, Jacob Eisenstein, and William W. Cohen. Time-aware language models as temporal knowledge bases, 2022.

[62] Yuqing Wang and Yun Zhao. Tram: Benchmarking temporal reasoning for large language models. *arXiv preprint arXiv:2310.00835*, 2023.

[63] Zheng Chu, Jingchang Chen, Qianglong Chen, Weijiang Yu, Haotian Wang, Ming Liu, and Bing Qin. Timebench: A comprehensive evaluation of temporal reasoning abilities in large language models. *arXiv preprint arXiv:2311.17667*, 2023.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.