A Survey on Large Language Model Integration with Databases for Enhanced Data Management and Survey Analysis

www.surveyx.cn

Abstract

The integration of large language models (LLMs) with databases marks a significant advancement in artificial intelligence, enhancing data management and processing capabilities across various sectors. This survey explores the transformative potential of LLM-database integration, highlighting its impact on market sentiment analysis, healthcare, and decision-making processes. By addressing challenges such as biases, hallucinations, and computational demands, the integration aims to improve the reliability and efficiency of AI systems. Key methodologies include dynamic human-like memory architectures and interdisciplinary frameworks, which optimize LLM performance for domain-specific tasks. Case studies demonstrate LLMs' ability to automate complex processes, such as technology acceptance analysis and robotic process automation, showcasing their versatility and scalability. Furthermore, advancements in natural language processing techniques, such as sentiment analysis and topic modeling, facilitate the extraction of nuanced insights from survey data, driving innovation in qualitative research. Despite technical limitations and data privacy concerns, the integration of LLMs with databases presents substantial opportunities for enhanced data storage, retrieval, and analysis. Future research directions focus on refining evaluation metrics, addressing biases, and expanding multimodal capabilities to ensure equitable and efficient AI applications. Overall, this integration redefines industry standards, paving the way for sophisticated AI-driven solutions and improved user experiences across multiple domains.

1 Introduction

Artificial intelligence (AI) has witnessed remarkable advancements in recent years, significantly reshaping various sectors, including healthcare, finance, and robotics. One of the most promising developments in this field is the emergence of large language models (LLMs), which have demonstrated exceptional capabilities in natural language understanding and generation. The integration of LLMs with databases represents a pivotal evolution in AI, enabling enhanced data processing, management, and analysis. This review paper aims to explore the context, relevance, and significance of LLM-database integration, alongside the structure of the survey that will guide the reader through the multifaceted implications of this integration. By examining the current literature, this paper seeks to illuminate the transformative potential of LLMs in various applications, providing insights into future research directions and innovations.

1.1 Context and Relevance

The integration of large language models (LLMs) with databases is pivotal in advancing artificial intelligence (AI) and data processing, addressing complex challenges in managing vast amounts of data. This integration enhances capabilities in market sentiment analysis on social media, overcoming

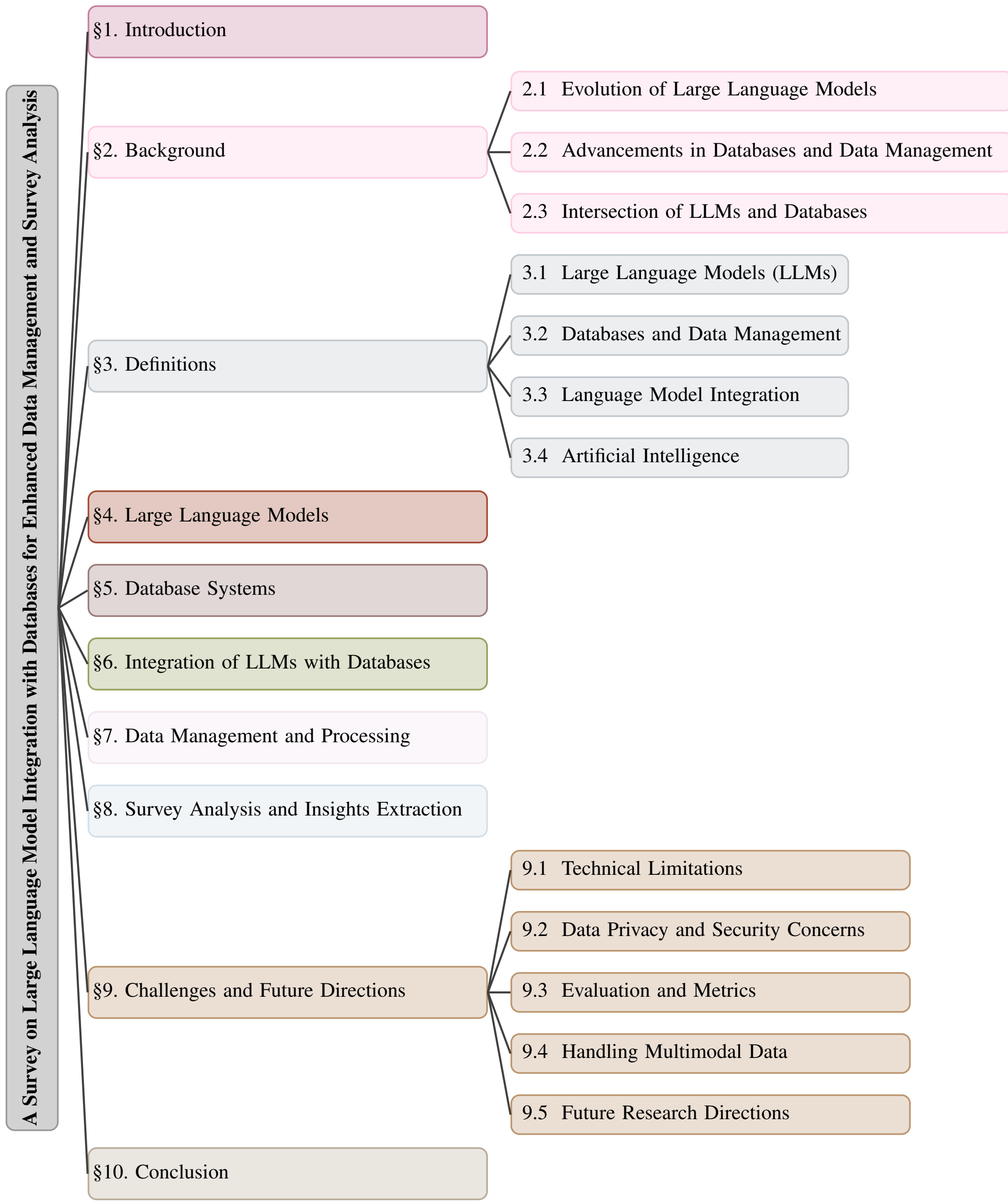


Figure 1: chapter structure

difficulties posed by financial jargon and the scarcity of high-quality labeled data [1]. Furthermore, the synergy between LLMs and optimization algorithms is crucial for decision-making processes in dynamic environments, highlighting their transformative potential [2].

LLMs like GPT-3 have revolutionized qualitative analysis, providing tools for deeper insights into complex datasets [3]. However, the need to address biases in LLM-generated content, such as demographic representation discrepancies, underscores the importance of integrating LLMs with databases to ensure fairness and accuracy [4]. Despite advances in LLM chatbots, challenges like hallucination, where models produce misleading information, necessitate robust integration frameworks to enhance reliability [5].

The study of working memory limitations in human cognition parallels the development of human-like memory architectures in LLM-based dialogue agents, emphasizing the relevance of database integration for improved cognitive abilities. In medicine, LLMs offer significant benefits, yet their successful implementation requires addressing challenges inherent to their integration with databases

AI-generated, for reference only.

[6]. Moreover, benchmarks like nanoLM are instrumental in facilitating model comparisons, enabling researchers to explore diverse designs and validate ideas efficiently [7].

The field of Explainable Artificial Intelligence (XAI) often caters to technically proficient users, presenting challenges in communicating complex XAI methods to non-experts [8]. Thus, integrating LLMs with databases is essential not only for enhancing data management and processing capabilities but also for ensuring accessibility and understanding across various user demographics. This integration is crucial for overcoming current inefficiencies and paving the way for innovative applications in AI, redefining industry standards, and enhancing the extraction of insights from complex datasets.

In conclusion, the integration of LLMs with databases is not merely a technological advancement; it is a necessary evolution that addresses the multifaceted challenges faced in data processing and management. By enhancing the capabilities of AI systems, this integration lays the groundwork for future innovations that will further improve the efficiency and accuracy of data-driven applications. The subsequent section will delve into the significance of this integration within the broader context of AI and data processing.

1.2 Significance in AI and Data Processing

The integration of large language models (LLMs) with databases marks a substantial advancement in the fields of artificial intelligence (AI) and data processing, offering transformative capabilities across diverse sectors. This synergy enables the creation of sophisticated evaluation frameworks such as ViLLM-Eval, which assess the knowledge and reasoning abilities of foundational models within specific cultural contexts, fostering model comparison and development [9]. In the robotics domain, LLMs enhance robots' ability to comprehend language instructions and respond to visual inputs, as exemplified by systems like the embodied robotic waiter [10].

In healthcare, the integration of LLMs significantly improves diagnostic accuracy and clinical decision-making, underscoring its transformative potential in medical applications [6]. Intelligent throat systems further demonstrate this potential by utilizing LLM processing to enhance patient communication [11]. The LABE benchmark highlights the importance of evaluating biases in AI-generated text, crucial for ensuring fairness and accuracy in AI applications [4].

Conversational AI benefits from LLM integration by addressing challenges related to factuality, conversationality, and latency in knowledge-grounded systems. Computational challenges are mitigated, and performance is enhanced in complex environments through the application of LLMs, which improve understanding and adaptability in various tasks. Understanding the working memory capabilities of LLMs like ChatGPT is vital for leveraging their human-like cognitive processes [12]. Furthermore, custom LLMs that generate audience-specific summaries of XAI methods make explainable AI more accessible, illustrating the transformative impact of LLM integration [8].

The integration of Large Language Models (LLMs) with databases has led to significant advancements that are reshaping the AI and data processing sectors. Notably, recent research has demonstrated the potential of fine-tuned LLMs to automate Systematic Literature Reviews (SLRs), enhancing academic research methodologies by improving the efficiency and accuracy of knowledge synthesis while maintaining high fidelity in factual accuracy. Additionally, a comprehensive survey of LLM inference serving systems has identified key innovations that optimize performance and scalability in real-world applications, further driving innovation and redefining traditional methodologies. Collectively, these developments not only streamline labor-intensive research processes but also enhance user experience and interaction dynamics across diverse applications, advocating for updated reporting guidelines to incorporate AI-driven methodologies for improved transparency and reliability in scholarly work [13, 14].

The significance of integrating LLMs with databases extends beyond mere technical improvements; it encapsulates a fundamental shift in how AI systems interact with data and users. This transformation holds the potential to redefine industry standards and enhance the overall quality of AI-driven solutions. The next section will outline the structure of this survey, detailing how each component contributes to a comprehensive understanding of LLM-database integration.

AI-generated, for reference only.

1.3 Structure of the Survey

The survey is organized into ten comprehensive sections, each contributing to a nuanced understanding of the integration of large language models (LLMs) with databases for enhanced data management and survey analysis. The introduction sets the stage by elucidating the context and relevance of LLM-database integration, highlighting its significance in AI and data processing. Following this, the background section offers a historical overview, tracing the evolution of LLMs and advancements in database systems, culminating in their intersection within artificial intelligence.

Subsequent sections provide precise definitions of key concepts, such as LLMs, databases, and artificial intelligence, establishing a foundational understanding for readers. The detailed exploration of LLM architecture and capabilities underscores their applicability in data management, while the examination of modern database systems highlights structural and functional aspects critical to integration efforts.

The survey focuses on advanced integration methodologies and technologies, specifically highlighting the application of fine-tuned Large Language Models (LLMs) for automating Systematic Literature Reviews (SLRs) and enhancing academic research methodologies. It includes real-world applications and case studies that demonstrate the practical implementation of these methodologies, such as the IIER framework for improved question-answering tasks through inter-chunk interactions, and the INformationINtensive (I N 2) training approach that addresses the challenges of long-context utilization in LLMs. These examples illustrate the significant benefits of integrating AI-driven processes into scholarly research, thereby improving efficiency, accuracy, and methodological transparency in literature reviews and other research tasks [15, 16, 14]. This is followed by an in-depth discussion on data management and processing improvements enabled by LLM integration, focusing on enhanced data processing, storage, retrieval, and automated analysis.

The survey analysis section delves into the role of LLMs in enriching survey analysis through natural language processing, detailing techniques for insight extraction and identifying challenges faced. Finally, the challenges and future directions section addresses technical limitations, data privacy concerns, evaluation metrics, and multimodal data handling, suggesting avenues for future research and innovation.

Each section is meticulously crafted to contribute to the overarching narrative, offering a holistic view of the transformative potential of integrating LLMs with databases, and paving the way for advancements in AI-driven data management and survey analysis.The following sections are organized as shown in Figure 1.

2 Background

2.1 Evolution of Large Language Models

The evolution of large language models (LLMs) highlights significant advancements in AI, characterized by innovations in architecture, training methods, and applications. Initial LLMs faced computational hurdles in managing large vocabularies and complex matrix operations, prompting the creation of energy-efficient models like SpikeLLM, which use spiking mechanisms to boost performance without extra energy demands [17, 18]. Autoregressive sampling techniques have been pivotal in achieving state-of-the-art results in natural language processing, enhancing the coherence of generated text [19]. Speculative decoding methods further refine these processes, addressing computational challenges inherent in large parameter models [20]. StreamBench benchmarks emphasize continuous improvement and adaptability in real-world applications [21].

LLM architectures have diversified into encoder-only, decoder-only, and encoder-decoder models, each tailored for specific tasks [22]. Advances in embedding techniques have enabled the transition to models capable of processing diverse data types, including time-series data [23]. Domain-specific pre-training, as seen in Traditional Chinese Medicine, highlights LLMs' adaptability to specialized fields [24]. Despite these advances, issues like generating inaccuracies persist, underscoring the importance of optimizing hyperparameters for improved model performance [25, 26]. Current alignment methods, reliant on human-annotated data, struggle with diverse online user preferences, indicating a need for more flexible approaches [27].

AI-generated, for reference only.

Integrating LLMs with Knowledge Graphs enhances reasoning and reduces hallucinations—instances of plausible yet incorrect information [28]. LLMs streamline ontology construction, traditionally a time-consuming and costly manual process [29]. Transitioning from BF16 to FP8 representations aims to improve training stability in reduced-precision settings, contributing to efficient model training [30].

LLMs' applicability extends to complex domains, such as multimodal health data integration, generating personalized health insights [31]. In psychology, LLMs enhance social media analytics and community monitoring, broadening their relevance in understanding human behavior [32]. However, a lack of empirical understanding of LLMs' impact on business model innovation remains a challenge [33].

A taxonomy of user intents for LLM interactions categorizes them into groups like 'Ask for Advice' and 'Information Retrieval', highlighting diverse applications [34]. Data quality and dataset practitioners' roles are crucial for LLM development, emphasizing high-quality datasets' importance [35]. The dynamic evolution of LLMs is vital for addressing knowledge processing complexities and enhancing capabilities in applications like economic beliefs and multimodal science question answering [36, 37]. Addressing user interaction challenges and balancing context length, accuracy, and performance remain ongoing tasks [38, 39].

2.2 Advancements in Databases and Data Management

Significant transformations in databases and data management are driven by increasing data complexity and volume. Traditional DBMS, like SQL and NoSQL, manage structured data effectively but struggle with complex queries, especially ESG data, prompting the development of graph databases for enhanced query capabilities [40]. IoT technologies further complicate data management due to the vast and diverse data they generate, leading to novel strategies for effective data integration and analysis [41]. Organizations increasingly require robust solutions for real-time data streams and actionable insights.

DBMS configuration tuning is challenging due to the configuration space's high dimensionality, necessitating automated solutions using machine learning to optimize performance [42]. These advancements enhance efficiency and reduce the burden on administrators. Integrating spiking neural networks into database systems, as seen in SpikeLLM, offers potential improvements in energy efficiency during large-scale data processing [18].

Dataset management has evolved, with initiatives like compiling TLDRs highlighting the importance of efficient data summarization and retrieval in research settings [43]. These developments underscore the ongoing evolution of database systems, essential for addressing modern data challenges and ensuring efficient data processing. As data landscapes evolve, innovative solutions remain critical.

2.3 Intersection of LLMs and Databases

The integration of LLMs and databases in AI enhances data management and processing. LLMs, such as GPT-3.5, automate metadata annotation, streamlining information organization and retrieval [44]. In Market Sentiment Analysis, LLMs label sentiment in Reddit posts, improving the speed and accuracy of financial market insights [1]. This integration addresses challenges of factual accuracy and hallucination, crucial for reliable AI systems [5].

Advanced algorithms optimize LLM performance, enhancing data management practices [2]. Prompt engineering techniques, like Zero-Shot and Chain-of-Thought, refine LLM-database interactions, improving information retrieval quality [45]. Challenges like position bias in LLMs highlight the need for robust integration strategies [46]. LLMs' cognitive capabilities, particularly working memory, are crucial for handling complex data interactions [12].

The nanoLM benchmark emphasizes advancements in model design and efficient data management, illustrating LLM-database convergence [7]. This integration enhances AI systems' capabilities, paving the way for sophisticated applications across domains, driving innovation, and improving data management efficiency. As the field evolves, LLM-database collaboration is likely to yield new methodologies and applications that enhance data utilization and accessibility.

AI-generated, for reference only.

3 Definitions

In examining the expansive domain of artificial intelligence, defining key concepts is critical to understanding its applications and impacts. This section focuses on Large Language Models (LLMs), central to the convergence of AI and data management. LLMs not only advance natural language processing but also demonstrate AI's transformative potential across multiple fields. As illustrated in Figure 2, the hierarchical structure of key concepts at the intersection of Artificial Intelligence and Data Management is depicted, emphasizing LLMs, their applications, and their integration with databases. This figure highlights the capabilities, challenges, and innovations in utilizing LLMs for data processing, user interaction, and complex analytics, as well as the synergy between AI and database technologies for enhanced data management. We will explore the characteristics, capabilities, and significance of LLMs in enhancing data-driven applications.

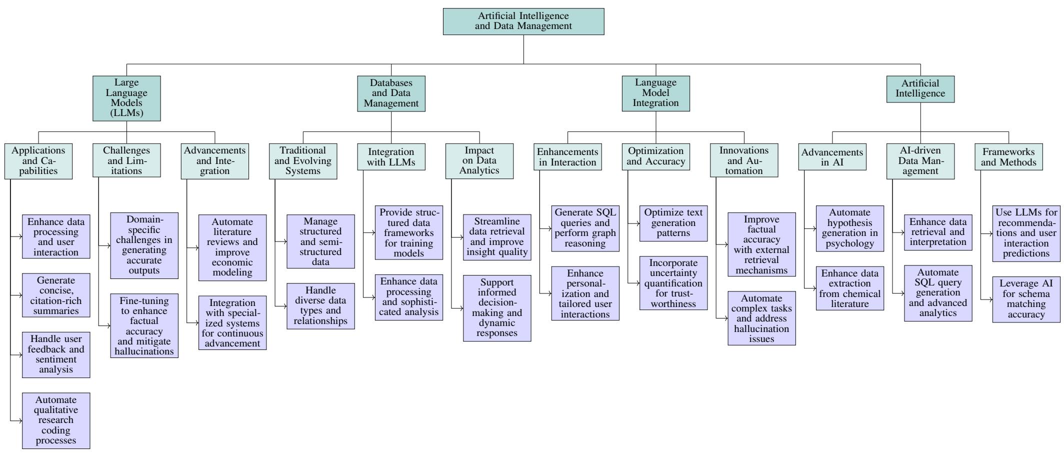


Figure 2: This figure illustrates the hierarchical structure of key concepts in the intersection of Artificial Intelligence and Data Management, focusing on Large Language Models (LLMs), their applications, integration with databases, and advancements in AI-driven methodologies. It highlights the capabilities, challenges, and innovations in utilizing LLMs for data processing, user interaction, and complex analytics, as well as the synergy between AI and database technologies for enhanced data management.

3.1 Large Language Models (LLMs)

Large Language Models (LLMs) are advanced AI systems designed to process and generate text akin to human language, enhancing data processing and user interaction in diverse domains [47]. They are essential in frameworks like RAGS4EIC for generating concise, citation-rich summaries, improving access to complex scientific data [48]. LLMs excel in handling user feedback and sentiment analysis, as seen in financial markets where they generate sentiment labels for market analysis [1].

As illustrated in Figure 3, the diverse applications of LLMs across various domains emphasize their critical roles in data processing, sentiment analysis, and healthcare. This categorization highlights LLMs' capabilities in enhancing data summarization, market analysis, and addressing domain-specific challenges, demonstrating their impact on improving efficiency and accuracy in these fields.

LLMs signify a paradigm shift in machine language processing and context understanding. In healthcare, benchmarks like the Personal Health Insights Agent (PHIA) assess LLMs' ability to address health queries, demonstrating their role in generating personalized insights [49]. Efficient metadata annotation further showcases their cost-effectiveness in data management [44]. However, domain-specific challenges, such as in transportation, highlight limitations in generating accurate outputs [50].

LLMs also enhance qualitative research through tools like QualiGPT, which automate coding processes, improving efficiency in qualitative analysis [3]. Cognitive benchmarks, such as n-back tasks, highlight LLMs' advanced capabilities in simulating human cognition [12]. Continuous advancements in training and integration with specialized systems ensure LLMs' relevance across fields.

AI-generated, for reference only.

Recent studies emphasize LLMs' role in automating literature reviews and improving economic modeling, while fine-tuning enhances factual accuracy and mitigates hallucination [13, 14]. Surveys on LLM inference serving highlight system-level innovations improving real-world application performance, underscoring LLMs' growing impact across sectors.

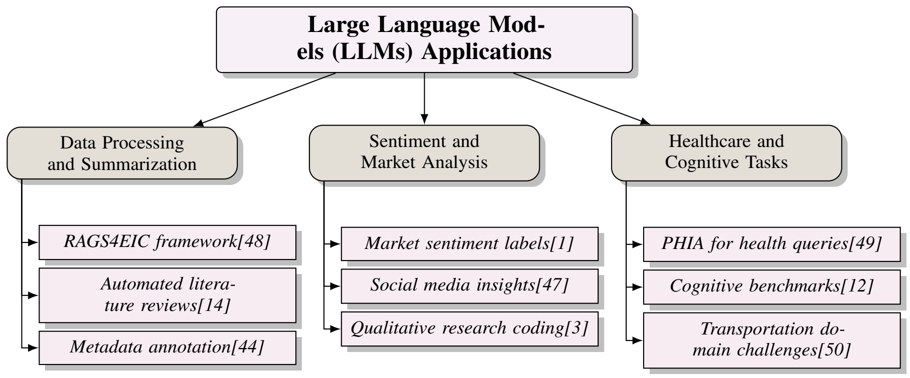


Figure 3: This figure illustrates the diverse applications of Large Language Models (LLMs) across various domains, highlighting their roles in data processing, sentiment analysis, and healthcare. The categorization emphasizes LLMs' capabilities in enhancing data summarization, market analysis, and addressing domain-specific challenges, demonstrating their impact on improving efficiency and accuracy in these fields.

3.2 Databases and Data Management

Databases, structured collections of data, are crucial for efficient data storage, retrieval, and management across domains. Traditional DBMS like SQL and NoSQL manage structured and semi-structured data, ensuring data integrity and security [40]. However, the rise of IoT and complex data necessitates evolving systems to handle diverse data types and relationships [41].

Data management involves practices and tools ensuring data quality and accessibility, critical for informed decision-making [42]. Integrating AI into data management enhances complex dataset analysis, paving the way for intelligent data handling solutions [18].

The relevance of databases to LLM integration lies in providing structured data frameworks for training and deploying models. This synergy enhances data processing, enabling sophisticated analysis of natural language data [44]. As LLMs evolve, databases play a pivotal role in supporting AI application scalability and efficiency.

The interplay between data management and LLMs fosters advanced data analytics, streamlining data retrieval and improving insight quality. This integration supports informed decision-making and dynamic responses to data trends. As real-time data processing demand grows, database technology evolution will be critical for next-generation AI applications.

3.3 Language Model Integration

Integrating language models with databases enhances LLMs' ability to interact with structured data, enabling complex tasks such as generating SQL queries and performing graph reasoning. The Persona-DB framework, for instance, enhances LLM personalization by structuring databases for tailored user interactions [51].

LLMs optimize database interactions by selecting rewrite rules for SQL queries, ensuring executability and equivalence [52]. Frameworks like Graph-ToolFormer empower LLMs for graph reasoning through prompts integrating external APIs, enhancing graph data processing [53]. PatternGPT further exemplifies LLMs' role in optimizing text generation patterns [54].

GenRec illustrates LLM use in generating recommendations by formatting user data as prompts, enhancing data retrieval accuracy [55]. The UAG method incorporates uncertainty quantification into KG-LLM systems, improving knowledge graph reasoning trustworthiness [28].

AI-generated, for reference only.

Advancements in LLM-database integration, such as RETA-LLM, improve factual accuracy and user experience by combining generative capabilities with external retrieval mechanisms. These innovations automate complex tasks and address hallucination issues, driving domain innovation and setting new research transparency standards [13, 56, 14].

3.4 Artificial Intelligence

Artificial Intelligence (AI) encompasses creating advanced algorithms and systems for tasks like learning, reasoning, and language processing. Recent advancements, particularly in integrating LLMs with knowledge frameworks, automate hypothesis generation in psychology and enhance data extraction from chemical literature [57, 58]. AI technologies improve data analysis efficiency across disciplines, revolutionizing methodologies and fostering discoveries [59, 14, 60].

LLM integration with databases exemplifies AI's advancement in data management. LLMs enhance data retrieval and interpretation, automating tasks like SQL query generation and advanced analytics. This synergy is evident in frameworks like Persona-DB for personalized user interactions and Graph-ToolFormer for graph reasoning [53]. AI optimizes database interactions, as seen in LLMR2's SQL query rewrite rule selection [52].

AI-driven methods, like GenRec, use LLMs for recommendations, predicting user interactions [55]. The UAG method enhances knowledge graph reasoning by incorporating uncertainty quantification [28]. AI's role in LLM-database integration improves data processing efficiency and drives domain innovation.

Frameworks like PromptMatcher leverage AI for schema matching accuracy through crafted prompts [61]. As AI evolves, its role in LLM-database integration will be pivotal in advancing data management and unlocking new AI-driven insights.

4 Large Language Models

The study of large language models (LLMs) involves a detailed analysis of their architectures, capabilities, and applications. Understanding these architectural frameworks is crucial as they underpin the diverse functionalities of LLMs. The following subsection will explore the specific architectural and design considerations that define LLMs, emphasizing innovative frameworks that enhance their performance across different domains.

4.1 Architecture and Design

The architecture of large language models (LLMs) is built on three main frameworks: auto-encoding models, which excel at capturing contextual information; auto-regressive models, recognized for their sequential generation capabilities; and sequence-to-sequence models, effective for tasks like translation. These frameworks offer distinct advantages, such as improved contextual understanding and generation accuracy, while facing challenges like hallucinations and integrating external knowledge [15, 54, 14]. LLMs are also categorized by their roles in reinforcement learning, serving as information processors, reward designers, decision-makers, and generators, highlighting their adaptability in complex environments [62].

In healthcare, LLMs are adapted to enhance communication and information processing, impacting patient care, medical research, and education [63]. TransGPT illustrates LLMs' application-specific architecture, processing both textual and visual data in transportation, demonstrating their adaptability to multimodal tasks [50]. Architectural frameworks like KG-enhanced LLMs and LLM-augmented KGs improve reasoning and data processing by leveraging the strengths of both LLMs and knowledge graphs [64].

LLMs are also utilized as search operators in optimization problems, generating algorithms through problem domain understanding, showcasing their functional diversity [2]. Recent advancements, such as fine-tuning and innovative training techniques like INformationINtensive (I N 2) training, improve LLM capabilities, addressing issues like the "lost-in-the-middle" challenge and automating complex tasks like systematic literature reviews [13, 16, 14].

AI-generated, for reference only.

4.2 Capabilities in Natural Language Processing

LLMs exhibit advanced capabilities in understanding and generating natural language, significantly enhancing applications across various domains. They improve memory recall and response generation in dialogue agents, enhancing language comprehension and production [65]. Frameworks like MPlug separate visual and language processing, allowing effective collaboration and improved task alignment [66].

As illustrated in Figure 4, the capabilities of large language models (LLMs) in natural language processing are multifaceted, encompassing key areas such as language understanding, task alignment, and tool learning integration. This figure provides a structured overview of how LLMs enhance language comprehension, facilitate multi-modal task alignment, and incorporate tool learning to improve performance and reliability.

LLMs handle complex language patterns, enabling tasks like sentiment analysis and sarcasm detection, crucial for interpreting public opinions from social media. Techniques like chain-of-thought prompting and frameworks such as Structure Guided Prompt and Venn Diagram Prompting enhance LLMs' performance on complex reasoning tasks, maintaining coherent reasoning chains and producing accurate, context-aware outputs [46, 67].

Tool learning integration significantly enhances LLM performance by improving accuracy, efficiency, and user interaction. Augmenting LLMs with retrieval systems mitigates hallucinations by providing factual external information, increasing reliability. Fine-tuning methodologies streamline processes like systematic literature reviews, demonstrating adaptability to academic research demands [56, 16, 68, 13, 14]. Streaming methods consistently improve LLM performance, indicating robust capabilities in natural language processing tasks.

These advancements highlight LLMs' transformative impact in natural language processing, enabling sophisticated understanding, generation, and interaction. LLMs automate literature review generation, enhancing research efficiency across disciplines. Fine-tuned LLMs streamline systematic literature review processes, ensuring high factual accuracy and mitigating hallucination issues. Retrieval-augmented techniques allow LLMs to produce reliable outputs by referencing external information, expanding applicability in academia and industry [13, 43, 56, 14].

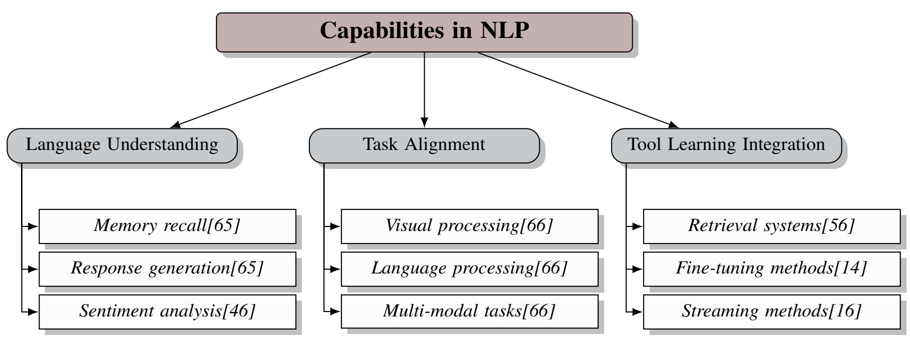


Figure 4: This figure illustrates the capabilities of large language models (LLMs) in natural language processing, highlighting key areas such as language understanding, task alignment, and tool learning integration. The figure provides a structured overview of how LLMs enhance language comprehension, facilitate multi-modal task alignment, and incorporate tool learning to improve performance and reliability.

4.3 Applications in Data Management

LLMs have transformed data management by offering sophisticated capabilities for handling complex datasets. The RAGS4EIC method uses LLMs to provide accurate and trustworthy summaries, enhancing data management efficiency and reliability [48]. The UAG method demonstrates LLMs' ability to deliver reliable, uncertainty-aware answers through structured reasoning, improving data-driven decision-making quality [28].

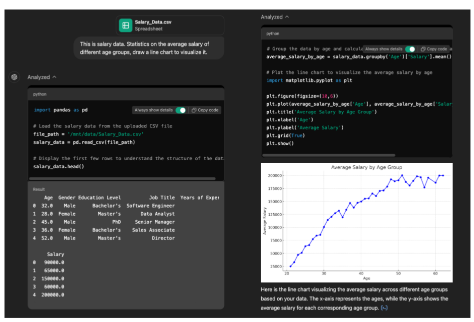
LLMs reduce reliance on human experts in ontology construction, dynamically generating relevant concepts and increasing efficiency [29]. Automated annotation of subject metadata illustrates

AI-generated, for reference only.

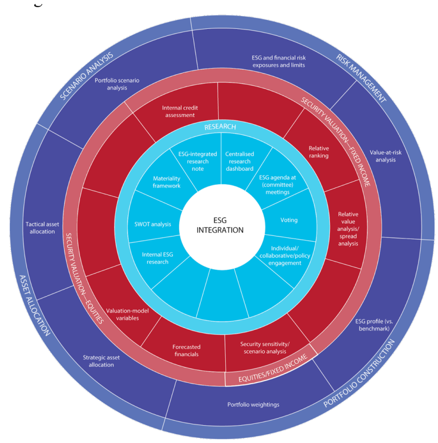
LLMs' role in streamlining data management, enhancing accuracy and speed [44]. In qualitative research, tools like QualiGPT automate thematic analysis, expediting workflows and improving insight accessibility [3].

The intelligent throat (IT) system captures throat muscle vibrations and carotid pulse signals, integrating with LLMs for intelligent speech synthesis, highlighting LLMs' potential in specialized domains [11]. Empirical examinations of LLM impacts on business models provide insights into practical applications, demonstrating transformative potential [33]. Models trained on balanced datasets outperform those on imbalanced datasets, emphasizing data balance importance in optimizing performance [69].

These applications illustrate LLMs' transformative impact on data management, driving innovation and improving accuracy and efficiency across sectors. LLM integration in data management improves reliability and quality, automates complex solutions, and enhances software requirements evaluation and refinement, leading to efficient outcomes in software development and research automation [70, 43, 35].



(a) Salary Data Analysis[71]



(b) ESG Integration Framework[40]

Figure 5: Examples of Applications in Data Management

As shown in Figure 5, LLMs are increasingly used in data management to streamline and enhance various applications. The "Salary Data Analysis" example demonstrates LLMs' role in analyzing and visualizing salary data, facilitating complex data analysis tasks efficiently. The "ESG Integration Framework" example highlights LLMs' role in structuring and managing ESG-related data, supporting comprehensive analysis and decision-making. These examples underscore LLMs' transformative potential in optimizing data management practices across sectors [71, 40].

5 Database Systems

The evolution of database systems, particularly with the integration of advanced technologies like Large Language Models (LLMs), has significantly enhanced their architecture and functionality. These advancements facilitate efficient data management and enable the extraction of insights from unstructured and historical data, addressing both current challenges and future opportunities in database administration and information retrieval [72, 43, 67, 73]. This section examines the characteristics and operational principles of modern database systems, focusing on their structure and functionality.

5.1 Structure and Functionality of Modern Database Systems

Modern database systems excel in managing vast data volumes, driven by design and operational advancements. The need to handle diverse data types and complex relationships, especially with the rise of IoT and AI, has led to the adoption of NoSQL databases alongside traditional relational systems, offering flexibility in managing semi-structured and unstructured data [40]. This shift reflects a trend towards optimizing performance across various data types and use cases.

AI-generated, for reference only.

Graph databases represent a significant advancement, offering enhanced capabilities for managing interconnected data structures, executing complex queries, and modeling data as nodes and edges, which is particularly effective in social networks and recommendation systems [40]. This facilitates efficient data access and deeper analytical insights, essential in today's data-driven environment.

The proliferation of IoT has driven the development of novel architectures like edge computing and distributed databases, which decentralize data processing to reduce latency and improve real-time access [41]. These systems efficiently handle dynamic data streams from IoT devices, enabling responsive and adaptive data management strategies.

Operationally, modern databases employ automated tuning solutions, leveraging machine learning and AI to optimize performance and reduce manual intervention [42]. Integrating spiking neural networks, as seen in SpikeLLM, offers potential improvements in energy efficiency and computational performance, crucial in large-scale data processing environments [18].

The continuous evolution of database systems, through innovative architectures and operational strategies, enables effective management of complex data environments. This evolution underscores the role of LLMs in automating maintenance tasks, enhancing response times, and reducing the burden on human DBAs [72, 74, 75, 35].

5.2 Challenges in LLM and Database Integration

Integrating LLMs with databases involves challenges related to computational demands, methodological limitations, and domain-specific intricacies. The substantial computational resources required for LLM training necessitate effective optimization strategies [2]. Efficient memory management remains a challenge, impacting seamless LLM-database integration [65].

As illustrated in Figure 6, the primary challenges in integrating LLMs with databases can be categorized into three main areas: computational demands, domain-specific issues, and benchmarking challenges. This figure highlights the need for optimization strategies and effective memory management, as well as the complexities associated with medical language and qualitative coding, which further complicate integration efforts.

In healthcare, the complexity of medical language poses challenges in ensuring unbiased models and maintaining data privacy [6]. Current qualitative analysis software often lacks automatic coding capabilities, highlighting a gap in tools supporting LLM integration for qualitative research [3].

Benchmarking LLMs faces challenges due to reliance on string-matching techniques, which inadequately capture language complexity [4]. Existing benchmarks often require costly, resource-intensive direct training on large models, necessitating efficient evaluation frameworks [7].

Addressing these challenges requires innovative solutions to enhance computational efficiency, refine methodological frameworks, and cater to specific domain requirements. This includes fine-tuned LLMs for automating Systematic Literature Reviews, improving factual accuracy, and retrieval-augmented approaches to mitigate hallucination issues [56, 35, 70, 13, 14].

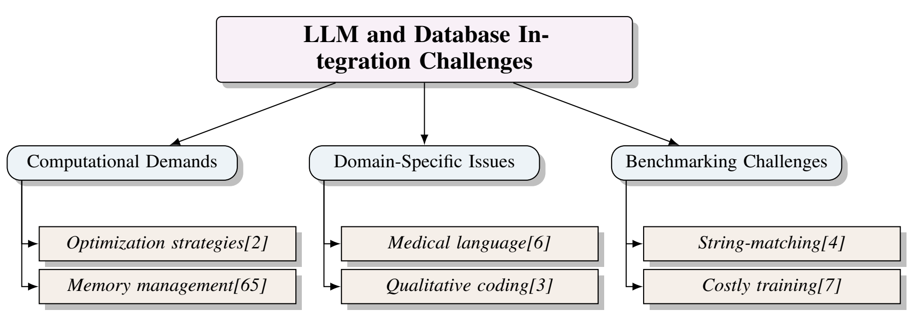


Figure 6: This figure illustrates the primary challenges in integrating Large Language Models (LLMs) with databases, categorized into computational demands, domain-specific issues, and benchmarking challenges. It highlights optimization strategies and memory management, the complexity of medical language and qualitative coding, and the inadequacies of current benchmarking methods.

AI-generated, for reference only.

5.3 Opportunities for Enhanced Data Management

Integrating LLMs with databases offers significant opportunities for enhancing data management. LLMs improve data retrieval and analysis, automating complex query generation and streamlining data access [52]. This reduces manual intervention, allowing data professionals to focus on higher-level analytical tasks.

As illustrated in Figure 7, the integration of LLMs not only enhances data retrieval but also significantly improves personalization in data-driven applications by adapting to user preferences. The Persona-DB framework exemplifies this improvement, demonstrating enhanced performance in cold-start scenarios through hierarchical organization of databases [51, 76].

Moreover, LLM integration with graph databases enhances data management by empowering LLMs to perform graph reasoning tasks, which are crucial for applications involving complex data structures [53]. The figure also emphasizes the importance of uncertainty quantification in LLM-driven systems, as demonstrated by the UAG method, which enhances reliability and trustworthiness [28].

Incorporating LLMs into schema matching processes, as seen in the PromptMatcher framework, improves data management accuracy and efficiency [61]. This streamlines data integration tasks, ensuring higher accuracy in data mapping and alignment.

Furthermore, LLM integration with database management systems revolutionizes data management by addressing challenges faced by traditional DBAs, particularly in managing numerous cloud database instances. LLMs like D-Bot provide timely diagnostics and optimization recommendations, thereby improving data management workflows [72, 13, 56, 35]. By leveraging LLM capabilities, organizations can optimize data management strategies, driving innovation and enhancing data-driven applications across sectors.

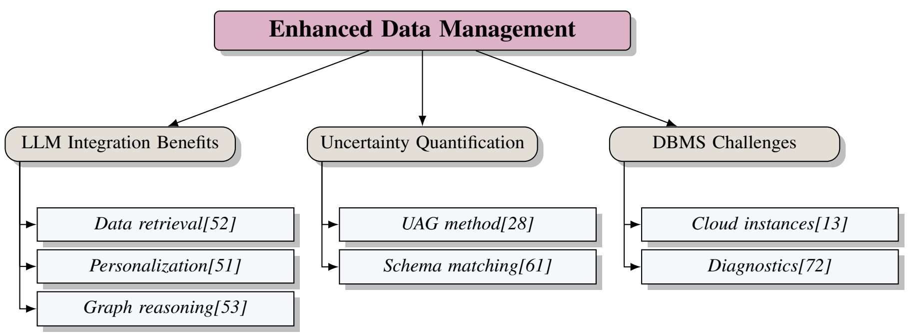


Figure 7: This figure illustrates the opportunities for enhanced data management through LLM integration, highlighting key benefits such as improved data retrieval, personalization, and graph reasoning capabilities. It also emphasizes the importance of uncertainty quantification in schema matching and the challenges faced by traditional database management systems, particularly in cloud environments.

6 Integration of LLMs with Databases

Category	Feature	Method
Methodologies for Integration	Cognitive Replication	DHMRC[65]
	Sequential Processing	MOWL[66]
Applications and Case Studies	LLM Utilization	LLMAS[77], SR[78], SGP[67], TRAD[79], UPAR[80]
	Data Management	UMLLM[81]
	User Interaction	Walert[82]

Table 1: This table provides a comprehensive summary of methodologies and applications related to the integration of large language models (LLMs) with databases. It categorizes various approaches and case studies, highlighting specific features and methods employed to enhance data processing, user interaction, and analytical performance across different domains. The listed references illustrate the transformative potential of LLMs in improving data management practices and facilitating advanced, automated solutions.

AI-generated, for reference only.

The integration of large language models (LLMs) with databases marks a significant advancement in data management and artificial intelligence, enhancing data processing efficiency and introducing innovative methodologies. Table 1 presents a detailed overview of methodologies and applications associated with the integration of large language models (LLMs) with databases, underscoring their significance in advancing data management and analytical processes. Additionally, Table 3 provides a comparative overview of three methodologies for integrating large language models (LLMs) with databases, emphasizing their integration processes, application domains, and performance enhancements. These methodologies leverage LLMs' strengths to improve interaction quality and analytical performance, particularly in natural language processing and spectral analysis, facilitating automated knowledge retrieval and enhancing the accuracy and traceability of responses. Techniques such as retrieval-augmented generation (RAG) and fine-tuning LLMs for specific tasks streamline labor-intensive processes like systematic literature reviews, addressing the challenges of handling vast academic literature [43, 14, 83]. This section explores the methodologies underpinning these advancements.

6.1 Methodologies for Integration

Integrating large language models (LLMs) with databases involves methodologies designed to enhance data processing and system performance across various applications. The Dynamic Human-like Memory Recall and Consolidation (DHMRC) approach integrates LLMs with memory processes to improve dialogue coherence, demonstrating LLMs' potential to mimic human memory dynamics and enhance interaction quality [65]. This approach emphasizes replicating human cognitive processes to improve user satisfaction and engagement.

In medical applications, interdisciplinary collaboration frameworks are crucial, focusing on dynamic model training to meet the specific needs of different medical disciplines [6]. These frameworks tailor LLMs to domain-specific requirements, enhancing their capability in managing complex medical data. Such adaptability is essential for improving diagnostic tools and supporting personalized medicine.

The MPlug method introduces a two-stage LLM integration process, initially aligning images and text using frozen components and subsequently fine-tuning with language-only and multi-modal instructions [66]. This method underscores LLMs' versatility in processing multimodal data, broadening their applicability in environments like education and training, where visual and textual information coexist.

The nanoLM benchmark provides a novel methodology for loss prediction, requiring a one-time hyperparameter search for smaller models, facilitating efficient integration strategies [7]. This benchmark highlights optimizing model training processes to enable seamless database integration, reducing computational burdens and promoting broader LLM adoption in resource-constrained settings.

These methodologies illustrate diverse strategies for integrating LLMs with databases, showcasing their transformative potential in driving innovation and improving data management practices across sectors. By leveraging LLMs' sophisticated capabilities, these approaches enable systems to manage intricate data structures and interactions, facilitating advanced, automated data-driven solutions. This is particularly evident in automating literature reviews, where retrieval-augmented generation and LLM fine-tuning significantly enhance the efficiency and accuracy of synthesizing research findings, transforming academic methodologies [43, 14].

6.2 Applications and Case Studies

The integration of large language models (LLMs) with databases has been implemented across various sectors, demonstrating their transformative impact on data management and analytical processes. In technology acceptance analysis, LLMs have replaced traditional survey methods, achieving consistency and accuracy in annotating user attitudes towards technology [77]. This application highlights LLMs' potential to refine data collection methodologies, enhancing survey analysis efficiency and minimizing human biases.

In information retrieval, the Self-Retrieval framework embeds the retrieval process within a single LLM, significantly improving performance over existing methods [78]. This approach demonstrates

AI-generated, for reference only.

Method Name	Application Areas	Methodological Innovations	Impact on Efficiency
LLMAS[77]	Technology Acceptance Research	Llm Annotation System	Improving Data Management
SR[78]	Information Retrieval	Self-Retrieval	Enhanced Productivity
TRAD[79]	Robotic Process Automation	Thought Retrieval	Enhanced Productivity
SGP[67]	Multi-step Reasoning	Structure Guided Prompt	Improved Reasoning Accuracy
UPAR[80]	Complex Reasoning	Upar Framework	Enhanced Accuracy
UMLLM[81]	Multi-modal Tasks	Task Tokens	Task Execution
Walert[82]	Computer Science Programs	Intent-Based And Retrieval-Augmented	Enhance User Interaction

Table 2: Overview of various methodologies integrating Large Language Models (LLMs) across different application areas, highlighting their methodological innovations and impact on efficiency. The table summarizes key frameworks such as LLMAS, SR, TRAD, SGP, UPAR, UMLLM, and Walert, illustrating their roles in enhancing data management, productivity, reasoning accuracy, and user interaction within their respective domains.

LLMs' capacity to streamline data accessibility and enhance user interaction in information-rich environments, from search engines to digital libraries.

The TRAD framework, applied in robotic process automation within a global business insurance company, exemplifies LLMs' utility in automating complex business processes [79]. This case study shows how LLMs optimize operational efficiency and support decision-making by automating routine tasks and providing intelligent insights, ultimately improving productivity and reducing costs.

In reasoning tasks, the Structure Guided Prompt framework enhances LLM performance in multi-step reasoning tasks by using structured prompts to improve cognitive processing [67]. This advancement emphasizes prompt engineering's role in augmenting LLM reasoning abilities, facilitating their application in complex analytical tasks.

The UPAR framework demonstrates significant improvements in datasets requiring intricate reasoning, showcasing LLMs' potential in tackling complex challenges [80]. By refining human-LLM interactions, UPAR fosters collaborative problem-solving, enhancing the quality of insights generated.

The UnifiedMLLM framework, validated through extensive experiments, demonstrates effectiveness, scalability, and generalization across tasks [81]. This framework exemplifies LLMs' versatility in handling diverse data types, underscoring their transformative impact on data management and analysis across domains.

Advancements such as VD Prompting enhance LLM performance in generating accurate answers from long-context documents [46]. Similarly, Walert employs Intent-Based and Retrieval-Augmented Generation methodologies to improve user interaction and response accuracy in conversational search applications [82]. These methodologies exemplify LLMs' innovative applications in enhancing user engagement and data processing efficiency.

These case studies and applications collectively demonstrate the transformative impact of LLM-database integration, driving innovation and improving data-driven processes across sectors. By harnessing LLMs' advanced capabilities, these integrations enhance data management efficiency, enabling sophisticated, automated, and data-driven solutions. Recent research highlights the automation of literature reviews through NLP techniques, including RAG with LLMs, significantly reducing manual effort in synthesizing academic literature. Fine-tuned LLMs improve systematic literature reviews' accuracy and reliability, addressing challenges like hallucination and ensuring transparency. Toolkits like RETA-LLM facilitate LLM customization for specific domains, enhancing interaction with information retrieval systems, underscoring LLMs' transformative potential in streamlining research methodologies across academic fields [56, 35, 13, 43, 14]. Table 2 provides a comprehensive summary of recent methodologies leveraging Large Language Models (LLMs) across diverse application areas, emphasizing their methodological advancements and resultant improvements in efficiency.

7 Data Management and Processing

Effective data management and processing are crucial in today's data-intensive environment, given the increasing data volume and complexity across various domains. Integrating advanced technologies, particularly retrieval-augmented large language models (LLMs), is becoming essential for organizations aiming to enhance data handling capabilities. These models improve response accuracy by

AI-generated, for reference only.

Feature	Dynamic Human-like Memory Recall and Consolidation (DHMRC)	Interdisciplinary Collaboration Frameworks	MPlug
Integration Process	Memory Integration	Dynamic Model Training	Two-stage Process
Application Domain	Dialogue Systems	Medical Applications	Multimodal Environments
Performance Enhancement	Improves Coherence	Tailors TO Disciplines	Broadens Applicability

Table 3: This table presents a comparative analysis of three distinct methodologies for integrating large language models (LLMs) with databases: Dynamic Human-like Memory Recall and Consolidation (DHMRC), Interdisciplinary Collaboration Frameworks, and MPlug. Each method is evaluated based on its integration process, application domain, and performance enhancement, highlighting their unique contributions to improving dialogue systems, medical applications, and multimodal environments.

using external information retrieval systems to mitigate hallucination and address specific in-domain queries that traditional LLMs may find challenging. Developments in LLM serving systems further facilitate efficient deployment in practical applications, underscoring their significance in modern data management practices [56, 13, 14, 35]. This integration streamlines workflows and introduces innovative methodologies for processing, storage, and retrieval, making it vital for researchers and practitioners to understand these advancements.

7.1 Enhanced Data Processing

Integrating large language models (LLMs) into data processing frameworks has markedly improved efficiency and accuracy across domains. This integration supports simultaneous sentiment and sarcasm detection, enhancing public opinion understanding and nuanced data interpretation [47]. The RELIC method exemplifies LLMs' transformative impact by enabling users to validate LLM-generated text, thus improving data management practices' reliability [84]. In traditional Chinese medicine (TCM), the TCMDA framework uses a tailored corpus for pre-training, significantly enhancing TCM task performance [24]. This highlights LLMs' potential to optimize workflows by adapting to domain-specific requirements, ensuring nuanced field-specific data is accurately captured and analyzed.

LLMs facilitate the creation of relevant datasets for evaluating retrieval-augmented generation (RAG) systems, enhancing data processing capabilities [75]. Automating dataset generation ensures systems are efficient and responsive to user needs, crucial in dynamic environments. These advancements illustrate LLM integration's transformative impact on data processing, enabling efficient, accurate handling of complex datasets. LLMs automate complex tasks like Systematic Literature Reviews (SLRs), maintaining high factual accuracy and addressing hallucination by integrating external sources. Organizations adopting these approaches can optimize workflows, ensure transparency, and adapt to increasing data volumes, setting new standards for research and application development [56, 35, 13, 43, 14].

7.2 Improved Data Storage and Retrieval

LLM integration with modern database systems has significantly advanced data storage and retrieval processes, enabling sophisticated data management solutions. LLMs automate complex query generation and optimization, streamlining retrieval processes and enhancing access efficiency [52]. This reduces manual query formulation reliance, expediting data access and minimizing error likelihood. Integrating LLMs with graph databases exemplifies improved storage solutions by managing complex data structures and relationships [53]. This is beneficial in domains requiring data interconnectivity, such as social networks, enhancing stored information utility.

The PromptMatcher framework demonstrates schema matching advancements, improving data integration task accuracy and efficiency [61]. Automating schema matching reduces resource needs, streamlining operations and improving data quality. Incorporating uncertainty quantification, as seen in the UAG method, enhances data retrieval reliability by providing uncertainty-aware answers [28]. This ensures accurate insights for informed decision-making. LLM integration offers transformative potential for data storage and retrieval, automating tasks like SLRs with high accuracy and efficiency. Robust data quality evaluations and innovative training methods ensure critical data points are not overlooked, setting new standards for transparency and reliability in research [16, 14, 35].

AI-generated, for reference only.

7.3 Automated Data Collection and Analysis

LLM integration into automated data collection and analysis processes represents a significant advancement in data-driven applications' efficiency and accuracy. LLMs automate data collection by generating complex queries, reducing manual data gathering time and effort [52]. This is beneficial in domains like healthcare and finance, where timely data collection is crucial. In qualitative research, LLMs like QualiGPT automate thematic analysis, expediting the research process and improving qualitative data analysis consistency [3]. LLMs enhance data analysis by providing uncertainty-aware answers through structured reasoning, as demonstrated by the UAG method [28]. This ensures reliable insights for strategic planning.

LLMs reduce reliance on human experts in ontology construction, dynamically generating relevant concepts and improving data categorization efficiency [29]. This is advantageous in rapidly evolving fields, ensuring data categorization systems remain relevant. LLMs' application in automated data collection and analysis exemplifies their transformative impact on data management practices, optimizing strategies for improved efficiency, accuracy, and reliability. This is particularly relevant across diverse sectors, as LLMs automate labor-intensive tasks like systematic literature reviews, ensuring high factual accuracy and addressing quality and source tracking challenges [13, 14, 35].

7.4 Resource Optimization and Management

LLM integration with databases enhances resource optimization and management, improving data-driven applications' efficiency and scalability. LLMs automate complex data processing tasks, reducing computational resource needs and enhancing system performance [2]. This is beneficial in large-scale data processing environments, optimizing resource utilization. The DHMRC framework exemplifies LLMs' resource management optimization by mimicking human cognitive processes, improving dialogue coherence without increasing computational demands [65]. This enhances interaction quality, fostering user engagement and satisfaction.

In medical applications, interdisciplinary frameworks emphasize dynamic model training for efficient resource allocation tailored to domain requirements [6]. Integrating LLMs ensures effective resource management, supporting complex medical data tasks. The MPlug method optimizes resources by separating visual and language processing, improving alignment and reducing computational overhead [66]. The nanoLM benchmark offers resource optimization strategies by detailing efficient model training processes, reducing training costs and resource consumption [7]. LLM integration with database systems enables automated, real-time diagnosis and optimization recommendations, addressing database management challenges. Systems like D-Bot utilize LLMs for continuous learning, providing maintenance insights. Retrieval-augmented LLMs enhance response accuracy, improving database management efficiency [72, 56]. LLMs enhance data management strategies' scalability and efficiency, supporting complex applications across sectors.

8 Survey Analysis and Insights Extraction

The integration of advanced methodologies in survey analysis is crucial for extracting meaningful insights from complex datasets. This section examines how natural language processing (NLP) techniques have transformed the interpretation of qualitative data, particularly through the use of large language models (LLMs). These models automate the analysis of open-ended survey responses, enhancing both the depth and efficiency of insights generated. The following subsection explores the role of NLP in modern survey analysis, emphasizing its importance for researchers and practitioners navigating large volumes of unstructured data.

8.1 Enhancement of Survey Analysis through NLP

Natural language processing (NLP) techniques have revolutionized survey analysis by automating the interpretation of open-ended responses, with large language models (LLMs) at the forefront of this innovation. LLMs automate the coding of qualitative data, allowing for a nuanced understanding of sentiments and themes with minimal manual intervention [3]. This automation streamlines analysis and mitigates human error, ensuring reliable insights.

AI-generated, for reference only.

Advanced NLP techniques like sentiment analysis and topic modeling extract key themes and patterns, providing insights into consumer behavior and preferences [47]. Sentiment analysis quantifies emotional responses, while topic modeling categorizes responses into coherent themes, simplifying interpretation. Chain-of-thought prompting further enhances LLMs' reasoning capabilities, enabling complex data processing and coherent insight generation [67]. These advancements highlight the need for sophisticated analytical tools in deriving actionable insights from qualitative data.

NLP also evaluates language agency, using LLMs to assess language's impact on user perceptions and interactions, informing effective communication strategies [4]. By analyzing survey response language nuances, researchers can tailor messaging to resonate with target audiences, enhancing engagement. This aligns with trends in personalized marketing and communication.

Frameworks like Persona-DB leverage LLMs to generate personalized insights, tailoring analysis to individual preferences through hierarchical data representation [34, 51, 77, 76]. This approach improves prediction accuracy and accommodates diverse user interactions, enriching user experience and understanding sentiments in various contexts. As user-centric strategies gain importance, deriving tailored insights from survey data becomes crucial.

Incorporating NLP techniques in survey analysis marks a significant advancement in data interpretation. LLMs automate labor-intensive tasks like systematic literature reviews, maintaining factual accuracy while streamlining knowledge synthesis. They also show promise in improving software requirements quality, reducing development costs, and enhancing software quality. As research volume grows, automated LLM systems efficiently generate literature reviews, transforming survey-driven insights into actionable information. This AI integration optimizes processes and sets new standards for methodological rigor and transparency [70, 43, 14].

8.2 Techniques for Insight Extraction

Insight extraction techniques have evolved with large language models (LLMs), offering advanced capabilities for complex dataset analysis. Sentiment analysis assesses the emotional tone of survey responses, providing insights into consumer attitudes and preferences, crucial for strategic decision-making [47]. By quantifying emotions, researchers better understand consumer behavior drivers, informing marketing strategies and product developments.

Topic modeling, another key technique, identifies themes within survey data. LLMs automatically categorize and summarize large text volumes, enhancing thematic analysis efficiency and accuracy [3]. Dynamic topic model adjustments based on emerging trends keep insights relevant in a rapidly evolving market.

Chain-of-thought prompting refines LLMs' ability to generate insights from survey data, enhancing reasoning capabilities for complex questions and coherent responses [67]. This is beneficial for multifaceted responses, allowing comprehensive data exploration.

Advanced NLP techniques, including named entity recognition and dependency parsing, enhance targeted information extraction, improving data analysis efficiency and accuracy in research contexts. This is vital in automating literature reviews and systematic reviews, ensuring high factual accuracy and supporting scholarly research methodologies [43, 14]. These techniques identify entities, relationships, and dependencies, providing structured survey data understanding, aiding clarity and reproducibility of findings.

LLMs in survey analysis processes transform insight extraction methodologies, as seen in literature review automation. Fine-tuned LLMs handle labor-intensive stages efficiently, maintaining factual accuracy and addressing hallucinations. This enhances research synthesis efficiency and reliability, highlighting AI's growing role in academic methodologies, prompting updates to reporting guidelines for AI-driven approaches [43, 85, 14]. Leveraging LLMs' capabilities optimizes survey analysis strategies, driving innovation and enhancing survey-driven insights' impact.

8.3 Challenges in Survey Analysis with LLMs

Applying large language models (LLMs) in survey analysis presents challenges, notably in accuracy and reliability of insights from complex datasets. A major issue is LLMs' tendency to produce hallucinations or inaccurate outputs, especially with ambiguous or poorly structured responses [5].

Robust validation frameworks are needed to ensure LLM-generated insights' reliability, as flawed insights can lead to misguided decisions and undermine research credibility.

Biases in LLMs affect fairness and representativeness of survey analysis. The LABE benchmark emphasizes evaluating and mitigating AI-generated text biases to ensure accurate, equitable insights [4]. This is critical in surveys involving diverse groups, where biased outputs can skew interpretations. Researchers must assess model bias potential and implement corrective measures to reflect true sentiments of all respondent groups.

Language agency complexity complicates LLM use in survey analysis. Current benchmarks often use simplistic evaluation methods, failing to capture language's nuanced influence on perceptions and interactions [4]. Developing sophisticated evaluation frameworks is essential to assess language's impact on survey responses. This requires a multi-faceted evaluation approach, incorporating qualitative and quantitative metrics for comprehensive understanding.

Integrating LLMs with qualitative analysis tools poses challenges in user-friendliness and cost-effectiveness. Existing software lacks efficient, accessible qualitative data coding capabilities [3]. Developing intuitive, affordable tools to leverage LLM capabilities is necessary. Democratizing these technologies broadens adoption across research fields, benefiting diverse researchers with advanced analytical techniques.

LLMs' computational demands are a barrier to widespread survey analysis adoption. Substantial resources are needed, limiting scalability and feasibility in large-scale projects [2]. Optimizing resource management and developing efficient models are crucial to overcoming these constraints. Balancing computational efficiency and analytical depth is paramount for institutions and organizations implementing LLMs.

Challenges in LLM survey analysis highlight the need for ongoing innovation to improve software requirements quality and automate systematic literature reviews. These advancements aim to reduce costs and improve quality, streamlining research methodologies. Refining LLM applications enhances accuracy, bias, language agency, usability, and computational efficiency, improving reliability and impact of LLM-driven survey insights [70, 14].

9 Challenges and Future Directions

The integration of large language models (LLMs) with databases faces multifaceted technical limitations that impede their scalability and effectiveness. These constraints encompass computational demands, alignment issues, and the interpretability of model outputs. Addressing these technical challenges is crucial for advancing LLM integration, necessitating innovative solutions to navigate complexities and contribute to this rapidly evolving field.

9.1 Technical Limitations

Integrating LLMs with databases encounters significant technical hurdles that challenge scalability and effectiveness. The computational intensity required for LLM training imposes substantial resource demands, limiting large-scale application deployment [65]. This burden is exacerbated by the complexity of aligning LLMs with human-like memory processes, affecting their ability to replicate nuanced cognitive functions. High-performance computing reliance increases costs and accessibility concerns, potentially widening the gap between large tech companies and academia.

Critical limitations include visual and language model alignment, as demonstrated by the MPlug framework, where misalignment leads to inefficiencies in generating coherent multi-modal outputs [66]. This poses challenges in applications requiring seamless integration of visual and textual data, impacting overall LLM-based system performance. Enhanced alignment techniques are essential for applications leveraging both visual and textual information, like autonomous driving and augmented reality.

Accuracy issues in loss predictions, highlighted by the nanoLM benchmark, undermine LLM pre-training and integration efforts [7]. More precise predictive models are needed to enhance LLM training reliability. Furthermore, ensuring LLMs generalize effectively across domains requires robust evaluation metrics for real-world scenario performance.

AI-generated, for reference only.

LLM output interpretability remains challenging as models struggle to generate specific insights from generalized data, affecting data-driven decision quality. This limitation is pronounced in human-robot interactions, where complex dialogues necessitate sophisticated communication approaches. Recent advancements, such as the Inter-Chunk Interaction Retrieval (IIER) system, improve contextual understanding by leveraging structural, keyword, and semantic interactions, enhancing human-robot communication quality [15, 86, 16, 87, 67].

Addressing technical limitations requires continuous research and development to refine integration methodologies, advance system architectures, and optimize resource management practices. This effort is crucial for enhancing AI-driven tools' accuracy and reliability in academic and software contexts while addressing ethical concerns like bias and misinformation [70, 86, 14]. Improving LLM adaptability and scalability ensures effective integration with databases and maximizes impact across diverse applications.

9.2 Data Privacy and Security Concerns

Integrating LLMs with databases raises significant data privacy and security challenges, necessitating robust protective measures. AI-generated misinformation risks compromising sensitive data integrity and confidentiality, emphasizing rigorous validation processes for AI output accuracy [6]. In healthcare, stringent privacy measures are crucial to prevent unauthorized access to confidential data and ensure compliance with legal and ethical standards.

LLMs' potential to inadvertently expose sensitive information necessitates comprehensive privacy safeguards to protect user data and maintain trust in AI-driven applications [6]. Safeguards must include technological solutions and organizational policies prioritizing data ethics and user consent.

LLMs generating queries from natural language pose risks, as adversaries may exploit system vulnerabilities to access sensitive information, especially when LLMs produce content mirroring copyrighted materials [88, 89]. Developing secure integration frameworks to mitigate potential threats and ensure safe processing of sensitive information is essential.

Addressing data privacy and security concerns is vital for mitigating risks related to copyright infringement and academic dishonesty, promoting responsible AI deployment in scholarly communication and education. With LLMs assisting in generating academic content, robust safeguards against misuse and protecting intellectual property rights are critical for maintaining academic integrity and public trust in AI applications [90, 85, 89]. Implementing rigorous validation processes and adhering to regulatory requirements can safeguard user data and enhance AI-driven data management practices' security and reliability.

9.3 Evaluation and Metrics

Benchmark	Size	Domain	Task Format	Metric
CN-Bench[91]	11,651	Hate Speech	Counter Narrative Generation	BERTScore, JudgeLM
SDAAP[83]	20,305	Spectral Analysis	Question Answering	BLEU, ROUGE
ARAGOG[92]	423	Artificial Intelligence	Question Answering	Retrieval Precision, Answer Similarity
FinBench[93]	103,000	Finance	Dialogue Generation	GPT-4 Evaluation Score
LLM-Eval[94]	7,300	Text Generation	Text Generation	Linguistic Acceptability, Output Content Quality
Masquerade-23[95]	2,400,000	Social Networks	Behavior Analysis	Toxicity Score, Content Similarity
SciTLDR[43]	5,400	Literature Review	Summarization	ROUGE-1, ROUGE-2
LLM-CryptoBench[96]	1,186	Cryptography	Cryptographic Misuse Detection	Actionability, Specificity

Table 4: The table presents a comprehensive overview of various benchmarks utilized for evaluating large language models (LLMs) across diverse domains. It includes essential details such as benchmark size, domain specificity, task format, and the metrics employed for performance assessment, thus providing a structured framework for understanding the multifaceted evaluation processes of LLMs.

Evaluating LLMs integration with databases requires comprehensive performance metrics and benchmarks tailored to specific applications. The F1-score is pivotal, balancing precision and recall, providing a robust measure of model performance on tasks where false positives and negatives

AI-generated, for reference only.

are significant [97]. A multi-faceted evaluation approach is essential to capture LLM performance intricacies.

LLM assessments involve various metrics, including accuracy, precision, recall, and AUC, capturing different aspects of model performance. Benchmarks like SciAssess evaluate LLMs on memorization, comprehension, and reasoning capabilities, while frameworks like TrustScore emphasize trustworthiness in LLM outputs [98, 99, 100, 101, 88]. These metrics highlight areas for improvement in data processing and retrieval. Table 4 illustrates the diverse benchmarks employed in assessing large language models, highlighting the importance of tailored evaluation metrics for different domains and tasks.

Benchmarking frameworks like nanoLM guide efficient loss prediction and hyperparameter optimization strategies [7]. Continuous evolution of these frameworks is vital for keeping pace with rapid LLM architecture advancements, ensuring evaluation practices remain relevant and rigorous.

LLM output interpretability influences insights' quality from integrated systems. Techniques like chain-of-thought prompting and structured reasoning paths improve clarity and relevance, enhancing LLMs' ability to provide accurate, context-aware responses [102, 1, 14, 67]. This is crucial in data-driven decision-making contexts, where meaningful insights lead to informed strategic choices.

Comprehensive evaluation strategies, including benchmarks like SciAssess and methodologies for LLM-as-a-Judge, enhance integration processes, streamlining labor-intensive tasks like systematic literature reviews and improving LLM reliability and accuracy in data management [99, 43, 14, 103]. This approach fosters innovation and maximizes LLMs' impact across academic and research domains, ensuring methodological transparency and addressing challenges like hallucination and bias.

9.4 Handling Multimodal Data

Handling multimodal data in LLM-database integration involves sophisticated approaches enhancing diverse data types' processing and analysis. Multimodal data, including text, images, audio, and video, requires models integrating and interpreting varied inputs for coherent outputs. The MPlug framework exemplifies this capability through a two-stage process aligning images and text, followed by fine-tuning with language-only and multi-modal instructions, improving efficiency [66].

The TransGPT framework adapts LLMs to multimodal tasks, processing textual and visual data for comprehensive insights in transportation [50]. Such advancements are crucial for applications like autonomous vehicles and smart cities, enhancing operational efficiencies and user experiences.

Graph reasoning tasks integration, as seen in Graph-ToolFormer, empowers models to process graph-structured data, enhancing interconnected multimodal inputs handling [53]. This capability is vital for applications involving complex data relationships, like social networks and recommendation systems.

VD Prompting enhances LLMs' performance in generating accurate answers from long-context documents, improving data processing efficiency and user engagement [46]. This methodology benefits academic research and content generation, where information volume can be overwhelming.

Advanced integration strategies leveraging fine-tuned LLMs enhance multimodal dataset handling efficiency and accuracy across sectors like academic research and environmental management. LLMs automate systematic literature reviews, improving transparency and reliability, while extracting critical information from unstructured historical documents, streamlining processes and reducing environmental risks [35, 92, 14, 73].

9.5 Future Research Directions

Future research on LLM-database integration should prioritize enhancing system robustness and applicability across domains. Refining tools like QualiGPT to explore capabilities and enhance human-AI collaboration in qualitative analysis is crucial [3]. Expanding datasets and exploring biases across demographic identities will improve language agency evaluation in LLMs, ensuring equitable AI applications [4].

Developing efficient models is essential, particularly in healthcare, where enhancing stakeholder collaboration and establishing regulatory frameworks facilitate responsible AI deployment [6]. Extending benchmarks to specialized domains and incorporating multi-hop retrieval techniques for complex queries are vital for improving LLM performance [5].

Future studies should explore cognitive tasks and refine benchmarks for better working memory capabilities evaluation, improving cognitive modeling [12]. Enhancing models like MPlug for document comprehension and scene text understanding is crucial for multi-modal applications [66].

Extending frameworks like nanoLM to encompass more tasks and integration scenarios will drive innovation in LLM-database systems, ensuring adaptability across applications [7]. Addressing these research directions will advance LLM-database integration, ensuring LLMs evolve to meet dynamic sector needs, enhancing capabilities and contributing to broader AI understanding and societal role.

10 Conclusion

Integrating large language models (LLMs) with databases marks a pivotal advancement in artificial intelligence, significantly enhancing the efficiency and capabilities of data management and processing across various sectors. This integration not only optimizes computational processes but also enriches user experiences through the development of sophisticated memory architectures. In the healthcare domain, LLMs have shown promise in improving patient care and health outcomes by facilitating timely access to accurate information, which is crucial for informed clinical decision-making.

Furthermore, user experience studies highlight the effectiveness of LLMs in conveying explainable artificial intelligence (XAI) methods, with a majority of users favoring these explanations over traditional approaches. This preference emphasizes the importance of designing AI systems that prioritize user comprehension and trust, particularly in applications where clarity and reliability are essential. By providing clearer insights and facilitating informed decision-making, LLMs foster greater acceptance and trust in AI technologies among users.

The integration of LLMs with databases thus represents more than a technical enhancement; it signifies a transformative shift in the way information is processed and understood across diverse fields. These advancements drive innovation and efficiency in data management, paving the way for the development of sophisticated AI-driven applications. The implications of this integration extend beyond technical improvements, influencing user engagement and fostering trust in AI systems. Ongoing research and development efforts are crucial to further expanding the capabilities and applications of LLMs within the dynamic landscape of artificial intelligence. Future studies should address existing challenges, such as biases in LLM outputs and the interpretability of their decision-making processes, ensuring that these technologies are not only effective but also ethical and equitable in their applications.

References

[1] Xiang Deng, Vasilisa Bashlovkina, Feng Han, Simon Baumgartner, and Michael Bendersky. What do llms know about financial markets? a case study on reddit market sentiment analysis, 2022.

[2] Sen Huang, Kaixiang Yang, Sheng Qi, and Rui Wang. When large language model meets optimization, 2024.

[3] He Zhang, Chuhao Wu, Jingyi Xie, Fiona Rubino, Sydney Graver, ChanMin Kim, John M. Carroll, and Jie Cai. When qualitative research meets large language model: Exploring the potential of qualigpt as a tool for qualitative coding, 2024.

[4] Yixin Wan and Kai-Wei Chang. White men lead, black women help? benchmarking language agency social biases in llms, 2024.

[5] Sina J. Semnani, Violet Z. Yao, Heidi C. Zhang, and Monica S. Lam. Wikichat: Stopping the hallucination of large language model chatbots by few-shot grounding on wikipedia, 2023.

[6] Mert Karabacak and Konstantinos Margetis. Embracing large language models for medical applications: opportunities and challenges. *Cureus*, 15(5), 2023.

[7] Yiqun Yao, Siqi fan, Xiusheng Huang, Xuezhi Fang, Xiang Li, Ziyi Ni, Xin Jiang, Xuying Meng, Peng Han, Shuo Shang, Kang Liu, Aixin Sun, and Yequan Wang. nanolm: an affordable llm pre-training benchmark via accurate loss prediction across scales, 2024.

[8] Philip Mavrepis, Georgios Makridis, Georgios Fatouros, Vasileios Koukos, Maria Margarita Separdani, and Dimosthenis Kyriazis. Xai for all: Can large language models simplify explainable ai?, 2024.

[9] Trong-Hieu Nguyen, Anh-Cuong Le, and Viet-Cuong Nguyen. Villm-eval: A comprehensive evaluation suite for vietnamese large language models, 2024.

[10] Tianyu Wang, Yifan Li, Haitao Lin, Xiangyang Xue, and Yanwei Fu. Wall-e: Embodied robotic waiter load lifting with large language model, 2023.

[11] Chenyu Tang, Shuo Gao, Cong Li, Wentian Yi, Yuxuan Jin, Xiaoxue Zhai, Sixuan Lei, Hongbei Meng, Zibo Zhang, Muzi Xu, Shengbo Wang, Xuhang Chen, Chenxi Wang, Hongyun Yang, Ningli Wang, Wenyu Wang, Jin Cao, Xiaodong Feng, Peter Smielewski, Yu Pan, Wenhui Song, Martin Birchall, and Luigi G. Occhipint. Wearable intelligent throat enables natural speech in stroke patients with dysarthria, 2024.

[12] Dongyu Gong, Xingchen Wan, and Dingmin Wang. Working memory capacity of chatgpt: An empirical study, 2024.

[13] Baolin Li, Yankai Jiang, Vijay Gadepally, and Devesh Tiwari. Llm inference serving: Survey of recent advances and opportunities, 2024.

[14] Teo Susnjak, Peter Hwang, Napoleon H. Reyes, Andre L. C. Barczak, Timothy R. McIntosh, and Surangika Ranathunga. Automating research synthesis with domain-specific large language model fine-tuning, 2024.

[15] Tiezheng Guo, Chen Wang, Yanyi Liu, Jiawei Tang, Pan Li, Sai Xu, Qingwen Yang, Xianlin Gao, Zhi Li, and Yingyou Wen. Leveraging inter-chunk interactions for enhanced retrieval in large language model-based question answering, 2024.

[16] Shengnan An, Zexiong Ma, Zeqi Lin, Nanning Zheng, and Jian-Guang Lou. Make your llm fully utilize the context. *arXiv preprint arXiv:2404.16811*, 2024.

[17] Nikolay Bogoychev, Pinzhen Chen, Barry Haddow, and Alexandra Birch. The ups and downs of large language model inference with vocabulary trimming by language heuristics, 2024.

[18] Xingrun Xing, Boyan Gao, Zheng Zhang, David A. Clifton, Shitao Xiao, Li Du, Guoqi Li, and Jiajun Zhang. Spikellm: Scaling up spiking neural network to large language models via saliency-based spiking, 2025.

AI-generated, for reference only.

[19] Ziteng Sun, Ananda Theertha Suresh, Jae Hun Ro, Ahmad Beirami, Himanshu Jain, and Felix Yu. Spectr: Fast speculative decoding via optimal transport, 2024.

[20] Xupeng Miao, Gabriele Oliaro, Zhihao Zhang, Xinhao Cheng, Zeyu Wang, Zhengxin Zhang, Rae Ying Yee Wong, Alan Zhu, Lijie Yang, Xiaoxiang Shi, Chunan Shi, Zhuoming Chen, Daiyaan Arfeen, Reyna Abhyankar, and Zhihao Jia. Specinfer: Accelerating generative large language model serving with tree-based speculative inference and verification, 2024.

[21] Cheng-Kuang Wu, Zhi Rui Tam, Chieh-Yen Lin, Yun-Nung Chen, and Hung yi Lee. Streambench: Towards benchmarking continuous improvement of language agents, 2024.

[22] Minghao Shao, Abdul Basit, Ramesh Karri, and Muhammad Shafique. Survey of different large language model architectures: Trends, benchmarks, and challenges, 2024.

[23] Chenxi Sun, Hongyan Li, Yaliang Li, and Shenda Hong. Test: Text prototype aligned embedding to activate llm's ability for time series, 2024.

[24] Guoxing Yang, Jianyu Shi, Zan Wang, Xiaohong Liu, and Guangyu Wang. Tcm-gpt: Efficient pre-training of large language models for domain adaptation in traditional chinese medicine, 2023.

[25] Samuel Marks and Max Tegmark. The geometry of truth: Emergent linear structure in large language model representations of true/false datasets, 2024.

[26] Matias Martinez. The impact of hyperparameters on large language model inference performance: An evaluation of vllm and huggingface pipelines, 2024.

[27] Guanying Jiang, Lingyong Yan, Haibo Shi, and Dawei Yin. The real, the better: Aligning large language models with online human behaviors, 2024.

[28] Bo Ni, Yu Wang, Lu Cheng, Erik Blasch, and Tyler Derr. Towards trustworthy knowledge graph reasoning: An uncertainty aware perspective, 2024.

[29] Maurice Funk, Simon Hosemann, Jean Christoph Jung, and Carsten Lutz. Towards ontology construction with language models, 2023.

[30] Joonhyung Lee, Jeongin Bae, Byeongwook Kim, Se Jung Kwon, and Dongsoo Lee. To fp8 and back again: Quantifying the effects of reducing precision on llm training stability, 2024.

[31] Justin Cosentino, Anastasiya Belyaeva, Xin Liu, Nicholas A. Furlotte, Zhun Yang, Chace Lee, Erik Schenck, Yojan Patel, Jian Cui, Logan Douglas Schneider, Robby Bryant, Ryan G. Gomes, Allen Jiang, Roy Lee, Yun Liu, Javier Perez, Jameson K. Rogers, Cathy Speed, Shyam Tailor, Megan Walker, Jeffrey Yu, Tim Althoff, Conor Heneghan, John Hernandez, Mark Malhotra, Leor Stern, Yossi Matias, Greg S. Corrado, Shwetak Patel, Shravya Shetty, Jiening Zhan, Shruthi Prabhakara, Daniel McDuff, and Cory Y. McLean. Towards a personal health large language model, 2024.

[32] Tianyu He, Guanghui Fu, Yijing Yu, Fan Wang, Jianqiang Li, Qing Zhao, Changwei Song, Hongzhi Qi, Dan Luo, Huijing Zou, and Bing Xiang Yang. Towards a psychological generalist ai: A survey of current applications of large language models and future prospects, 2023.

[33] Jochen Wulf and Juerg Meierhofer. Towards a taxonomy of large language model based business model transformations, 2023.

[34] Jiayin Wang, Weizhi Ma, Peijie Sun, Min Zhang, and Jian-Yun Nie. Understanding user experience in large language model interactions, 2024.

[35] Crystal Qian, Emily Reif, and Minsuk Kahng. Understanding the dataset practitioners behind large language model development, 2024.

[36] Leland Bybee. Surveying generative ai's economic expectations, 2023.

[37] Lei Wang, Yi Hu, Jiabang He, Xing Xu, Ning Liu, Hui Liu, and Heng Tao Shen. T-sciq: Teaching multimodal chain-of-thought reasoning via mixed large language model signals for science question answering, 2023.

AI-generated, for reference only.

[38] Ben Wang, Jiqun Liu, Jamshed Karimnazarov, and Nicolas Thompson. Task supportive and personalized human-large language model interaction: A user study, 2024.

[39] Pai Zeng, Zhenyu Ning, Jieru Zhao, Weihao Cui, Mengwei Xu, Liwei Guo, Xusheng Chen, and Yizhou Shan. The cap principle for llm serving: A survey of long-context large language model serving, 2024.

[40] Partha Sen and Sumana Sen. Graph database while computationally efficient filters out quickly the esg integrated equities in investment management, 2024.

[41] Qinghua Guan, Jinhui Ouyang, Di Wu, and Weiren Yu. Citygpt: Towards urban iot learning, analysis and interaction with multi-agent system, 2024.

[42] Jiale Lao, Yibo Wang, Yufei Li, Jianping Wang, Yunjia Zhang, Zhiyuan Cheng, Wanghu Chen, Mingjie Tang, and Jianguo Wang. Gptuner: A manual-reading database tuning system via gpt-guided bayesian optimization, 2024.

[43] Nurshat Fateh Ali, Md. Mahdi Mohtasim, Shakil Mosharrof, and T. Gopi Krishna. Automated literature review using nlp techniques and llm-based retrieval-augmented generation, 2024.

[44] Shiwei Zhang, Mingfang Wu, and Xiuzhen Zhang. Utilising a large language model to annotate subject metadata: A case study in an australian national research data catalogue, 2023.

[45] Ziyang Chen and Stylios Moscholios. Using prompts to guide large language models in imitating a real person's language style, 2024.

[46] Sakshi Mahendru and Tejul Pandit. Venn diagram prompting : Accelerating comprehension with scaffolding effect, 2024.

[47] Jiahao Wang and Amer Shalaby. Transit pulse: Utilizing social media as a source for customer feedback and information extraction with large language model, 2024.

[48] Karthik Suresh, Neeltje Kackar, Luke Schleck, and Cristiano Fanelli. Towards a rag-based summarization agent for the electron-ion collider, 2024.

[49] Mike A. Merrill, Akshay Paruchuri, Naghmeh Rezaei, Geza Kovacs, Javier Perez, Yun Liu, Erik Schenck, Nova Hammerquist, Jake Sunshine, Shyam Tailor, Kumar Ayush, Hao-Wei Su, Qian He, Cory Y. McLean, Mark Malhotra, Shwetak Patel, Jiening Zhan, Tim Althoff, Daniel McDuff, and Xin Liu. Transforming wearable data into health insights using large language model agents, 2024.

[50] Peng Wang, Xiang Wei, Fangxu Hu, and Wenjuan Han. Transgpt: Multi-modal generative pre-trained transformer for transportation, 2024.

[51] Chenkai Sun, Ke Yang, Revanth Gangi Reddy, Yi R. Fung, Hou Pong Chan, Kevin Small, ChengXiang Zhai, and Heng Ji. Persona-db: Efficient large language model personalization for response prediction with collaborative data refinement, 2025.

[52] Zhaodonghui Li, Haitao Yuan, Huiming Wang, Gao Cong, and Lidong Bing. Llm-r2: A large language model enhanced rule-based rewrite system for boosting query efficiency, 2024.

[53] Jiawei Zhang. Graph-toolformer: To empower llms with graph reasoning ability via prompt augmented by chatgpt, 2023.

[54] Le Xiao and Xin Shan. Patterngpt :a pattern-driven framework for large language model text generation, 2023.

[55] Jianchao Ji, Zelong Li, Shuyuan Xu, Wenyue Hua, Yingqiang Ge, Juntao Tan, and Yongfeng Zhang. Genrec: Large language model for generative recommendation, 2023.

[56] Jiongnan Liu, Jiajie Jin, Zihan Wang, Jiehan Cheng, Zhicheng Dou, and Ji-Rong Wen. Reta-llm: A retrieval-augmented large language model toolkit, 2023.

[57] Song Tong, Kai Mao, Zhen Huang, Yukun Zhao, and Kaiping Peng. Automating psychological hypothesis generation with ai: when large language models meet causal graph, 2024.

AI-generated, for reference only.

[58] Yingqiang Ge, Wenyue Hua, Kai Mei, Juntao Tan, Shuyuan Xu, Zelong Li, Yongfeng Zhang, et al. Openagi: When llm meets domain experts. *Advances in Neural Information Processing Systems*, 36:5539–5568, 2023.

[59] Lishan Zhang, Han Wu, Xiaoshan Huang, Tengfei Duan, and Hanxiang Du. Automatic deductive coding in discourse analysis: an application of large language models in learning analytics, 2024.

[60] Kexin Chen, Hanqun Cao, Junyou Li, Yuyang Du, Menghao Guo, Xin Zeng, Lanqing Li, Jiezhong Qiu, Pheng Ann Heng, and Guangyong Chen. An autonomous large language model agent for chemical literature data mining, 2024.

[61] Longyu Feng, Huahang Li, and Chen Jason Zhang. Prompt-matcher: Leveraging large models to reduce uncertainty in schema matching results, 2025.

[62] Yuji Cao, Huan Zhao, Yuheng Cheng, Ting Shu, Yue Chen, Guolong Liu, Gaoqi Liang, Junhua Zhao, Jinyue Yan, and Yun Li. Survey on large language model-enhanced reinforcement learning: Concept, taxonomy, and methods, 2024.

[63] Jan Clusmann, Fiona R Kolbinger, Hannah Sophie Muti, Zunamys I Carrero, Jan-Niklas Eckardt, Narmin Ghaffari Laleh, Chiara Maria Lavinia Löffler, Sophie-Caroline Schwarzkopf, Michaela Unger, Gregory P Veldhuizen, et al. The future landscape of large language models in medicine. *Communications medicine*, 3(1):141, 2023.

[64] Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[65] Yuki Hou, Haruki Tamoto, and Homei Miyashita. "my agent understands me better": Integrating dynamic human-like memory recall and consolidation in llm-based agents, 2024.

[66] Qinghao Ye, Haiyang Xu, Guohai Xu, Jiabo Ye, Ming Yan, Yiyang Zhou, Junyang Wang, Anwen Hu, Pengcheng Shi, Yaya Shi, et al. mplug-owl: Modularization empowers large language models with multimodality. *arXiv preprint arXiv:2304.14178*, 2023.

[67] Kewei Cheng, Nesreen K. Ahmed, Theodore Willke, and Yizhou Sun. Structure guided prompt: Instructing large language model in multi-step reasoning by exploring graph structure of the text, 2024.

[68] Changle Qu, Sunhao Dai, Xiaochi Wei, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, Jun Xu, and Ji-Rong Wen. Tool learning with large language models: A survey, 2024.

[69] Vincent Jung and Lonneke van der Plas. Understanding the effects of language-specific class imbalance in multilingual fine-tuning, 2024.

[70] Sebastian Lubos, Alexander Felfernig, Thi Ngoc Trang Tran, Damian Garber, Merfat El Mansi, Seda Polat Erdeniz, and Viet-Man Le. Leveraging llms for the quality assurance of software requirements, 2024.

[71] Maojun Sun, Ruijian Han, Binyan Jiang, Houduo Qi, Defeng Sun, Yancheng Yuan, and Jian Huang. A survey on large language model-based agents for statistics and data science, 2024.

[72] Xuanhe Zhou, Guoliang Li, and Zhiyuan Liu. Llm as dba. *arXiv preprint arXiv:2308.05481*, 2023.

[73] Zhiwei Ma, Javier E. Santo, Greg Lackey, Hari Viswanathan, and Daniel O'Malley. Information extraction from historical well records using a large language model, 2024.

[74] Ruoxi Sun, Sercan Ö. Arik, Alex Muzio, Lesly Miculicich, Satya Gundabathula, Pengcheng Yin, Hanjun Dai, Hootan Nakhost, Rajarishi Sinha, Zifeng Wang, and Tomas Pfister. Sql-palm: Improved large language model adaptation for text-to-sql (extended), 2024.

[75] Tristan Kenneweg, Philip Kenneweg, and Barbara Hammer. Retrieval augmented generation systems: Automatic dataset creation, evaluation and boolean agent setup, 2024.

[76] Yiren Liu, Pranav Sharma, Mehul Jitendra Oswal, Haijun Xia, and Yun Huang. Personaflow: Boosting research ideation with llm-simulated expert personas, 2024.

[77] Pawel Robert Smolinski, Joseph Januszewicz, and Jacek Winiarski. Scaling technology acceptance analysis with large language model (llm) annotation systems, 2024.

[78] Qiaoyu Tang, Jiawei Chen, Zhuoqun Li, Bowen Yu, Yaojie Lu, Cheng Fu, Haiyang Yu, Hongyu Lin, Fei Huang, Ben He, Xianpei Han, Le Sun, and Yongbin Li. Self-retrieval: End-to-end information retrieval with one large language model, 2024.

[79] Ruiwen Zhou, Yingxuan Yang, Muning Wen, Ying Wen, Wenhao Wang, Chunling Xi, Guoqiang Xu, Yong Yu, and Weinan Zhang. Trad: Enhancing llm agents with step-wise thought retrieval and aligned decision, 2024.

[80] Hejia Geng, Boxun Xu, and Peng Li. Upar: A kantian-inspired prompting framework for enhancing large language model capabilities, 2023.

[81] Zhaowei Li, Wei Wang, YiQing Cai, Xu Qi, Pengyu Wang, Dong Zhang, Hang Song, Botian Jiang, Zhida Huang, and Tao Wang. Unifiedmllm: Enabling unified representation for multimodal multi-tasks with large language model, 2024.

[82] Sachin Pathiyan Cherumanal, Lin Tian, Futoon M. Abushaqra, Angel Felipe Magnossao de Paula, Kaixin Ji, Danula Hettiachchi, Johanne R. Trippas, Halil Ali, Falk Scholer, and Damiano Spina. Walert: Putting conversational search knowledge into action by building and evaluating a large language model-powered chatbot, 2024.

[83] Jiheng Liang, Ziru Yu, Zujie Xie, and Xiangyang Yu. A quick, trustworthy spectral knowledge qa system leveraging retrieval-augmented generation on llm, 2024.

[84] Furui Cheng, Vilém Zouhar, Simran Arora, Mrinmaya Sachan, Hendrik Strobelt, and Mennatallah El-Assady. Relic: Investigating large language model responses using self-consistency, 2024.

[85] Andrew Gray. Chatgpt "contamination": estimating the prevalence of llms in the scholarly literature, 2024.

[86] Jesse G Meyer, Ryan J Urbanowicz, Patrick CN Martin, Karen O'Connor, Ruowang Li, Pei-Chen Peng, Tiffani J Bright, Nicholas Tatonetti, Kyoung Jae Won, Graciela Gonzalez-Hernandez, et al. Chatgpt and large language models in academia: opportunities and challenges. *BioData Mining*, 16(1):20, 2023.

[87] Nhat Tran and Diane Litman. Enhancing knowledge retrieval with topic modeling for knowledge-grounded dialogue, 2024.

[88] Xiaonan Li, Changtai Zhu, Linyang Li, Zhangyue Yin, Tianxiang Sun, and Xipeng Qiu. Llatrieval: Llm-verified retrieval for verifiable generation, 2024.

[89] Weijie Zhao, Huajie Shao, Zhaozhuo Xu, Suzhen Duan, and Denghui Zhang. Measuring copyright risks of large language model via partial information probing, 2024.

[90] Suriya Prakash Jambunathan, Ashwath Shankarnarayan, and Parijat Dube. Convnlp: Image-based ai text detection, 2024.

[91] Irune Zubiaga, Aitor Soroa, and Rodrigo Agerri. A llm-based ranking method for the evaluation of automatic counter-narrative generation, 2024.

[92] Matouš Eibich, Shivay Nagpal, and Alexander Fred-Ojala. Aragog: Advanced rag output grading, 2024.

[93] Ziao Wang, Jianning Wang, Junda Wu, and Xiaofeng Zhang. An effective data creation pipeline to generate high-quality financial instruction data for large language model, 2023.

[94] Rishav Hada, Varun Gumma, Adrian de Wynter, Harshita Diddee, Mohamed Ahmed, Monojit Choudhury, Kalika Bali, and Sunayana Sitaram. Are large language model-based evaluators the solution to scaling up multilingual evaluation?, 2024.

AI-generated, for reference only.

[95] Siyu Li, Jin Yang, and Kui Zhao. Are you in a masquerade? exploring the behavior and impact of large language model driven social bots in online social networks, 2023.

[96] Zohaib Masood and Miguel Vargas Martin. Beyond static tools: Evaluating large language models for cryptographic misuse detection, 2024.

[97] Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. Bloomberggpt: A large language model for finance, 2023.

[98] Marco AF Pimentel, Clément Christophe, Tathagata Raha, Prateek Munjal, Praveen K Kanithi, and Shadab Khan. Beyond metrics: A critical analysis of the variability in large language model evaluation frameworks, 2024.

[99] Hengxing Cai, Xiaochen Cai, Junhan Chang, Sihang Li, Lin Yao, Changxin Wang, Zhifeng Gao, Hongshuai Wang, Yongge Li, Mujie Lin, Shuwen Yang, Jiankun Wang, Mingjun Xu, Jin Huang, Xi Fang, Jiaxi Zhuang, Yuqi Yin, Yaqi Li, Changhong Chen, Zheng Cheng, Zifeng Zhao, Linfeng Zhang, and Guolin Ke. Sciassess: Benchmarking llm proficiency in scientific literature analysis, 2024.

[100] Danna Zheng, Danyang Liu, Mirella Lapata, and Jeff Z. Pan. Trustscore: Reference-free evaluation of llm response trustworthiness, 2024.

[101] Kun Zhou, Yutao Zhu, Zhipeng Chen, Wentong Chen, Wayne Xin Zhao, Xu Chen, Yankai Lin, Ji-Rong Wen, and Jiawei Han. Don't make your llm an evaluation benchmark cheater. *arXiv preprint arXiv:2311.01964*, 2023.

[102] Junjie Wang, Dan Yang, Binbin Hu, Yue Shen, Wen Zhang, and Jinjie Gu. Know your needs better: Towards structured understanding of marketer demands with analogical reasoning augmented llms, 2024.

[103] Jiawei Gu, Xuhui Jiang, Zhichao Shi, Hexiang Tan, Xuehao Zhai, Chengjin Xu, Wei Li, Yinghan Shen, Shengjie Ma, Honghao Liu, et al. A survey on llm-as-a-judge. *arXiv preprint arXiv:2411.15594*, 2024.

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.