A Survey of Educational Knowledge Graphs, Deep Learning, Large Language Models, and Multimodal Integration

Abstract

This survey paper explores the integration of advanced computational frameworks in educational settings, focusing on educational knowledge graphs, deep learning, large language models (LLMs), multimodal approaches, semantic web, ontology, and data integration. The paper examines recent advancements and challenges in multimodal machine learning, particularly the unification of LLMs and knowledge graphs to enhance factual knowledge recall and content generation. It highlights the importance of transparency in LLMs and reviews the construction and application of multimodal knowledge graphs (MMKGs), emphasizing text and image integration. The survey also investigates generative AI applications in coding and data analysis within computational social sciences, aiming to lower barriers in these fields. Additionally, it analyzes recursive modality changes in generative AI models, focusing on content stability and divergence. The study covers ADHD-related information organization using network-based approaches and evaluates recent LLM developments, addressing training strategies and applications. The scope includes multimodal pretrained models, large models, and instruction tuning strategies, while excluding broader AI technology adoption discussions. Potential limitations include dataset reliance and scalability issues. The survey aims to provide a comprehensive overview of integrated educational technologies, facilitating exploration of multimodal datasets to predict concept learnability and advancing multimodal understanding through new dataset formats.

1 Introduction

1.1 Significance of Integration

The integration of advanced computational technologies in education is essential for managing complex educational data characterized by multimedia, multistructure, multisource, multimodal, and multiversion dimensions [1]. This integration enhances the understanding of educational content across various modalities—text, images, audio, and video—thereby improving educational technologies [2]. The increasing reliance on remote teaching and recorded lectures further emphasizes the necessity of multimodal data integration for effective student learning [3].

Multimodal large language models (MLLMs) play a crucial role in advancing educational methodologies by processing extensive image sequences and improving reasoning capabilities through data synthesis. They address existing knowledge gaps and challenges, underscoring their relevance in educational contexts [4]. Incorporating structural information into incomplete multimodal knowledge graphs (MKGs) enhances reasoning performance, which is vital for educational applications [5].

Combining knowledge graphs (KGs) with large language models (LLMs) mitigates LLMs' limitations in recalling factual knowledge and generating knowledge-grounded content, thus enhancing interpretability and factual accuracy in educational applications [6]. Multi-Task Learning (MTL) tech-

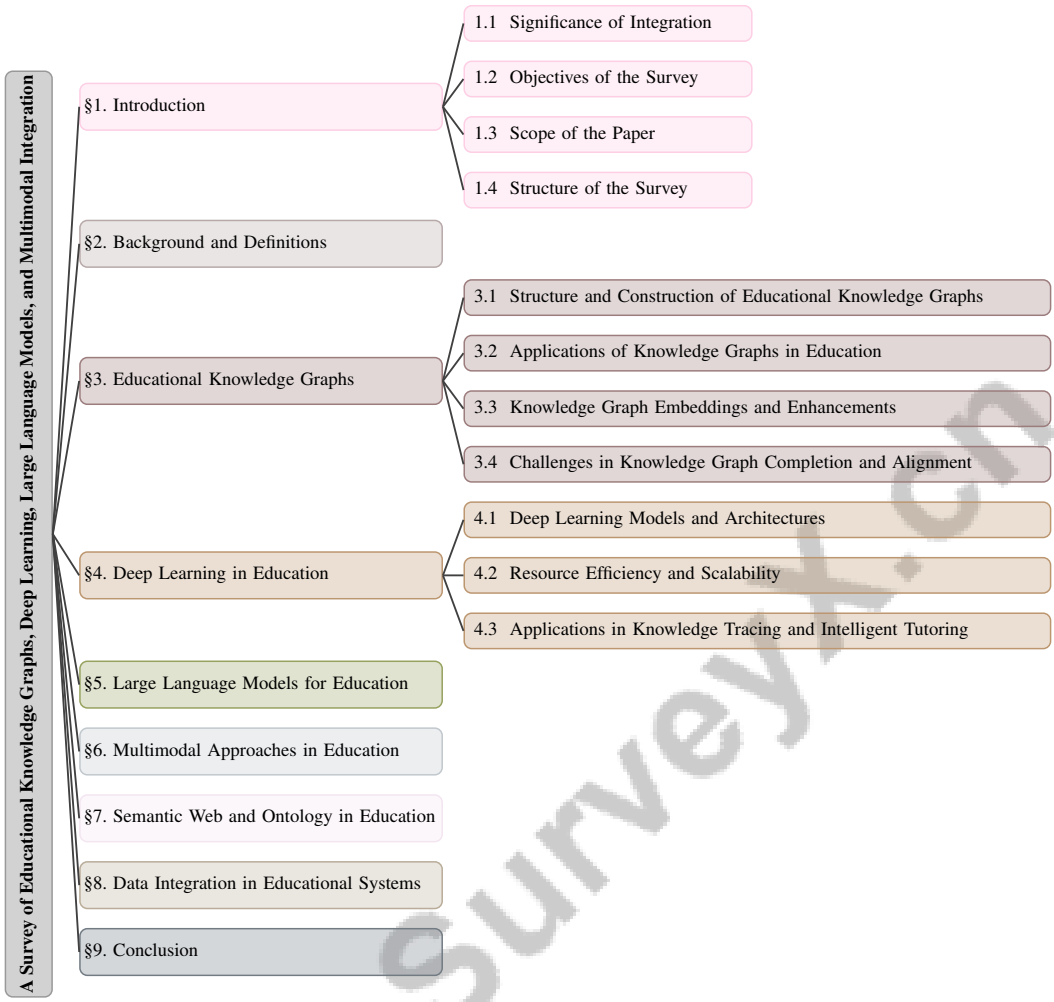


Figure 1: chapter structure

niques leverage shared information to boost learning efficiency across educational fields, showcasing the transformative potential of integrated educational technologies [7].

The Bloom Library exemplifies the importance of expanding multimodal resources for non-dominant language communities, promoting language diversity and indigenous perspectives within the NLP research community [8]. The demand for effective assessments in programming education, such as high-quality multiple-choice questions (MCQs), further highlights the need for integrated educational technologies [9].

Integrating multimodal data, knowledge graphs, and advanced machine learning models fosters a comprehensive understanding of complex educational data interactions, ultimately enhancing the educational experience by providing more effective, comprehensive, and accessible learning opportunities. Curriculum learning significantly contributes to rapid learning and effective pedagogy, emphasizing the importance of structured educational approaches [10].

1.2 Objectives of the Survey

This survey aims to explore the integration of advanced computational frameworks in educational environments, focusing on educational knowledge graphs, deep learning, large language models, multimodal approaches, semantic web, ontology, and data integration. A primary objective is to investigate recent advancements in multimodal machine learning, addressing foundational challenges and methods to enhance educational methodologies [11]. The survey will critically examine the

unification of LLMs and KGs, emphasizing strategies to overcome LLMs' limitations in capturing factual knowledge and the challenges in constructing and evolving KGs.

Another significant aim is to develop a human-centered roadmap for transparency approaches tailored to diverse stakeholder needs, exploring the challenges of achieving transparency in LLMs. The survey will comprehensively examine the construction and application of multimodal knowledge graphs (MMKGs), systematically addressing challenges, progress, and emerging opportunities in this field. It will particularly emphasize the integration of textual and visual data to enhance machine understanding and reasoning capabilities, while identifying open research problems in MMKG development [12, 5, 13]. Additionally, the survey will assess advancements in multimodal language models (MLMs) for evaluating image-text quality, addressing limitations of existing methods such as CLIPScore.

Furthermore, the survey will explore generative AI applications in coding, data analysis, and educational tools within computational social sciences, highlighting its potential to lower barriers in these fields [8]. It aims to analyze the effects of recursive modality changes in generative AI models, focusing on content stability and divergence. Enhancing visual literacy among data consumers by providing tools for accurate visual data interpretation is another significant objective. Constructing a multimodal knowledge graph to systematically organize and analyze ADHD-related information using network-based approaches is also a priority.

Finally, the survey endeavors to provide a comprehensive overview of recent developments in LLMs, addressing knowledge gaps in their training strategies and applications. Through these objectives, the survey aims to present a holistic view of the transformative potential of integrated educational technologies, facilitating the exploration of multimodal datasets to predict concept learnability in machine learning settings and advancing research in multimodal understanding through new dataset formats [9].

1.3 Scope of the Paper

This survey focuses on the integration of educational knowledge graphs, deep learning, large language models, and multimodal approaches within educational environments, emphasizing their transformative potential while acknowledging inherent limitations. The scope is confined to advanced computational frameworks that facilitate the synthesis and application of multimodal educational data, deliberately excluding broader discussions on AI technology adoption and organizational culture [14]. It includes representation, translation, alignment, fusion, and co-learning in multimodal machine learning, while excluding specific applications [15].

The survey encompasses multimodal pretrained models, large models, and instruction tuning strategies, while excluding applications outside text, images, audio, and video domains [4]. It covers frameworks for unifying LLMs and KGs, such as KG-enhanced LLMs, LLM-augmented KGs, and Synergized LLMs + KGs, while excluding unrelated methodologies [16]. The study is limited to a pair of generative AI tools, with a call for further research to explore the effects of different models and configurations [17].

Topics related to the definitions, construction, and application of multimodal knowledge graphs (MMKGs) are included, focusing specifically on the integration of visual and textual data, while excluding areas unrelated to multi-modal knowledge representation and applications [13]. Limitations include the scalability of prompts, as increasing the number of issues to detect leads to longer and more complex prompts that may exceed LLMs' processing capabilities [18].

The scope also focuses on ADHD, its complex symptomatology, and the integration of biological, cognitive, and behavioral data through knowledge graphs [19]. It covers diverse topics related to LLMs, including architectural innovations, training strategies, fine-tuning, and multi-modal LLMs, while excluding smaller language models and niche applications [20]. Additionally, the paper explores the effectiveness of the MLM filter in improving data quality for image-text datasets, particularly in educational contexts [21].

The survey addresses challenges of transparency in LLMs, excluding technical details about specific LLM architectures and training processes [22]. It examines MTL techniques categorized into five key areas: regularization, relationship learning, feature propagation, optimization, and pre-training, while excluding unrelated learning paradigms like single-task learning [7]. The scope further includes

evaluating MLLMs' reasoning capabilities and categorizing multimodal reasoning tasks, while excluding detailed analyses of specific MLLM architectures [11].

Potential limitations include reliance on specific datasets and challenges in generalizing findings across different contexts [23]. The benchmark is designed to address challenges in interpreting intricate visual data and deducing relationships between images and text in multimodal documents [24]. Furthermore, the survey excludes non-computational methods and traditional social science research methodologies [25].

The survey aims to deliver an in-depth analysis of the transformative potential and inherent limitations of integrated educational technologies, particularly focusing on synthesizing multimodal data—audio, text, and visual elements—to enhance educational methodologies. This approach reflects a growing recognition of the need for innovative tools and frameworks that effectively leverage diverse data types to improve teaching and learning experiences in the evolving landscape of online education [26, 27, 3, 28, 15].

1.4 Structure of the Survey

This survey is meticulously structured to provide a comprehensive exploration of integrating advanced computational frameworks within educational environments. It begins with an **Introduction** that outlines the significance of educational knowledge graphs, deep learning, large language models, and multimodal approaches, while delineating the objectives and scope of the survey to establish a foundational understanding of the paper's intentions.

Following the introduction, the survey delves into the **Background and Definitions**, offering precise explanations of key concepts such as educational knowledge graphs, deep learning, large language models, multimodal systems, semantic web, ontology, and data integration, setting the stage for deeper discussions.

The survey then transitions into a detailed examination of **Educational Knowledge Graphs**, discussing their structure, applications, and the challenges faced in knowledge graph completion and alignment. This study investigates the pivotal role of knowledge graphs in organizing and integrating diverse educational data sources, including structured and unstructured information, thereby significantly improving accessibility and interoperability of educational resources across digital platforms and enhancing the effectiveness of educational technologies, such as large language models and multimodal applications [29, 13, 30].

In the section on **Deep Learning in Education**, the focus shifts to the application of deep learning techniques, highlighting key models and methods and their impact on educational data processing and analysis. This is followed by a discussion on **Large Language Models for Education**, which explores their capabilities, applications, and integration with knowledge graphs, addressing both challenges and potential solutions in educational contexts.

The survey further investigates **Multimodal Approaches in Education**, analyzing how integrating multiple data modalities enhances understanding and accessibility of educational content. This section draws on insights from benchmarks that evaluate multimodal embeddings and recursive modality changes, providing a nuanced understanding of multimodal integration.

This exploration delves into the significant role of Semantic Web technologies and Ontology in Education, highlighting how these frameworks enhance the structuring and integration of diverse educational data sources, thereby promoting interoperability and facilitating seamless data sharing across various digital platforms and applications [30, 31, 3, 28, 29]. This is complemented by an examination of **Data Integration in Educational Systems**, addressing challenges and solutions related to integrating diverse educational data sources, with an emphasis on technological solutions and evaluation methods.

Finally, the survey concludes with a **Conclusion** that summarizes key findings, discusses implications of integrating these advanced technologies, and highlights potential future research directions and applications. Throughout the survey, a hierarchical discourse-aware approach is employed to generate summaries and align various modalities, ensuring a coherent and comprehensive narrative [28]. The inclusion of detailed metrics for assessing model performance further enriches the evaluation framework, providing a robust basis for analyzing the transformative potential of these integrated educational technologies [32].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Background and Definitions

The integration of advanced computational frameworks in education hinges on a thorough grasp of educational knowledge graphs, deep learning, large language models (LLMs), multimodal systems, the semantic web, ontologies, and data integration. These concepts are pivotal for leveraging technologies like automated academic paper interpretation and knowledge tracing models, which enhance personalized learning by utilizing diverse data sources [28, 31, 33, 30]. Collectively, they underpin the development of modern educational technologies, enabling effective representation, processing, and integration of varied educational data.

Educational knowledge graphs provide structured representations of relationships and entities, crucial for data interoperability and accessibility, thereby facilitating personalized learning through the linkage of disparate information [19]. The dual quaternion approach to knowledge graph embeddings, which includes translation and rotation models, enriches their semantic depth [34].

Deep learning, characterized by multi-layered neural networks, is employed in education to analyze complex datasets, offering insights into student behavior and learning patterns. Challenges in implementing these models in electronic markets are similarly applicable in educational contexts [14].

LLMs, sophisticated AI models trained on extensive text corpora, excel in generating human-like text. When combined with multimodal data, they form multimodal large language models (MLLMs), capable of processing and synthesizing text, images, and audio, enriching educational content [4]. However, integrating these modalities presents challenges, particularly in evaluating learning outcomes and addressing biases, underscoring the complexity of developing efficient MLLMs [35].

Multimodal systems, which utilize diverse data types, provide a comprehensive understanding of educational content, especially in contexts requiring synthesis of lengthy multimodal documents [36]. Challenges in multimodal summarization, particularly for complex documents like financial reports, highlight the need for advanced LLM capabilities [37]. Effective data integration across modalities, as discussed in multimodal machine learning, is crucial for successful educational content synthesis [15].

The semantic web and ontologies are foundational for structuring and integrating educational data. Semantic web technologies enable the creation of a machine-understandable data web, promoting seamless data sharing and interoperability. Ontologies offer formal knowledge representations within a domain, aiding in the alignment and integration of educational resources [13].

Data integration, which merges data from various sources to create a unified view, is vital in educational systems where data is often dispersed across multiple platforms. Effective strategies enhance the accessibility and usability of educational content, supporting informed decision-making [1]. Challenges of information loss and divergence during recursive modality changes, particularly when transitioning between text and image formats, underscore the need for robust data integration frameworks [17].

These foundational concepts are essential for understanding the subsequent sections of this survey, providing the building blocks for cohesive educational systems that support diverse learning environments. Integrating these technologies into educational frameworks promises to transform the delivery and consumption of educational content, fostering dynamic and personalized learning experiences.

3 Educational Knowledge Graphs

As the landscape of educational technologies continues to evolve, the role of educational knowledge graphs (EKGs) has become increasingly significant. The graphs function as essential frameworks that integrate and represent a wide array of educational data sources—including unstructured and structured text, relational databases, and API data—thereby enhancing learning experiences through improved contextual understanding and accurate feedback facilitated by advanced models like knowledge graphs and large language models. [29, 30]. To fully appreciate the intricacies involved in the development of EKGs, it is essential to examine their structure and construction. This discussion

will delve into the methodologies and frameworks that underpin the creation of EKGs, laying the groundwork for understanding their applications and potential impact in educational contexts.

3.1 Structure and Construction of Educational Knowledge Graphs

The construction of educational knowledge graphs (EKGs) involves a systematic integration of diverse methodologies designed to effectively represent and process educational data. Central to this endeavor is the Cross-Data Knowledge Graph Construction (CD-KG) method, which includes essential steps such as data preprocessing, semantic clustering, automatic cluster labeling, and relation discovery [30]. This approach ensures the organization of data into meaningful clusters, facilitating the discovery of relations that enhance the semantic richness of the knowledge graph.

Figure 2 illustrates the hierarchical structure of Educational Knowledge Graphs (EKGs), highlighting the primary construction methods, their applications and challenges, and the technical barriers faced in their development. The construction methods include CD-KG, multimodal knowledge graph (MMKG) integration, and the DynoNet model. Applications and challenges are represented by multilingual MMNER, ADHD modeling, and curriculum learning, while technical barriers focus on issues such as data isolation, complex data warehousing, and multimodal reasoning.

The development of multimodal knowledge graphs (MMKGs) further exemplifies the complexity of EKG construction, with methods divided into labeling images with symbols (A-MMKG) and grounding symbols to images (N-MMKG) [13]. This dual approach allows for the integration of visual and textual data, enriching the educational content and improving the learning experience. The MMNERD dataset, consisting of 42,908 sentences across four languages and two modalities, underscores the importance of multilingual and multimodal integration in EKGs [38].

Incorporating dynamic elements into EKGs is exemplified by the DynoNet model, which combines knowledge graphs with dynamic embeddings to capture both structured and unstructured elements of dialogue [39]. This dynamic approach is critical for constructing EKGs that can adapt to the evolving nature of educational content, ensuring relevance and utility over time.

The integration of scientific literature and clinical data to represent complex interrelationships, as demonstrated in ADHD research, highlights the potential of EKGs to model intricate educational phenomena [19]. Additionally, the dataset constructed from Wikipedia, incorporating 14,489 entities and 434,675 image-sentence pairs, serves as a testament to the scale and scope of data required for comprehensive EKG construction [40].

Addressing technical barriers such as dataset isolation and incompatibility with data processing tools is crucial for the widespread application of EKGs [41]. The framework for warehousing complex data, which integrates various data types and sources using metadata and XML, provides a robust foundation for EKG construction [1].

Curriculum-aware algorithms, which enhance the benefits of curriculum learning, can inform the structure of EKGs by promoting adaptive learning pathways tailored to individual learner needs [10]. Furthermore, a framework for categorizing multimodal reasoning tasks can guide the integration of various data modalities within EKGs, ensuring a comprehensive representation of educational content [11].

3.2 Applications of Knowledge Graphs in Education

Knowledge graphs (KGs) have become pivotal in transforming educational technologies by structuring and integrating diverse data types to enhance learning experiences. One significant application is in the development of multimodal classification algorithms, which are employed to identify teaching activities during lectures. This facilitates easier access to specific educational content, thereby optimizing learning processes [3]. The integration of KGs in such systems underscores their role in improving the accessibility and organization of educational resources.

In enhancing the performance of knowledge-intensive tasks, KGs play a crucial role, particularly in named entity recognition and question answering. By providing a structured framework for integrating factual knowledge, KGs improve the accuracy and reliability of these tasks, which is essential for educational applications [6]. This enhancement is particularly valuable in environments where the precision of information is critical.

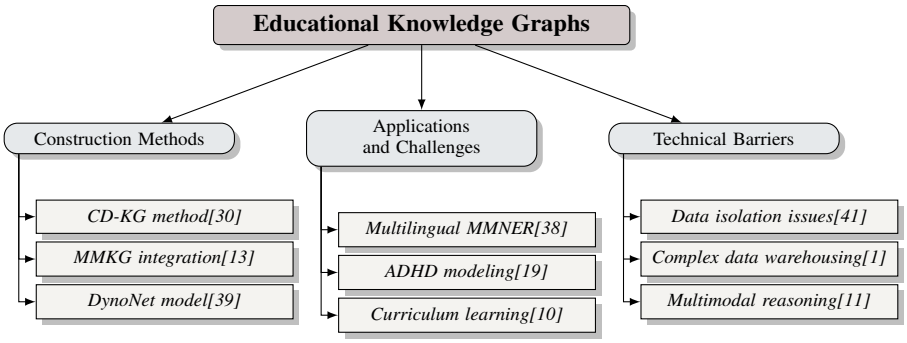


Figure 2: This figure illustrates the hierarchical structure of Educational Knowledge Graphs (EKGs), highlighting the primary construction methods, their applications and challenges, and the technical barriers faced in their development. The construction methods include CD-KG, MMKG integration, and DynoNet. Applications and challenges are represented by multilingual MMNER, ADHD modeling, and curriculum learning. Technical barriers are focused on data isolation, complex data warehousing, and multimodal reasoning.

The Bloom Library exemplifies the utility of KGs in managing multimodal datasets, covering a wide range of languages and including extensive collections of stories, image-caption pairs, and audio files [8]. This extensive dataset facilitates research and development in multilingual and multimodal educational applications, promoting inclusivity and diversity in educational content.

Furthermore, KGs are instrumental in the development of educational assessment tools, such as those evaluating multiple-choice questions (MCQs) aligned with Bloom's taxonomy. By structuring these assessments within a KG framework, educators can ensure that they cover various cognitive levels, thereby supporting comprehensive evaluations of student learning [9]. This application highlights the potential of KGs to enhance the design and effectiveness of educational assessments.

Knowledge graphs are essential to the evolution of educational technologies, as they provide comprehensive frameworks for organizing, integrating, and reasoning about diverse educational data. These graphs facilitate the embedding of both structural and literal information, enhancing the accuracy and relevance of educational applications such as question-answering systems powered by large language models (LLMs). Furthermore, they enable the construction of multi-modal knowledge graphs that incorporate various data types, thereby improving machine understanding and supporting the transition from traditional to digital or blended educational environments. [29, 13, 30]. Their applications span a wide range of educational contexts, from enhancing multimodal reasoning tasks to improving the accuracy and inclusivity of educational content, ultimately supporting more effective and personalized learning experiences.

3.3 Knowledge Graph Embeddings and Enhancements

The evolution of knowledge graph (KG) embeddings has introduced several innovative methodologies aimed at enhancing the representation, reasoning, and application of knowledge graphs, particularly in educational contexts. One such advancement is the DualE framework, which combines rotation and translation operations to improve the representation of relations within KGs [34]. This dual approach enhances the semantic richness and accuracy of relational embeddings, facilitating more precise knowledge representation.

In the realm of educational knowledge graphs, the integration of structural and literal information through joint embedding learning has been shown to significantly enhance the quality of embeddings [29]. By capturing both the structural relationships and the literal context of educational concepts, this approach addresses the challenge of sparse data, thereby improving the efficacy of knowledge tracing and prediction models.

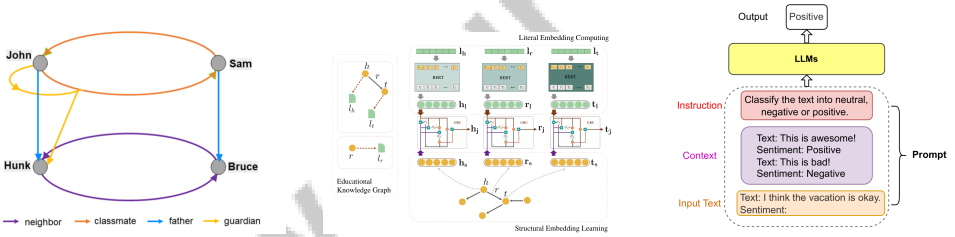
The development of reusable software packages, such as PyKEEN, has further contributed to the field by providing a standardized benchmark for training and evaluating KG embeddings [42]. This tool allows researchers and practitioners to compare different models and configurations, promoting the adoption of best practices and the refinement of embedding techniques across various domains.

Hybrid architectures, such as the MKGformer, introduce a multi-level fusion of visual and textual representations, enabling the completion of multimodal knowledge graph tasks [43]. This hybrid approach leverages the strengths of both modalities to enhance the comprehensiveness and depth of knowledge representations, addressing the limitations of traditional, unimodal embedding techniques.

The mPLUG-Owl3 model employs hyper attention blocks to process long image sequences, contributing to the enhancement of knowledge graphs by integrating detailed visual information [44]. This integration is particularly beneficial in educational settings, where visual data plays a crucial role in conveying complex concepts and facilitating deeper understanding.

Prerequisite-driven approaches, such as the Prerequisite-driven Deep Knowledge Tracing with Constraint modeling (PDKT-C), utilize ordering pairs to model relationships between concepts, effectively regularizing knowledge tracing predictions [33]. This methodology addresses data sparseness by imposing structured constraints, thereby enhancing the predictive accuracy and reliability of educational KGs.

Recent advancements in knowledge graph (KG) embeddings and enhancements highlight the transformative potential of knowledge graphs to deliver more comprehensive, multimodal, and context-aware representations. These developments include the construction of Multi-modal Knowledge Graphs (MMKGs), which integrate diverse data types such as text and images, thereby improving machine understanding of real-world concepts. Additionally, innovations like hyper-relational KG embeddings, which capture complex relationships beyond traditional triplet structures, and the introduction of Structure Guided Multi-modal Pre-trained Transformers for knowledge graph reasoning, further facilitate the effective utilization of structural information in KGs. Collectively, these advancements underscore the ability of KGs to enhance applications in areas such as recommendation systems and visual question answering, positioning them as vital tools for achieving human-level machine intelligence. [5, 45, 13]. By integrating various data modalities and leveraging innovative embedding techniques, KGs continue to facilitate a wide range of applications in educational and other domains, ultimately supporting more effective and personalized learning experiences.



(a) Social Network Diagram[34]

(b) Structural Embedding Learning: A Comprehensive Approach to Knowledge Graph Embedding[29]

(c) Text Classification Model with Prompt and Contextual Information[16]

Figure 3: Examples of Knowledge Graph Embeddings and Enhancements

As shown in Figure 3, In the realm of educational knowledge graphs, knowledge graph embeddings and their enhancements play a pivotal role in advancing the understanding and utility of complex data structures. The provided examples illustrate various facets of this field, showcasing how different models and diagrams contribute to this area of study. The "Social Network Diagram" exemplifies the visualization of relationships within a network, highlighting the interconnectedness of individuals like John, Sam, and Bruce through directed arrows. This visual representation is crucial for understanding social dynamics and interactions within a knowledge graph. Meanwhile, the "Structural Embedding Learning" example offers a comprehensive approach to embedding educational knowledge graphs into a latent space, thereby facilitating the analysis of entities and their relationships. This process is essential for transforming complex graph structures into a format that can be easily manipulated and analyzed. Lastly, the "Text Classification Model with Prompt and Contextual Information" demonstrates the application of knowledge graph enhancements in the realm of natural language processing, where input text is classified based on sentiment, guided by contextual information. Together, these examples underscore the diverse applications and enhancements of knowledge graph embeddings, particularly in educational contexts, thereby enriching the ways in which information is processed and utilized. [**?**]cao2021dual,yao2019jointembeddinglearningeducational,pan2024unifying)

AI-generated, for reference only.

3.4 Challenges in Knowledge Graph Completion and Alignment

The completion and alignment of knowledge graphs (KGs) present numerous challenges, particularly within educational contexts where data complexity and diversity are pronounced. A significant challenge in KG completion is the inference of missing relations between entities, which is crucial for constructing comprehensive and accurate graphs. Current models often struggle to represent multiple relations and complex patterns such as symmetry, inversion, and composition, which are essential for capturing the nuanced relationships within educational data [34].

Entity alignment, which involves identifying pairs of entities from different KGs that represent the same real-world entity, is further complicated by the heterogeneity of data sources and the variability in entity representation. This task is exacerbated by the inadequacy of existing data warehousing methods to handle the diverse structures and sources of complex data, leading to inefficiencies in data integration and alignment [1].

In educational knowledge graphs, the incorporation of the vast array of literals associated with entities poses a challenge, as it is difficult to maintain the relevance of structural information while accommodating these literals [29]. The reliance on sufficient data for effective hypergraph recovery also limits the applicability of certain methods, as such data may not always be available in practical scenarios [46].

The integration of multimodal data introduces additional obstacles, particularly in the context of multimodal knowledge graphs (MKGs). Existing methods often fail to fully leverage the structural information inherent in MKGs, resulting in suboptimal reasoning capabilities [5]. The modality gap between unimodal networks and unlabeled multimodal data further complicates the integration process, as current approaches do not adequately address this gap [2].

Moreover, the need for different model architectures for various tasks and the introduction of noise from irrelevant visual data present significant challenges in the development of hybrid models for MKGs [43]. Existing benchmarks also provide limited functionalities, restricting the comprehensive evaluation of different model components and their combinations [42].

Addressing these challenges requires innovative methodologies and evaluation frameworks capable of accommodating the complexity and diversity of educational data. By addressing the challenges associated with knowledge graph construction and representation, particularly in the context of educational applications, these graphs can be leveraged more effectively to enhance educational technologies. This advancement supports the development of personalized and dynamic learning experiences by integrating rich literal information and multi-modal data, thus improving the machine's understanding of educational content and facilitating tailored learning pathways for students. [29, 13]

4 Deep Learning in Education

Category	Feature	Method
Deep Learning Models and Architectures	Multimodal Processing	SGMPT[5], ODL[41]
	Knowledge Enhancement	PDKT-C[33]
Resource Efficiency and Scalability	Incremental Layer Strategies	LW-FedMML[47]
	Distributed and Parallel Techniques	SINGA[48]
	Cross-Domain Learning	BGNN[49]
Applications in Knowledge Tracing and Intelligent Tutoring	Uncertainty and Reliability	UAG[50]
	Data Integration	MCA[3], MKE[2]

Table 1: This table provides a comprehensive summary of various deep learning methods and their applications in educational contexts. It categorizes these methods into three primary areas: deep learning models and architectures, resource efficiency and scalability, and applications in knowledge tracing and intelligent tutoring. Each category is further detailed with specific features and the corresponding methodologies, highlighting the diverse approaches employed to enhance educational technologies.

The incorporation of deep learning techniques into educational environments has significantly transformed the processing and utilization of educational data. This section delves into the models and architectures that are at the forefront of this transformation, examining their unique contributions to enhancing educational experiences. As illustrated in **??**, the hierarchical structure of deep learning's role in education highlights key models, resource efficiency, and applications in knowledge tracing

and intelligent tutoring. Table 1 presents an organized overview of the deep learning methods and their applications within the educational domain, emphasizing the integration of advanced models, resource-efficient strategies, and intelligent tutoring systems. Additionally, Table 2 presents a comparative overview of various deep learning models, specifically CNNs, RNNs, and GANs, outlining their unique features and applications within the educational sector. The diagram categorizes traditional and advanced architectures, addresses resource challenges, and showcases multimodal integration for personalized learning experiences, thereby providing a comprehensive overview of how these elements interact to foster improved educational outcomes.

4.1 Deep Learning Models and Architectures

Deep learning models have fundamentally advanced educational technologies by offering sophisticated methods for analyzing complex educational data. Traditional architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are complemented by Generative Adversarial Networks (GANs) and Graph Neural Networks (GNNs), each contributing distinct capabilities in educational settings [41].

In educational knowledge tracing, models like Deep Knowledge Tracing (DKT), Dynamic Key-Value Memory Networks (DKVMN), and Self-Attentive Knowledge Tracing (SAKT) excel in predictive performance but often face issues with interpretability, necessitating a balance between accuracy and meaningful insights [10]. Hierarchical graph relations in models such as HGKT enhance both prediction accuracy and interpretability [33].

GNNs address labeling gaps and improve representations of educational interactions, facilitating knowledge transfer and reasoning capabilities. The Structure Guided Multimodal Pretrained Transformer (SGMPT) integrates textual, visual, and structural features to enhance knowledge graph reasoning [5].

Multimodal contexts benefit from models like the Enhanced Vision Model for Text-Rich Content (EVM-TRC), which processes text-rich educational materials through dataset preprocessing and fine-tuning with instructional data [7]. Techniques such as late fusion and early fusion highlight the adaptability of deep learning models to diverse educational data.

The Self-Supervised Multimodal Versatile Networks (MMV) architecture processes visual, audio, and textual data, embedding them into a shared vector space for alignment. The 2M-NER model's success in multimodal named entity recognition underscores the effectiveness of such models [41].

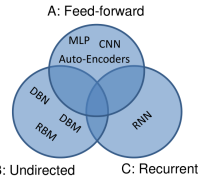
Deep learning has made strides in multimodal reasoning, as seen in models like Google's Gemini 1.5 Flash and OpenAI's GPT-4o, which integrate Chain-of-Thought reasoning and Visual Question Answering to enhance language model accuracy [51, 52, 53]. These advancements underscore deep learning's potential to enrich educational content.

The expansion of deep learning models is enhancing educational technologies, improving the understanding of complex data and supporting personalized learning experiences. These models process vast data across domains, benefiting educators and learners through tailored tools that adapt to individual needs [54, 55]. The emphasis on data quality and diversity is crucial for the performance of Multimodal Large Language Models (MLLMs), paving the way for integrated educational technologies.

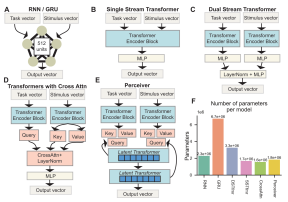
As illustrated in Figure 4, understanding diverse deep learning models and architectures is crucial for leveraging their potential in education. The Venn diagram elucidates the commonalities among Feed-forward, Undirected, and Recurrent models, while the transformer model comparison highlights language understanding nuances. The flowchart categorizes machine learning algorithms, showcasing the complexity and versatility of deep learning in educational contexts [48, 56, 54].

4.2 Resource Efficiency and Scalability

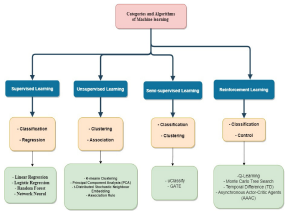
Deploying deep learning models in education faces challenges related to resource efficiency and scalability. High computational demands and complex model implementations limit accessibility [48]. The LW-FedMML framework addresses these issues by reducing memory usage, floating-point operations, and communication costs, exemplifying advancements in resource-efficient federated learning [47].



(a) Venn diagram showing the intersection of three categories: Feed-forward, Undirected, and Recurrent[48]



(b) Comparison of Transformer Models for Language Understanding[56]



(c) Categories and Algorithms of Machine Learning[54]

Figure 4: Examples of Deep Learning Models and Architectures

Deep learning models require large labeled datasets, posing scalability barriers [54]. Graph structures' irregularity complicates traditional methods, necessitating scalable models for large graphs [57]. Interdisciplinary integration and human-machine interaction further challenge scalability in educational technologies [14].

Multimodal large language models (MLLMs) face scalability challenges due to high computational costs and tuning complexities [4]. Bridged-GNNs improve classification performance through knowledge transfer, enhancing resource efficiency [49].

4.3 Applications in Knowledge Tracing and Intelligent Tutoring

Deep learning has significantly advanced knowledge tracing and intelligent tutoring systems, improving predictions and learning outcomes. Integrating retrieval techniques with large language models (LLMs) enhances output accuracy, offering insights into student learning [58]. Combining knowledge graphs with LLMs augments systems by capturing factual knowledge for precise tracing and tutoring [16].

Models like the Structure-Guided Multimodal Pretrained Transformer (SGMPT) and MKGformer leverage multimodal data to enrich educational experiences [5, 43]. Uncertainty-aware reasoning frameworks generate reliable answers from knowledge graphs, benefiting intelligent tutoring systems [50]. The Multimodal Knowledge Extraction (MKE) framework highlights the importance of learning from unlabeled multimodal data [2].

Incorporating audio signals and transcriptions enhances teaching activity classification, supporting effective tutoring interventions [3]. Models like DistilBERT and Wav2Vec2 demonstrate applicability in multilingual contexts, transforming assessment practices in programming education [8, 9].

Deep learning integration in knowledge tracing and intelligent tutoring transforms educational technologies, emphasizing resource efficiency, multimodal integration, and advanced reasoning capabilities. These applications offer solutions for predicting student learning outcomes and customizing instructional strategies, promoting dynamic and personalized learning environments. Innovations such as the multi-modal automated academic paper interpretation system (MMAPIS) and enhanced vision models significantly contribute to adaptive educational frameworks [28, 55, 3].

As illustrated in Figure 5, this figure highlights the key advancements in knowledge tracing and intelligent tutoring systems, emphasizing the integration of deep learning, innovative models, and educational frameworks. The decision tree model tracks skill mastery probabilities, providing insights into learning progression, while the Open Entity Discovery Framework enhances content delivery, creating personalized learning experiences [59, 30]. The role of retrieval techniques, multimodal data, and uncertainty-aware reasoning in enhancing educational technologies is also underscored, reflecting the transformative impact of these advancements.

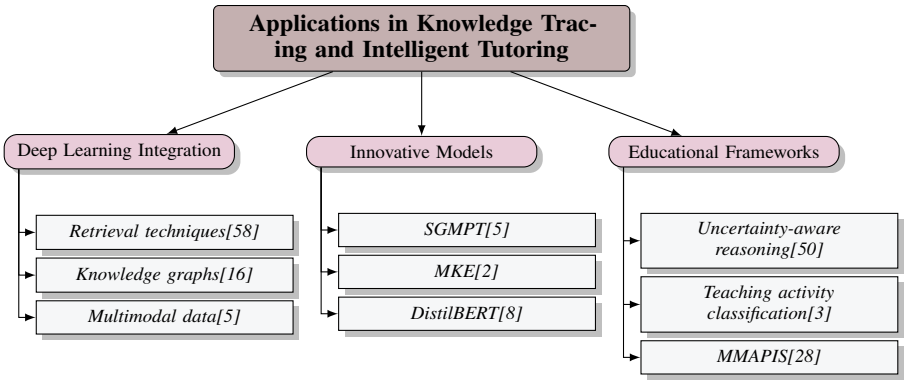


Figure 5: This figure illustrates the key advancements in knowledge tracing and intelligent tutoring systems, highlighting the integration of deep learning, innovative models, and educational frameworks. It emphasizes the role of retrieval techniques, multimodal data, and uncertainty-aware reasoning in enhancing educational technologies.

Feature	Convolutional Neural Networks (CNNs)	Recurrent Neural Networks (RNNs)	Generative Adversarial Networks (GANs)
Model Type	Feed-forward	Recurrent	Generative
Key Feature	Image Processing	Sequence Modeling	Data Generation
Application Domain	Educational Content	Temporal Data	Simulations

Table 2: This table provides a comparative analysis of three prominent deep learning models: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs). It highlights their model types, key features, and specific application domains within the educational context. This comparison underscores the distinct capabilities and contributions of each model to educational technologies.

5 Large Language Models for Education

5.1 Capabilities and Applications of Large Language Models

Large Language Models (LLMs) are integral to enhancing educational technologies, offering improvements in content generation, personalized learning, and assessment methods [31]. They excel in processing and synthesizing diverse data modalities, enriching educational content. For example, mPLUG-Owl3 effectively manages long image sequences, highlighting LLMs' potential in handling complex multimodal tasks [44].

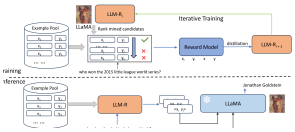
The development of LLMs reflects a progression from statistical models to sophisticated neural architectures, culminating in models with emergent abilities suitable for complex educational problem-solving [60, 61]. The M3Exam benchmark, which includes multilingual and multimodal content, demonstrates LLMs' efficacy in educational assessments [62].

LLMs enhance reasoning tasks through instruction tuning and multi-task learning, as evidenced by studies on their multimodal capabilities [11]. The ICL-D3IE approach illustrates LLMs' proficiency in constructing diverse demonstrations for effective information extraction [63].

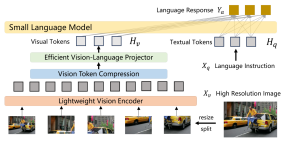
In personalized learning, models such as MarineGPT emphasize the importance of fine-tuning on domain-specific datasets to enhance applicability [64]. Additionally, LLMs facilitate the automation of multiple-choice question generation, reducing educators' workloads and improving assessment quality [9].

Benchmarks evaluating Multimodal Large Language Models (MLLMs) offer insights into the performance of both open-source and closed-source models across various tasks [65]. The exploration of efficient training strategies and the impact of context length further enhance LLM utility in education [20].

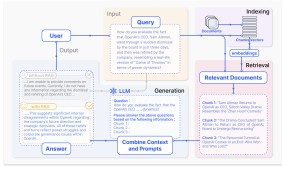
As shown in Figure 7, this figure illustrates the capabilities and applications of Large Language Models (LLMs) in education, multimodal tasks, and advanced model development, highlighting their role in educational enhancement, handling complex multimodal tasks, and advancing model capabilities through emergent abilities and efficient training. LLMs significantly enhance educational



(a) Iterative training of language models for multi-choice questions[66]



(b) Efficient Vision-Language Transformer for Language-Independent Image Captioning[35]



(c) Query Processing Flowchart for Evaluating OpenAI CEO's Dismissal and Rehiring[58]

Figure 6: Examples of Capabilities and Applications of Large Language Models

technology by providing innovative solutions to complex challenges. The iterative training of language models for multiple-choice questions optimizes accuracy in assessments, while the efficient vision-language transformer facilitates language-independent image captioning, promoting inclusivity in educational materials. Moreover, the query processing flowchart demonstrates LLMs' application in analyzing real-world scenarios, fostering critical thinking among learners [66, 35, 58].

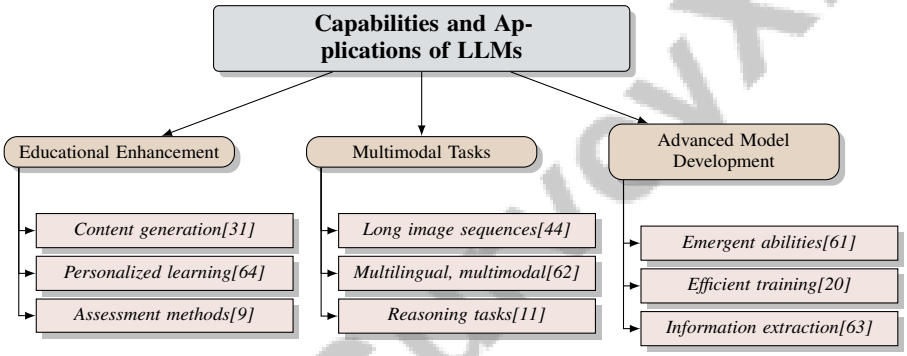


Figure 7: This figure illustrates the capabilities and applications of Large Language Models (LLMs) in education, multimodal tasks, and advanced model development, highlighting their role in educational enhancement, handling complex multimodal tasks, and advancing model capabilities through emergent abilities and efficient training.

5.2 Integration with Knowledge Graphs

Integrating LLMs with knowledge graphs (KGs) significantly advances educational technologies by enhancing interpretative and reasoning capabilities. This synergy leverages KGs' structured knowledge representation to mitigate LLMs' limitations, such as factual recall challenges [31]. Embedding LLMs within KGs fosters a nuanced understanding of complex educational content, improving reasoning processes.

The incorporation of multimodal datasets enriches this integration, facilitating a holistic approach to data interpretation. Wang et al. emphasize the potential of Multimodal Large Language Models (MLLMs) in educational settings, particularly for integrating diverse data modalities [11]. This capability is essential for processing and synthesizing information from various sources in educational contexts.

Innovative frameworks like ICL-D3IE exemplify in-context learning for document information extraction, guiding LLMs in entity label prediction through diverse demonstrations [63]. This method showcases how LLMs can be effectively integrated with KGs to enhance educational information extraction and organization.

LLMs' ability to manage vision-language tasks through prompting, as shown by Hakimov et al., further highlights their versatility in processing multimodal data [67]. This integration is particularly beneficial in educational environments where visual and textual data converge to create richer learning experiences.

The fusion of LLMs with KGs offers significant opportunities for advancing educational technologies. By leveraging advanced reasoning frameworks, multimodal datasets, and dynamic updating capabilities, this integration supports the development of personalized and adaptive learning environments. The evolution of LLMs and multimodal automated academic interpretation systems promises to revolutionize educational methodologies, facilitating tailored learning experiences and enhancing engagement through interactive content while emphasizing critical thinking and ethical considerations in AI use in education [28, 31, 55, 68].

5.3 Challenges and Solutions in Educational Contexts

The deployment of LLMs in education presents challenges related to reliability, interpretability, and ethical implications. A significant concern is the propensity for LLMs to generate hallucinations and inaccuracies, particularly with complex or non-domain-specific queries, as observed in the MarineGPT model's limitations with non-marine topics [64]. This highlights the necessity for robust mechanisms to ensure output reliability in educational applications.

High computational demands for training LLMs pose barriers to widespread adoption in educational contexts [60]. Aligning these models with human values and developing effective evaluation methods for their performance in educational settings further complicate this challenge.

The modality and task gap restricts LLMs' effectiveness in Document Information Extraction (DIE) tasks [63], limiting their ability to leverage diverse data modalities essential for comprehensive educational applications. Additionally, the inability of models to interpret ambiguous language instructions can lead to misinterpretations, affecting educational content delivery [69].

Ethical implications and biases inherent in LLMs also pose significant challenges, as current studies often inadequately address these issues, risking misapplications in education [31]. Addressing these ethical concerns is crucial for responsible LLM deployment.

Furthermore, the reliance on English-only datasets limits the generalizability of findings and restricts LLM applicability across diverse educational contexts [67]. This underscores the need for inclusive datasets to enhance LLM adaptability in global education.

To address these challenges, developing flexible, transparent, and resource-efficient frameworks is essential. Enhancing LLM interpretability and ethical alignment while employing diverse datasets can improve generalizability. This approach tackles existing challenges such as event recall, integration of new information, and domain-specific inaccuracies, thereby increasing LLM output accuracy and relevance in education. Techniques like Retrieval-Augmented Generation (RAG) and Knowledge Graph integration can further enrich LLM performance, transforming educational methodologies and creating dynamic, personalized learning experiences [28, 18, 22, 30].

6 Multimodal Approaches in Education

6.1 Multimodal Representation Techniques

Integrating multimodal data in education requires sophisticated techniques to synthesize information from varied sources, enhancing learning experiences. Innovative methods like mPLUG-Owl3's hyper attention blocks adaptively select visual features based on textual context, optimizing the fusion of visual and textual data [44]. KOSMOS-2 links textual descriptions to visual bounding boxes, facilitating grounded language understanding [70], while AdaptVision's dynamic image partitioning module adjusts visual token input to meet specific image demands [71]. The rate-adaptive coding mechanism adjusts coding rates based on modality importance, optimizing communication efficiency and inference accuracy [72].

These techniques highlight the importance of integrating diverse data modalities to advance educational technologies. Dynamic feature selection, grounded language understanding, adaptive image partitioning, and rate-adaptive coding collectively enhance learning environments' personalization and effectiveness. For instance, CSEAL employs a recurrent neural network to monitor learners' knowledge progression, optimizing the learning experience, while AdaptVision adjusts visual tokens based on image resolution, improving scene comprehension in MLLMs [3, 68, 71]. These advancements are set to transform educational methodologies, delivering tailored learning experiences.

6.2 Multimodal Model Training and Evaluation

Training and evaluating multimodal models are crucial for maximizing diverse data modalities' potential to enhance learning. The OPERA framework reduces reliance on summary tokens, decreasing hallucinations in generated outputs and improving reliability [73]. Attention mechanisms, as seen in transforming text and audio into gloss notations, bridge text-audio gaps, facilitating seamless integration in educational applications [74]. Graph-based approaches like SynerGraph merge textual and visual data to elucidate complex interactions, supporting comprehensive educational technologies [75].

The Mammoth model exemplifies minimalist design in MLLMs, reducing computational overhead while preserving data integrity [76]. Training and evaluation necessitate balancing computational efficiency, data integration, and interpretability. Advanced frameworks enable educational technologies to leverage diverse data from audio, text, and visual elements, fostering dynamic, personalized, and effective learning experiences. This enhances instructional material accessibility and academic content interpretation through automated summarization and categorization systems [28, 3].

6.3 Challenges in Multimodal Integration

Integrating multimodal systems in education presents challenges due to data modalities' complexity and diversity. Dependency on complete modality sets during training can lead to biased predictions and failures when modalities are missing. Modality mismatch, where available modalities at inference differ from training, exacerbates this issue. Recent advancements, such as LLMs with in-context learning, enhance multimodal systems' adaptability and generalizability by utilizing text as a unified semantic space [77, 28, 78, 79].

Existing benchmarks' inadequacy in evaluating MLLMs' reasoning abilities further complicates integration [11]. The inability of filtering methods to capture subtle image-text differences affects performance, compounded by distribution shifts from teacher forcing during fine-tuning [80]. High-quality benchmarks are scarce, failing to measure contextual visual concreteness effectively [23]. Challenges include integrating high-intelligence models during optimization, as seen in VISUAL-O1, and methods like AdaptVision struggling with high-resolution document images [69, 71].

Addressing these challenges is essential for advancing educational technologies, enabling more effective, adaptable tools that harness diverse data sources. This includes automated academic paper interpretation systems and multimodal classification algorithms enhancing online learning material accessibility [28, 55, 3]. Overcoming these obstacles can lead to more dynamic, personalized learning experiences, transforming educational methodologies and outcomes.

6.4 Applications and Case Studies

Multimodal approaches in education, enhanced by LLMs and comprehensive datasets, improve data processing and synthesis. The MS-GQA benchmark highlights subtask dependencies in model selection for multimodal reasoning, enhancing AI systems' robustness in educational contexts [81]. The BDoG framework achieves state-of-the-art results in multimodal reasoning tasks by addressing opinion trivialization and distraction, bridging the gap between text and other modalities [82]. Multimodal Prompt Optimization enhances MLLMs' reasoning abilities, establishing new benchmarks [80].

In federated learning, the 3FM algorithm excels in scenarios with missing modalities, enhancing adaptability in distributed educational environments [83]. The MLM filter enhances educational dataset quality through multimodal approaches [21]. The IMKGA-SM model excels in multimodal knowledge graph link prediction, facilitating educational content analysis and knowledge discovery [84]. MOLBIND aligns multiple molecular modalities, improving zero-shot learning tasks [85].

MERLOT achieves state-of-the-art results in video QA tasks, benefiting educational settings where video content conveys complex concepts [86]. VISUAL-O1 enhances handling ambiguous instructions, showcasing practical multimodal applications [69]. Despite advancements, a gap remains in multimodal models matching human performance in multi-image instruction-following tasks. Addressing this gap is crucial for developing educational technologies that effectively interpret complex instructions, enhancing interactive learning. Future research should integrate multimodal and temporal knowledge into KGLLMs, improving educational content understanding [6].

These applications and case studies illustrate multimodal approaches' transformative potential in education. Leveraging comprehensive datasets and advanced models, educational technologies can provide dynamic, personalized learning experiences, revolutionizing educational methodologies and outcomes. Future research should expand datasets and adapt models to improve performance across diverse languages and contexts [87].

7 Semantic Web and Ontology in Education

7.1 Role of Semantic Web Technologies in Education

Semantic web technologies play a crucial role in managing educational data by enhancing interoperability and accessibility across various resources. These technologies enable seamless integration and retrieval of educational content through structured, machine-readable data, facilitating efficient data management in educational settings [13]. Ontologies, a core component, provide formal representations of knowledge, defining relationships among concepts to create structured datasets. This approach supports the alignment and integration of educational resources and aids in developing personalized learning pathways tailored to individual student needs [19].

The combination of semantic web technologies with educational knowledge graphs enhances data management by organizing and connecting disparate information sources. Knowledge graphs leverage the semantic richness of ontologies to improve accessibility and usability of educational content, enabling intelligent tutoring systems to deliver accurate, context-aware responses to student inquiries [6]. Additionally, these technologies foster adaptive educational environments by allowing real-time updates and retrieval of data, crucial in modern education where knowledge evolves rapidly. By integrating diverse data sources, educational systems can adapt to learners' evolving needs, enhancing digital and blended learning environments' effectiveness. They also support multimodal information incorporation, improving knowledge delivery accuracy and fostering personalized learning experiences [28, 30].

7.2 Ontology and Structuring Educational Data

Ontologies are vital for structuring educational data, providing a formal framework that defines relationships and hierarchies among educational concepts. This structured representation enhances integration and interoperability of diverse resources, facilitating effective data management and retrieval. Leveraging ontologies, educational technologies can establish a semantic layer that organizes complex content, supporting personalized learning pathways and adaptive environments [6].

Integrating ontologies with educational knowledge graphs exemplifies their role in enhancing data accessibility and usability. By organizing and connecting disparate information sources, this integration allows intelligent tutoring systems to deliver accurate and context-aware responses to students [6]. The semantic enrichment provided by ontologies ensures educational data is meaningfully structured, fostering effective interactions between learners and content.

Moreover, ontologies facilitate dynamic updates and retrieval of educational data, crucial in rapidly evolving educational settings. Offering a flexible framework for incorporating new information, ontologies help educational systems remain relevant and responsive to diverse learner needs, particularly in multimodal technologies where coherent data representation is essential [13].

7.3 Interoperability and Data Sharing

Semantic web technologies significantly enhance interoperability and data sharing within educational systems by providing a standardized framework for effective communication among diverse data sources. These technologies create a web of machine-readable and semantically enriched data, facilitating seamless integration and retrieval of educational content across platforms [13]. By employing ontologies and linked data principles, they ensure educational resources are accessible and interoperable, enabling seamless information exchange among various educational applications and systems.

The structuring of educational data through ontologies is pivotal for achieving interoperability, providing a formal knowledge representation that aligns and integrates disparate resources [6]. This

structured approach fosters the development of interoperable educational systems capable of sharing and utilizing data from multiple sources, enhancing accessibility and usability.

Furthermore, semantic web technologies support real-time updating and retrieval of educational data, essential for maintaining relevance and effectiveness in rapidly changing knowledge environments. By facilitating new information integration, these technologies ensure educational content remains current and aligned with learning objectives, particularly in the context of multimodal educational technologies, where robust semantic foundations are crucial for coherent data representation [13].

8 Data Integration in Educational Systems

8.1 Technological Solutions for Data Integration

Integrating diverse educational data sources requires sophisticated technological solutions that manage and synthesize information across various modalities. Hybrid architectures like MKGformer employ multi-level fusion to process visual and textual data, enhancing knowledge graph completion tasks [43]. This method leverages the strengths of different data types, promoting a comprehensive understanding of educational content and facilitating effective data integration.

Frameworks utilizing semantic web technologies and ontologies provide standardized methodologies for data representation, enhancing interoperability and promoting efficient data sharing across platforms. By integrating multimodal data processing, hierarchical classification, and advanced querying techniques, these frameworks enable seamless access to varied data types, fostering improved educational outcomes and collaboration [41, 26, 1, 30, 28]. The use of ontologies to define relationships among educational concepts ensures data accessibility and semantic enrichment, supporting dynamic and personalized learning experiences.

Advanced machine learning models further enhance multimodal data integration by processing and analyzing information from text, images, and other forms. By combining large language models (LLMs) with knowledge graphs (KGs), researchers develop frameworks that synthesize diverse educational data from structured and unstructured sources, improving accuracy and contextual relevance in educational applications. Techniques such as Retrieval-Augmented Generation (RAG) enrich the digital educational landscape by enhancing factual context in user query responses [29, 30].

8.2 Evaluation and Benchmarking

Benchmark	Size	Domain	Task Format	Metric
M3Exam[62]	12,317	Education	Multiple Choice	Accuracy
MS-GQA[81]	8,426	Visual Question Answering	Model Selection	Successful Execution Rate
MMR[78]	126,803	Recommendation Systems	Classification	Accuracy, AUC
MM-Vet[88]	200	Multimodal Evaluation	Open-ended Question Answering	LLM-based evaluator
WanJuan[89]	2,000,000	Multimodal Learning	Model Training	Accuracy, F1-score
MM-NIAH[36]	12,000	Multimodal Document Comprehension	Retrieval	Accuracy, Soft Accuracy
FedMultimodal[90]	10,000	Emotion Recognition	Classification	F1, Accuracy
TRINS[91]	39,153	Visual Question Answering	Visual Captioning	BLEU, CIDEr

Table 3: This table presents a comprehensive overview of prominent benchmarks used in the evaluation of multimodal and multilingual data integration solutions. It details the size, domain, task format, and metrics associated with each benchmark, providing insights into their applicability for assessing large language models and multimodal learning systems. These benchmarks serve as crucial tools for evaluating the effectiveness and precision of educational data processing and integration.

Evaluating and benchmarking data integration solutions in educational systems are crucial for ensuring their effectiveness and efficiency. A comprehensive evaluation framework should address dimensions such as data quality, integration precision, and system scalability to accommodate growing datasets and diverse user needs, particularly concerning complex scholarly content like electronic theses and dissertations (ETDs) and the demand for automated academic paper interpretation tools [28, 26]. Standardized benchmarks provide a consistent basis for comparing different solutions. Table 3 provides a detailed enumeration of standardized benchmarks utilized in the evaluation of

data integration solutions within educational systems, highlighting their significance in assessing the performance of large language models across diverse modalities.

Benchmarks like M3Exam, which encompasses multilingual and multimodal content, offer a robust platform for assessing large language models (LLMs) in handling diverse educational data [62]. Such benchmarks are essential for evaluating data integration performance and insights into processing and synthesizing information from various modalities.

Additionally, benchmarks like MS-GQA, addressing subtask dependencies in multimodal reasoning, highlight the need for nuanced evaluation criteria in educational contexts [81]. These benchmarks ensure effective processing of individual data types while synthesizing complex multimodal interactions.

Beyond standardized benchmarks, evaluation processes should incorporate metrics to assess the interpretability and reliability of integrated data. Techniques such as the OPERA framework, which minimizes hallucinations in generated outputs, exemplify the necessity for evaluation methods focused on the accuracy and trustworthiness of data integration solutions [73].

The evaluation and benchmarking of data integration solutions require a comprehensive approach that integrates standardized benchmarks, interpretability metrics, and performance assessments, similar to psychometric principles used in large language model evaluations. This approach should leverage advanced methodologies, such as multimodal data processing and hierarchical summarization techniques, to enhance the interpretation and usability of academic literature, thereby supporting the integration of diverse educational data sources [28, 92]. By adopting these methods, educational technologies can ensure effective integration and synthesis of diverse data, fostering dynamic, personalized, and effective learning environments.

9 Conclusion

9.1 Challenges and Future Directions

The integration of advanced computational frameworks in education introduces challenges that require targeted research to optimize their effectiveness. A key challenge lies in aligning and integrating diverse data modalities to develop robust educational technologies. Current models often suffer from imbalanced data across modalities, leading to biased representations and diminished robustness in noisy environments. Future research should focus on enhancing multimodal alignment strategies and developing improved evaluation benchmarks for Multimodal Large Language Models (MLLMs) to accurately assess reasoning capabilities across modalities.

Scalability and resource efficiency are critical for deep learning models in educational applications. Despite progress in efficient MLLMs, maintaining performance while managing resource constraints remains a challenge. Although frameworks like SINGA have tackled usability and scalability issues, further advancements are needed, especially for multimedia applications. Future studies could explore layer-wise approaches or other progressive training methodologies to enhance efficiency and performance.

The classification of minority categories poses another challenge due to limited training samples, adversely affecting multimodal classification systems. Future research should aim to improve classification performance for these categories by expanding benchmarks to include additional datasets and multimodal techniques. Refining datasets, expanding modalities, and applying benchmarks to new model architectures are crucial areas for further exploration.

In federated learning, future research should expand the range of supported applications, refine modality fusion methods, and address privacy concerns in multimodal environments. Expanding models like Miko to encompass diverse domains, languages, and behavioral contexts could enhance their applicability.

Addressing limitations in current knowledge tracing models, particularly their inability to capture the nuances of student learning behaviors in dynamic environments, is vital for developing more interpretable and adaptive educational technologies. Future developments should focus on enhancing system capabilities by incorporating broader external knowledge and improving responsiveness in interpretation processes.

Advancements in visual text recognition are necessary to overcome limitations associated with low-resolution images and assumptions of single visual text entities. Future research should develop methods for semi-supervised and unsupervised learning, enhance model interpretability, and address computational challenges in deep learning. Exploring combinatorial optimization techniques could further improve the understanding of relationships between retrieved examples, enhancing multimodal integration.

The integration of knowledge distillation and external knowledge sources could significantly boost the capabilities of multimodal large language models. Future research should refine retrieval processes, evaluate methods across diverse datasets, and address challenges like hallucination in multimodal outputs.

A notable limitation of current methods is their reliance on sampling approaches for processing videos and audios, which can result in information loss. Future research should refine multimodal interaction technologies and explore ways to enhance accessibility and affordability in education. Additionally, improving modality alignment methods, enhancing model transparency, and establishing ethical guidelines for deploying MLLMs in healthcare are crucial areas for future exploration.

9.2 Future Directions and Research Opportunities

Future research should prioritize the exploration of domain-specific Knowledge Graph-enhanced Large Language Models (KGLLMs) to better address the needs of specialized educational applications. This focus can significantly improve the precision and relevance of educational tools, particularly in fields requiring specialized knowledge.

Developing robust frameworks for integrating large language models (LLMs) with knowledge graphs (KGs) presents another promising avenue, with potential applications across various domains. Investigating the effects of utilizing different models in recursive multimodal generative processes and exploring complex scenes could shed light on mechanisms of information loss and enhance the robustness of educational content generation.

Expanding the application of hybrid architectures, such as MKGformer, to additional tasks in natural language processing and information retrieval can enhance capabilities and efficiency in multimodal contexts. This expansion will support more comprehensive data integration strategies, fostering dynamic and personalized learning environments.

In multimodal AI, future research should optimize tuning strategies and broaden the range of modalities integrated into these models. This approach will address existing challenges in multimodal integration and improve the adaptability of educational technologies to diverse learning contexts.

Exploring task-agnostic training and the potential for zero-shot learning in Multi-Task Learning (MTL) represents another significant opportunity. These emerging trends can enhance MTL's flexibility and applicability, supporting versatile educational applications.

Future research should also investigate alternative representations of concept interdependencies, aiming to jointly optimize the extraction of prerequisite relations and knowledge tracing processes. This focus can lead to more accurate and context-aware educational assessments and interventions.

Moreover, optimizing in-context learning processes, particularly in demonstration selection and updating, could extend the framework to other information extraction tasks, enhancing the versatility of educational technologies.

Finally, potential applications of advanced computational frameworks include biodiversity monitoring, marine research collaboration, and educational tools for enhancing understanding of marine biology. These applications underscore the transformative potential of integrating advanced technologies into educational settings, facilitating dynamic, personalized, and effective learning experiences.

References

[1] Jérôme Darmont, Omar Boussaid, Jean-Christian Ralaivao, and Kamel Aouiche. An architecture framework for complex data warehouses, 2007.

[2] Zihui Xue, Sucheng Ren, Zhengqi Gao, and Hang Zhao. Multimodal knowledge expansion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 854–863, 2021.

[3] Oscar Sapena and Eva Onaindia. Multimodal classification of teaching activities from university lecture recordings, 2023.

[4] Soyeon Caren Han, Feiqi Cao, Josiah Poon, and Roberto Navigli. Multimodal large language models and tunings: Vision, language, sensors, audio, and beyond, 2024.

[5] Ke Liang, Sihang Zhou, Yue Liu, Lingyuan Meng, Meng Liu, and Xinwang Liu. Structure guided multi-modal pre-trained transformer for knowledge graph reasoning, 2023.

[6] Linyao Yang, Hongyang Chen, Zhao Li, Xiao Ding, and Xindong Wu. Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling, 2024.

[7] Jun Yu, Yutong Dai, Xiaokang Liu, Jin Huang, Yishan Shen, Ke Zhang, Rong Zhou, Eashan Adhikarla, Wenxuan Ye, Yixin Liu, Zhaoming Kong, Kai Zhang, Yilong Yin, Vinod Namboodiri, Brian D. Davison, Jason H. Moore, and Yong Chen. Unleashing the power of multi-task learning: A comprehensive survey spanning traditional, deep, and pretrained foundation model eras, 2024.

[8] Colin Leong, Joshua Nemecek, Jacob Mansdorfer, Anna Filighera, Abraham Owodunni, and Daniel Whitenack. Bloom library: Multimodal datasets in 300+ languages for a variety of downstream tasks, 2022.

[9] Jacob Doughty, Zipiao Wan, Anishka Bompelli, Jubahed Qayum, Taozhi Wang, Juran Zhang, Yujia Zheng, Aidan Doyle, Pragnya Sridhar, Arav Agarwal, Christopher Bogart, Eric Keylor, Can Kultur, Jaromir Savelka, and Majd Sakr. A comparative study of ai-generated (gpt-4) and human-crafted mcqs in programming education, 2023.

[10] Luca Saglietti, Stefano Sarao Mannelli, and Andrew Saxe. An analytical theory of curriculum learning in teacher-student networks, 2022.

[11] Yiqi Wang, Wentao Chen, Xiaotian Han, Xudong Lin, Haiteng Zhao, Yongfei Liu, Bohan Zhai, Jianbo Yuan, Quanzeng You, and Hongxia Yang. Exploring the reasoning abilities of multimodal large language models (mllms): A comprehensive survey on emerging trends in multimodal reasoning, 2024.

[12] Abhirama Subramanyam Penamakuri and Anand Mishra. Visual text matters: Improving text-kvqa with visual text entity knowledge-aware large multimodal assistant, 2024.

[13] Xiangru Zhu, Zhixu Li, Xiaodan Wang, Xueyao Jiang, Penglei Sun, Xuwu Wang, Yanghua Xiao, and Nicholas Jing Yuan. Multi-modal knowledge graph construction and application: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 36(2):715–735, 2022.

[14] Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning. *Electronic Markets*, 31(3):685–695, 2021.

[15] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2):423–443, 2018.

[16] 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.

[17] Javier Conde, Tobias Cheung, Gonzalo Martínez, Pedro Reviriego, and Rik Sarkar. Analyzing recursiveness in multimodal generative artificial intelligence: Stability or divergence?, 2024.

[18] Leo Yu-Ho Lo and Huamin Qu. How good (or bad) are llms at detecting misleading visualizations?, 2024.

[19] Hakan T. Otal, Stephen V. Faraone, and M. Abdullah Canbaz. A new perspective on adhd research: Knowledge graph construction with llms and network based insights, 2024.

[20] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. A comprehensive overview of large language models. *arXiv preprint arXiv:2307.06435*, 2023.

[21] Weizhi Wang, Khalil Mrini, Linjie Yang, Sateesh Kumar, Yu Tian, Xifeng Yan, and Heng Wang. Finetuned multimodal language models are high-quality image-text data filters, 2024.

[22] Q. Vera Liao and Jennifer Wortman Vaughan. Ai transparency in the age of llms: A human-centered research roadmap, 2023.

[23] Jack Hessel, David Mimno, and Lillian Lee. Quantifying the visual concreteness of words and topics in multimodal datasets, 2018.

[24] Junjie Wang, Yin Zhang, Yatai Ji, Yuxiang Zhang, Chunyang Jiang, Yubo Wang, Kang Zhu, Zekun Wang, Tiezhen Wang, Wenhao Huang, Jie Fu, Bei Chen, Qunshu Lin, Minghao Liu, Ge Zhang, and Wenhu Chen. Pin: A knowledge-intensive dataset for paired and interleaved multimodal documents, 2024.

[25] Yongjun Zhang. Generative ai has lowered the barriers to computational social sciences, 2023.

[26] Muntabir Hasan Choudhury, Lamia Salsabil, William A. Ingram, Edward A. Fox, and Jian Wu. Etdpc: A multimodality framework for classifying pages in electronic theses and dissertations, 2023.

[27] Ruochen Zhao, Hailin Chen, Weishi Wang, Fangkai Jiao, Xuan Long Do, Chengwei Qin, Bosheng Ding, Xiaobao Guo, Minzhi Li, Xingxuan Li, and Shafiq Joty. Retrieving multimodal information for augmented generation: A survey, 2023.

[28] Feng Jiang, Kuang Wang, and Haizhou Li. Bridging research and readers: A multi-modal automated academic papers interpretation system, 2024.

[29] Siyu Yao, Ruijie Wang, Shen Sun, Derui Bu, and Jun Liu. Joint embedding learning of educational knowledge graphs, 2019.

[30] Tuan Bui, Oanh Tran, Phuong Nguyen, Bao Ho, Long Nguyen, Thang Bui, and Tho Quan. Cross-data knowledge graph construction for llm-enabled educational question-answering system: A case study at hcmut, 2024.

[31] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274, 2023.

[32] Jiahui Geng, Yova Kementchedjhieva, Preslav Nakov, and Iryna Gurevych. Multimodal large language models to support real-world fact-checking, 2024.

[33] Penghe Chen, Yu Lu, Vincent W Zheng, and Yang Pian. Prerequisite-driven deep knowledge tracing. In *2018 IEEE international conference on data mining (ICDM)*, pages 39–48. IEEE, 2018.

[34] Zongsheng Cao, Qianqian Xu, Zhiyong Yang, Xiaochun Cao, and Qingming Huang. Dual quaternion knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 6894–6902, 2021.

[35] Yizhang Jin, Jian Li, Yexin Liu, Tianjun Gu, Kai Wu, Zhengkai Jiang, Muyang He, Bo Zhao, Xin Tan, Zhenye Gan, Yabiao Wang, Chengjie Wang, and Lizhuang Ma. Efficient multimodal large language models: A survey, 2024.

[36] Weiyun Wang, Shuibo Zhang, Yiming Ren, Yuchen Duan, Tiantong Li, Shuo Liu, Mengkang Hu, Zhe Chen, Kaipeng Zhang, Lewei Lu, et al. Needle in a multimodal haystack. *Advances in Neural Information Processing Systems*, 37:20540–20565, 2024.

[37] Tianyu Cao, Natraj Raman, Danial Dervovic, and Chenhao Tan. Characterizing multimodal long-form summarization: A case study on financial reports, 2024.

[38] Dongsheng Wang, Xiaoqin Feng, Zeming Liu, and Chuan Wang. 2m-ner: Contrastive learning for multilingual and multimodal ner with language and modal fusion, 2024.

[39] He He, Anusha Balakrishnan, Mihail Eric, and Percy Liang. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings. *arXiv preprint arXiv:1704.07130*, 2017.

[40] Yangning Li, Tingwei Lu, Yinghui Li, Tianyu Yu, Shulin Huang, Hai-Tao Zheng, Rui Zhang, and Jun Yuan. Mesed: A multi-modal entity set expansion dataset with fine-grained semantic classes and hard negative entities, 2023.

[41] Conghui He, Wei Li, Zhenjiang Jin, Chao Xu, Bin Wang, and Dahua Lin. Opendatalab: Empowering general artificial intelligence with open datasets, 2024.

[42] Mehdi Ali, Max Berrendorf, Charles Tapley Hoyt, Laurent Vermue, Sahand Sharifzadeh, Volker Tresp, and Jens Lehmann. Pykeen 1.0: a python library for training and evaluating knowledge graph embeddings. *Journal of Machine Learning Research*, 22(82):1–6, 2021.

[43] Xiang Chen, Ningyu Zhang, Lei Li, Shumin Deng, Chuanqi Tan, Changliang Xu, Fei Huang, Luo Si, and Huajun Chen. Hybrid transformer with multi-level fusion for multimodal knowledge graph completion, 2023.

[44] Jiabo Ye, Haiyang Xu, Haowei Liu, Anwen Hu, Ming Yan, Qi Qian, Ji Zhang, Fei Huang, and Jingren Zhou. mplug-owl3: Towards long image-sequence understanding in multi-modal large language models. In *The Thirteenth International Conference on Learning Representations*, 2024.

[45] Paolo Rosso, Dingqi Yang, and Philippe Cudré-Mauroux. Beyond triplets: hyper-relational knowledge graph embedding for link prediction. In *Proceedings of the web conference 2020*, pages 1885–1896, 2020.

[46] Yang Chen, Cong Fang, Zhouchen Lin, and Bing Liu. Relational learning in pre-trained models: A theory from hypergraph recovery perspective, 2024.

[47] Ye Lin Tun, Chu Myaet Thwal, Minh N. H. Nguyen, and Choong Seon Hong. Resource-efficient federated multimodal learning via layer-wise and progressive training, 2024.

[48] Wei Wang, Gang Chen, Haibo Chen, Tien Tuan Anh Dinh, Jinyang Gao, Beng Chin Ooi, Kian-Lee Tan, and Sheng Wang. Deep learning at scale and at ease, 2016.

[49] Wendong Bi, Xueqi Cheng, Bingbing Xu, Xiaoqian Sun, Li Xu, and Huawei Shen. Bridged-gnn: Knowledge bridge learning for effective knowledge transfer, 2023.

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

[51] Zhuosheng Zhang, Aston Zhang, Mu Li, Hai Zhao, George Karypis, and Alex Smola. Multimodal chain-of-thought reasoning in language models. *arXiv preprint arXiv:2302.00923*, 2023.

[52] Abhinav Arun, Dipendra Singh Mal, Mehul Soni, and Tomohiro Sawada. Towards a unified multimodal reasoning framework, 2023.

[53] Pan Lu, Swaroop Mishra, Tanglin Xia, Liang Qiu, Kai-Wei Chang, Song-Chun Zhu, Oyvind Tafjord, Peter Clark, and Ashwin Kalyan. Learn to explain: Multimodal reasoning via thought chains for science question answering. *Advances in Neural Information Processing Systems*, 35:2507–2521, 2022.

[54] Mohammad Mustafa Taye. Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Computers*, 12(5):91, 2023.

[55] Adithya TG, Adithya SK, Abhinav R Bharadwaj, Abhiram HA, and Surabhi Narayan. Enhancing vision models for text-heavy content understanding and interaction, 2024.

[56] Takuya Ito, Soham Dan, Mattia Rigotti, James Kozloski, and Murray Campbell. On the generalization capacity of neural networks during generic multimodal reasoning, 2024.

[57] Ziwei Zhang, Peng Cui, and Wenwu Zhu. Deep learning on graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 34(1):249–270, 2020.

[58] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2023.

[59] Sein Minn, Jill-Jenn Vie, Koh Takeuchi, Hisashi Kashima, and Feida Zhu. Interpretable knowledge tracing: Simple and efficient student modeling with causal relations, 2021.

[60] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.

[61] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*, 2022.

[62] Wenxuan Zhang, Sharifah Mahani Aljunied, Chang Gao, Yew Ken Chia, and Lidong Bing. M3exam: A multilingual, multimodal, multilevel benchmark for examining large language models, 2023.

[63] Jiabang He, Lei Wang, Yi Hu, Ning Liu, Hui Liu, Xing Xu, and Heng Tao Shen. Icl-d3ie: In-context learning with diverse demonstrations updating for document information extraction, 2023.

[64] Ziqiang Zheng, Jipeng Zhang, Tuan-Anh Vu, Shizhe Diao, Yue Him Wong Tim, and Sai-Kit Yeung. Marinegpt: Unlocking secrets of ocean to the public, 2023.

[65] Chenhao Zhang, Xi Feng, Yuelin Bai, Xinrun Du, Jinchang Hou, Kaixin Deng, Guangzeng Han, Qinrui Li, Bingli Wang, Jiaheng Liu, Xingwei Qu, Yifei Zhang, Qixuan Zhao, Yiming Liang, Ziqiang Liu, Feiteng Fang, Min Yang, Wenhao Huang, Chenghua Lin, Ge Zhang, and Shiwen Ni. Can mllms understand the deep implication behind chinese images?, 2024.

[66] Liang Wang, Nan Yang, and Furu Wei. Learning to retrieve in-context examples for large language models. *arXiv preprint arXiv:2307.07164*, 2023.

[67] Sherzod Hakimov and David Schlangen. Images in language space: Exploring the suitability of large language models for vision language tasks, 2023.

[68] Qi Liu, Shiwei Tong, Chuanren Liu, Hongke Zhao, Enhong Chen, Haiping Ma, and Shijin Wang. Exploiting cognitive structure for adaptive learning. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 627–635, 2019.

[69] Minheng Ni, Yutao Fan, Lei Zhang, and Wangmeng Zuo. Visual-o1: Understanding ambiguous instructions via multi-modal multi-turn chain-of-thoughts reasoning, 2024.

[70] Zhiliang Peng, Wenhui Wang, Li Dong, Yaru Hao, Shaohan Huang, Shuming Ma, and Furu Wei. Kosmos-2: Grounding multimodal large language models to the world. *arXiv preprint arXiv:2306.14824*, 2023.

[71] Yonghui Wang, Wengang Zhou, Hao Feng, and Houqiang Li. Adaptvision: Dynamic input scaling in mllms for versatile scene understanding, 2024.

[72] Yangshuo He, Guanding Yu, and Yunlong Cai. Rate-adaptive coding mechanism for semantic communications with multi-modal data, 2023.

[73] Qidong Huang, Xiaoyi Dong, Pan Zhang, Bin Wang, Conghui He, Jiaqi Wang, Dahua Lin, Weiming Zhang, and Nenghai Yu. Opera: Alleviating hallucination in multi-modal large language models via over-trust penalty and retrospection-allocation, 2024.

[74] Sen Fang, Sizhou Chen, Yalin Feng, Xiaofeng Zhang, and Teik Toe Teoh. Bridging the gap between text, audio, image, and any sequence: A novel approach using gloss-based annotation, 2024.

[75] Mert Burabak and Tevfik Aytekin. Synergraph: An integrated graph convolution network for multimodal recommendation, 2024.

[76] Qi She, Junwen Pan, Xin Wan, Rui Zhang, Dawei Lu, and Kai Huang. Mammothmoda: Multi-modal large language model, 2024.

[77] Yun-Da Tsai, Ting-Yu Yen, Pei-Fu Guo, Zhe-Yan Li, and Shou-De Lin. Text-centric alignment for multi-modality learning, 2024.

[78] Maciej Pawłowski, Anna Wróblewska, and Sylwia Sysko-Romańczuk. Does a technique for building multimodal representation matter? – comparative analysis, 2022.

[79] Yang Luo, Zangwei Zheng, Zirui Zhu, and Yang You. How does the textual information affect the retrieval of multimodal in-context learning?, 2024.

[80] Weiyun Wang, Zhe Chen, Wenhai Wang, Yue Cao, Yangzhou Liu, Zhangwei Gao, Jinguo Zhu, Xizhou Zhu, Lewei Lu, Yu Qiao, and Jifeng Dai. Enhancing the reasoning ability of multimodal large language models via mixed preference optimization, 2024.

[81] Xiangyan Liu, Rongxue Li, Wei Ji, and Tao Lin. Towards robust multi-modal reasoning via model selection. *arXiv preprint arXiv:2310.08446*, 2023.

[82] Changmeng Zheng, Dayong Liang, Wengyu Zhang, Xiao-Yong Wei, Tat-Seng Chua, and Qing Li. A picture is worth a graph: A blueprint debate paradigm for multimodal reasoning, 2024.

[83] Minh Tran, Roochi Shah, and Zejun Gong. 3fm: Multi-modal meta-learning for federated tasks, 2023.

[84] Yilin Wen, Biao Luo, and Yuqian Zhao. Imkga-sm: Interpretable multimodal knowledge graph answer prediction via sequence modeling, 2023.

[85] Teng Xiao, Chao Cui, Huaisheng Zhu, and Vasant G. Honavar. Molbind: Multimodal alignment of language, molecules, and proteins, 2024.

[86] Rowan Zellers, Ximing Lu, Jack Hessel, Youngjae Yu, Jae Sung Park, Jize Cao, Ali Farhadi, and Yejin Choi. Merlot: Multimodal neural script knowledge models, 2021.

[87] Aozhu Chen, Ziyuan Wang, Chengbo Dong, Kaibin Tian, Ruixiang Zhao, Xun Liang, Zhanhui Kang, and Xirong Li. Chinaopen: A dataset for open-world multimodal learning, 2023.

[88] Weihao Yu, Zhengyuan Yang, Linjie Li, Jianfeng Wang, Kevin Lin, Zicheng Liu, Xinchao Wang, and Lijuan Wang. Mm-vet: Evaluating large multimodal models for integrated capabilities. *arXiv preprint arXiv:2308.02490*, 2023.

[89] Conghui He, Zhenjiang Jin, Chao Xu, Jiantao Qiu, Bin Wang, Wei Li, Hang Yan, Jiaqi Wang, and Dahua Lin. Wanjuan: A comprehensive multimodal dataset for advancing english and chinese large models, 2023.

[90] Tiantian Feng, Digbalay Bose, Tuo Zhang, Rajat Hebbar, Anil Ramakrishna, Rahul Gupta, Mi Zhang, Salman Avestimehr, and Shrikanth Narayanan. Fedmultimodal: A benchmark for multimodal federated learning, 2023.

[91] Ruiyi Zhang, Yanzhe Zhang, Jian Chen, Yufan Zhou, Jiuxiang Gu, Changyou Chen, and Tong Sun. Trins: Towards multimodal language models that can read, 2024.

[92] Qixiang Fang, Daniel L. Oberski, and Dong Nguyen. Patch! psychometrics-assisted benchmarking of large language models: A case study of proficiency in 8th grade mathematics, 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.