Building Trustworthy Knowledge: Data Quality and Validation Protocols for Educational Knowledge Graphs

Ana Pasarić, Danijela Jakšić, Patrizia Poščić
University of Rijeka
Faculty of Informatics and Digital Technologies
Radmile Matejčić 2, 51000 Rijeka, Croatia
ana.pasaric1@student.uniri.hr, {danijela.jaksic, patrizia}@inf.uniri.hr

Abstract. *Todays education struggles with rapid technological change, the need for interdisciplinary learning, and demands for personalization. Knowledge Graphs (KGs) offer a crucial AI solution, organizing vast information and enabling advanced reasoning. In education, KGs are transformative, forming the foundation of smart learning by enhancing personalized learning, curriculum development, and assessment. However, smart learning's effectiveness relies entirely on KG data quality. The diverse and dynamic nature of educational data highlights a critical need for rigorous data validation. Populating KGs often faces a knowledge acquisition bottleneck, impeding consistent, high-quality information. This review therefore assesses current educational data management methods, evaluating existing protocols to determine if new or improved standards are essential for ensuring a continuous flow of accurate knowledge into KGs, thereby fostering more effective intelligent educational systems.*

Keywords: knowledge graphs, education, AI

1 Introduction

The digital age is defined by an unprecedented explosion of data, with an estimated 149 zettabytes created, captured, copied, and consumed globally in 2024, a figure projected to surge to 394 zettabytes by the end of 2028[1]. This exponential growth, largely fuelled by the proliferation of Internet of Things (IoT) devices, the demand for real-time processing, and widespread cloud adoption, underscores a critical need for advanced information organization (Qin et al., 2020). A substantial portion of this burgeoning data, estimated at 80% remains unstructured, presenting both significant challenges and opportunities for effective utilization[2].

The modern education landscape faces complex challenges driven by rapid technological advancements, the growing need for interdisciplinary connections, and an escalating demand for personalized learning experiences (Fettach et al., 2022, Hu et al., 2023). Keeping pace requires continuous adaptation from educators and students, demanding ongoing skill development, frequent curriculum updates, and significant investment in new resources (Paneque et al., 2023, Yang Y. et al., 2024). As technology and industry standards shift, educators must promptly adjust content, develop innovative tech-driven approaches, and integrate ethical considerations (Abu-Rasheed et al., 2025, Fan et al., 2024, Yan et al., 2024). In response to these pervasive data and educational complexities, Knowledge Graphs (KGs) have emerged as indispensable structures within Artificial Intelligence (AI), expert systems, data management, and intelligent applications (Abu-Salih & Alotaibi, 2024, Deng et al., 2021).

KGs serve as structured repositories that formalize domain expertise and facilitate sophisticated reasoning, playing a critical role across diverse fields such as medicine, e-commerce, and, particularly, education (Qu et al., 2024). Within the educational sector, KGs underpin advancements in personalized learning, curriculum design, content recommendation systems, innovative teaching methods and muti-dimensional assessment driving the increasing global attention given to smart learning (Li et. al., 2022, Chen Z. et al., 2024).

Smart learning, or smart education, has rapidly evolved since the 2010s, shifting from traditional paradigms to emphasize technology-embedded, self-directed, motivated, adaptive, and resource-rich learning experiences (Qin et al., 2020, Qu et al., 2024). Its key features include being learner-centric, personalized, adaptive, interactive, collaborative, context-aware, and ubiquitous, enabling anytime,

[1]https://www.statista.com/statistics/871513/worldwide-data-created/

[2]https://www.forbes.com/sites/traceywelsonrossman/2019/01/28/i-see-data-forge-ai-mines-the-worlds-unstructured-data/?ctpv=searchpage

anywhere knowledge acquisition via smart devices (Mzwri & Turcsányi-Szabo, 2023).

However, smart learning's effectiveness critically depends on the quality of knowledge within KGs. This presents a significant challenge: knowledge acquisition (Tian et al. 2022). Populating KGs, whether through laborious manual encoding or automated extraction from diverse data, often leads to a knowledge acquisition bottleneck (Paneque et al., 2023). Therefore, robust protocols are essential to manage the quality, consistency, and validity of knowledge within KGs, as these directly impact intelligent system reliability.

Our objective is to evaluate current methodologies' efficacy in addressing the unique data challenges of educational contexts. This review will assess if existing protocols adequately handle the scale, diversity, and dynamic nature of educational data, and determine the need for new or improved protocols to ensure the consistent acquisition of high-quality, verifiable knowledge within educational KGs. This work aims to contribute to the development of more accurate, adaptive, and impactful intelligent educational systems.

2 Methodology

This systematic literature review (SLR) critically evaluates how well current methodologies address the unique data challenges in education. Our main goal is to see if existing approaches are robust enough to handle the sheer scale, diversity, and dynamic nature of educational data. We're also assessing whether new or significantly improved protocols are needed to ensure we consistently acquire high-quality, verifiable knowledge within educational Knowledge Graphs (KGs). Ultimately, this work aims to contribute to more accurate, adaptive, and impactful intelligent educational systems.

To gather relevant articles, we initiated an extensive search on Scopus, limiting results to English-language publications (2020-2025). Our initial keyword query, (knowledge AND graph AND in AND education), yielded approximately 22 articles. Recognizing that a single database search can limit comprehensive coverage and introduce bias, we expanded our screening process to include Google Scholar. For instance, the exact phrase allintitle: "knowledge graph education" was utilized to specifically target articles where "knowledge graph" and "education" appeared together in the title, thereby enhancing the precision and relevance of the retrieved results. This expansion aimed to cast a wider net for relevant literature and ensure a more robust and diverse dataset. By thoroughly reviewing citations and references from the initially retrieved papers on Google Scholar, we added another 75 articles to our dataset. This preliminary phase resulted in a total of 97 records.

A rigorous screening process followed to eliminate redundant or irrelevant articles. We carefully examined the titles and abstracts of all 97 records against our pre-defined inclusion criteria, leading to the exclusion of 65 records. In the final stage of the literature review, we identified 32 papers that met our qualifications for inclusion in this comprehensive review of Knowledge Graph use in an educational context.

3 The Expansive Role of Knowledge Graphs in Modern Education

Knowledge graphs (KGs) are fundamentally changing education, proving themselves to be indispensable tools for teaching and learning. They are increasingly employed to enhance personalized learning experiences, optimize curriculum design, facilitate concept mapping, and power educational content recommendation systems. KGs further support diverse pedagogical approaches, including personalized and collaborative learning, by recommending suitable activities and resources, thereby improving group knowledge building (Chen L. et al., 2024, Edge et al., 2024, Gligorea et al., 2023).

KGs are highly versatile in education, applied across disciplines from history to medicine, and spanning all levels from primary to higher education (Agrawal et al., 2022, Fan et al., 2024, Hou et al., 2025, Qu et al., 2024). Their utility includes recommending personalized learning resources, predicting performance, generating questions, retrieving materials, and improving course management and curriculum coherence (Bernasconi et al., 2025, Deng et al, 2021, Qi et al., 2025). KGs dynamically enhance education by acting as continuously updated information hubs. They provide real-time access to current trends and foster interdisciplinary connections in course design. Crucially, KGs facilitate personalized learning by analyzing student interactions and feedback to tailor content, optimize materials, and boost engagement and innovative thinking (Qin et al., 2020, Fan et al., 2024, Hu et al., 2023, Li et al., 2022). Ultimately, KGs are proving transformative across diverse educational domains. They provide structured, interconnected knowledge that addresses key challenges, with applications ranging from data integration in Learning Management Systems (LMSs) to specialized pedagogical approaches in subjects like history, computer science, and creative arts (Mzwri & Turcsányi-Szabo, 2023, Xue, 2023, 32).

A core similarity among many proposed KG-based educational models is their focus on semantic enrichment and the creation of structured representations of knowledge. For instance, the e-LION model (Paneque et al., 2023) addresses the challenge of heterogeneous LMS data by transforming raw information into a standardized RDF knowledge graph, enabling richer data analysis and supporting AI algorithms to identify implicit student interaction patterns. Similarly, in primary and secondary history education, a proposed model (Yu et al., 2021) uses a semantically enriched KG to define historical events by attributes like time and place, fostering deeper understanding and cross-disciplinary exploration. The cybersecurity education model (Deng et al., 2021) also relies on a semantically enriched KG built from public information (Wikipedia articles, YouTube videos) and advanced word embedding to guide students through problem-based learning scenarios. This shared emphasis on building well-defined, interconnected knowledge structures underpins their ability to provide context-aware and personalized learning experiences.

The scope of these models extends across various learning experiences and addresses different educational gaps. While e-LION focuses on administrative support and data analytics within the broader e-learning ecosystem (Paneque et al., 2023), other models directly enhance student learning outcomes. The history education model (Yu et al., 2021) aims for improved retention and deeper understanding of complex historical concepts. The cybersecurity model (Deng et al., 2021) targets practical problem-solving skills and increased engagement in a highly specialized technical field. Furthermore, the MEduKG model (Li et al., 2022) pushes the boundaries by constructing a multi-modal educational knowledge graph, integrating not just text but also images and teacher speech to create a more comprehensive and engaging learning experience, addressing the limitation of predominantly text-based KGs. This multi-modal approach is a significant step towards covering the richness of real classroom interactions that other models, focused primarily on textual data or structured content, might not fully capture. Another model (Hu et al., 2023) for professional introduction courses exemplifies systematizing complex content into hierarchical knowledge trees and then KGs, demonstrating applicability beyond traditional academic subjects to vocational training.

Finally, while proposed models show significant advancements, they also indicate opportunities for further integration. KG construction varies, from multi-agent crawling in cybersecurity (Deng et al., 2021) to combining machine learning and human expertise for text-based KGs (Agrawal et al., 2022). The integration of Large Language Models (LLMs) with KGs, exemplified by Graphusion for NLP education (Yang R. et al., 2024) and IntelliChain for Socratic math teaching (Qi et al., 2025), powerfully enhances content generation and dialogue-based learning, an area some earlier models underutilized. Similarly, models like Optimal Knowledge Component Extracting (OKCE) (Choi et al., 2023) and KG-PLPPM (Hou et al., 2025) for personalized learning paths complement existing KGs by focusing on knowledge component relationships and addressing specific learning gaps.

The overarching theme is clear: KGs, by offering versatile applications from data integration to highly personalized and multi-modal learning experiences, are continuously expanding their footprint across the educational landscape.

4 Data Quality and Validation in Educational KGs: Addressing Challenges of Sparse and Incomplete Data

While KGs are undeniably powerful tools for revolutionizing education, their effective implementation faces several key challenges. These include insufficient data for construction, difficulties in extending applications across diverse subject areas, and the critical need for more comprehensive user feedback and evaluation processes.

One primary challenge lies in KG construction, which typically involves both top-down (domain modeling and rule definition) and bottom-up (knowledge acquisition and extraction) approaches. Building high-quality educational KGs necessitates a huge quantity of clean, prepared course-related data (Agrawal et al., 2023, Qin et al., 2020, Elkaimbillah et al., 2021). However, acquiring and integrating this data, especially from diverse sources like unstructured texts, can be complex and resource-intensive. Existing methods for knowledge extraction heavily rely on Natural Language Processing (NLP), text mining, and machine learning techniques, including Conditional Random Fields (CRF), Support Vector Machines (SVM), Bi-directional Long Short-Term Memory (BiLSTM) networks, and Hidden Markov Models (HMM) (Abu-Salih, 2021, Bernasconi et al., 2025, Elkaimbillah et al., 2021, Fettach et al. 2022). Despite these advanced techniques, the sheer volume and heterogeneity of educational content often result in insufficient data for fully comprehensive and automated KG development. Furthermore, while generic KGs like Wikidata provide broad encyclopedic content, developing domain-specific KGs tailored for particular educational problems, requires specialized data and expertise, which might not always be readily available or easily scalable across numerous subject areas (Abu-Salih & Alotaibi, 2024, Agrawal et al., 2023, Chen Z. et al., 2024).

Another significant challenge is extending KG applications across diverse subject areas and ensuring their quality and reliability. While knowledge graph embeddings (KGEs) like TransE, TransH, and TransR effectively transform entities and relationships into low-dimensional vector representations for numerical analysis, ensuring their relevance and accuracy across subjects like history, mathematics, or art and design remains complex (Bernasconi et al., 2025, Elkaimbillah et al. 2021, Yu et al., 2021). More critically, the implementation of robust data check and validation protocols is paramount for controlling knowledge input into KBs. These protocols encompass syntactic and structural validation (e.g., OWL ontology axioms, RDF graph well-formedness), semantic consistency checks (e.g., logical consistency, redundancy detection), and plausibility checks (leveraging heuristics or external knowledge) (Qu et al., 2024, Weichselbraun et al., 2022, Yang R. et al., 2024).

Table 1. Educational Knowledge Graphs: Usage, Technical Means, and Evaluation Protocols

Ref.	KG usage	Technical means	Evaluation protocol
(22)	Smart education applications	NER, BILSTM-CRF, Word2vec, Neo4j	Teaching quality and student learning outcomes
(32)	Personalized learning	Semi-automatic annotation and extraction	N/A
(10)	Problem-based learning (PBL)	XLNet, KG, interactive GUI	Student learning outtcome
(31)	Recommender systems	LightGCN	Recall@K, NDCG@K
(4)	Personalized learning using chatbot for question answering	Bi-LSTM-CRF, BERT), KG, SPARQL	Completeness consistency, accuracy, application impact
(18)	Intelligent tutoring system	EduBERT, BiLSTM-CRF, speech-fusion	Precision, recall, F1-score f
(17)	Learning Resources and Guiding Recommendations	TransE Neo4j	Sudent learning outcomes
(27)	intelligent tutoring systems in e-learning	OWL 2 DL, RDF, SPARQL, KG	Accuracy, precision, recall, F1 and interpretability
(20)	Inteligent tutoring systems in e-learning	OWL2, RDF, SWRL	Accuracy, precision, recall, F1 and interpretability
(9)	Personal learning path reccommendation	LASSO, Random Forest, DKT, LSTM	AUC
(30)	Knowledge recommendation	MKR	LRS, LKG, LREG, L2
(29)	TutorQA (Question Answering)	Graphusion	TutorQA benchmark
(7)	Collaborative problem solving (CPS)	Knowledge extraction & structural analysis	Student knowledge identification, visualization, and activation patterns
(13)	Learning resource, curriculum design and educaional assesment	Node feature extraction method	Curriculum, personalized learning, teaching methods, assessment
(6)	Personalized learning	IQASs	Multi-faceted quantitative & user-centered qualitative assessment
(1)	Personal learning path reccommendation	N/A	Relevance, accuracy, ADC, Modularity
(21)	Personalized learning via chain-of-thought dialogue	LLMs, knowledge graphs, multi-agent system	Teaching outcomes, pedagogical dialogue, alignment with objectives
(16)	Personal learning path reccommendation	TransR Euclidean distance V-DINA	Learning path recommendation accuracy via KG semantic and user history score similarity
(11)	Global QA (Question Answering)	GraphRAG	RAG answer evaluation metrics

Furthermore, process-oriented controls like comprehensive review and approval workflows, versioning, access control, and robust feedback mechanisms are essential for continuous quality enhancement. The emergence of "hallucination" in Large Language Models (LLMs) underscores this challenge, driving research into KG augmentation strategies—Knowledge-Aware Inference, Learning, and Validation—to enhance LLM performance and ensure factual accuracy by providing external, structured knowledge and robust fact-checking (Abu-Salih, 2021, Agrawal et al., 2023, Qi et al., 2025). However, despite these advancements, ensuring consistent application and comprehensive coverage of these validation protocols across every unique educational context and domain still represents a considerable implementation hurdle.

The effective implementation of KGs in education faces persistent challenges related to data quality, including issues of completeness, consistency, and noise. A significant hurdle is the management of sparse and incomplete data within many educational KGs, coupled with the persistent challenge of ensuring truthfulness and semantic accuracy. This necessitates robust protocols for verifying the factual correctness and meaningfulness of knowledge extracted from potentially noisy or ambiguous sources, such as diverse free-text curricula or unvalidated online resources (Mzwri & Turcsányi-Szabo, 2023, Waagmeester et al., 2020).

Human expert validation is a consistent requirement across proposed models. For example, some approaches (Abu-Salih & Alotaibi, 2024, Agrawal et al., 2022) implicitly integrate data validation through rigorous ontology development and schema constraints, where human domain experts refine extracted entities and relations, define types, and apply validation rules to ensure information accuracy and consistency. Similarly, other models (Abu-Rasheed et al., 2025, Qi et al., 2025) explicitly employ a human-in-the-loop approach, with teachers validating LLM-extracted and classified topics and sub-topics. This expert verification, combined with quantitative accuracy data, establishes high-quality and reliable semantic relations, effectively acting as a robust fact-checking mechanism.

Beyond human oversight, the proposed models integrate robust computational and experimental validation to ensure the quality and impact of their KGs. For instance, the MEduKG model (Li et al., 2022) employs rigorous experimental evaluation, utilizing standard metrics such as Accuracy, Precision, Recall, and F1-Score to validate concept recognition and relation extraction, further predicting entities in unlabelled text as a quality assurance measure. Similarly, the proposed KG model (Hu et al., 2023) for professional introduction courses leverages the transE algorithm to achieve over 90% accuracy in predicting relationships, with its educational impact validated by significant improvements in student scores. The e-LION model (Paneque et al., 2023) prioritizes extensive data consolidation and cleaning to resolve semantic inconsistencies, validating its predictive capabilities through rigorous metrics for student grade prediction and time-series forecasting. While the cybersecurity model (Deng et al., 2021) lacks explicit systematic fact-checking for all KG output, it does incorporate content vetting for appropriateness and factual correctness during resource selection.

Finally, emerging research, particularly Knowledge-Aware Validation (Yang R. et al., 2024) for LLM augmentation, explicitly leverages KGs as a robust fact-checking mechanism for generated information, a strategy that could significantly bolster the validation protocols of educational KG models currently lacking explicit fact-checking components.

5 Addressing Critical Gaps in Educational Knowledge Graph Protocols

Despite their significant potential, existing educational KGs are hindered by critical limitations, primarily concerning data quality, including completeness, consistency, and noise. A persistent challenge is maintaining truthfulness and semantic accuracy, especially when dealing with large, continuously evolving datasets common in education. This necessitates urgent development of standardized protocols for KG creation and management to prevent misinformation, ensure comprehensive coverage, and enable effective scalability.

The adaptability to dynamic domains presents another critical limitation. Educational landscapes are constantly evolving with new pedagogical theories, research findings, and curriculum updates (Hu et al., 2023, Mzwri & Turcsányi-Szabo, 2023, Yang R. et al., 2024). Current protocols often fail to easily accommodate these rapid changes, resulting in KGs that quickly become outdated or incomplete. Furthermore, the difficulty in integrating heterogeneous and multi-modal knowledge is pronounced. Educational KGs draw from disparate sources, including structured course syllabi, unstructured research papers, and even multimedia content (Qin et al., 2020, Waagmeester et al., 2020). Ensuring semantic consistency across these varied formats, particularly when concepts are described differently across sources, presents a considerable technical hurdle.

A significant gap observed in many proposed KG models pertains to the explicit implementation of data quality and truthfulness protocols. While efforts are often directed towards robust construction methods or validating the model's utility through learning outcomes (Choi et al., 2023, Deng et al., 2021, Hu et al. 2023, Yang Y. et al. 2024), there is a frequent absence of systematic, granular fact-checking mechanisms for the KG's output itself. For instance, some models might mention general "vetting" of source material, but lack detailed, transparent frameworks for how the factual accuracy of each knowledge triplet or inferred relationship is rigorously verified against external ground truths. This omission is particularly concerning given that educational KGs frequently draw from diverse, potentially noisy, or unvalidated online sources, creating a vulnerability for the introduction and propagation of misinformation (Li et al., 2022, Mzwri & Turcsányi-Szabo, 2023, Waagmeester et al., 2020). The lack of a universal, standardized protocol for assessing factual correctness across varied educational KGs not only leads to inconsistencies but also undermines the verifiable trustworthiness of these crucial learning resources.

Furthermore, proposed models often struggle with the inherent challenges of sparse and incomplete data in educational contexts, alongside the broader issue of scalability. Educational content can be highly specialized and fragmented, making the comprehensive construction of KGs incredibly resource intensive (Abu-Rasheed et al., 2025, Bernasconi et al. 2025, Yang Y. et al., 2022, Yang Y. 2024). While techniques like knowledge graph completion are employed (e.g., TransE in Hu et al., 2023), many models do not fully detail scalable, systematic strategies for identifying and filling data gaps in a way that can transcend specific domains or large datasets. This problem is compounded when considering the semantic accuracy across diverse subject areas as KGs grow. Maintaining consistent definitions and relationships (e.g., how "concept" is understood in mathematics versus fine arts) becomes exponentially more complex with scale. The reliance on human experts for validation (Abu-Rasheed et at., 2025, Agrawal et al., 2022, Qi et al., 2025, Yu et al., 2021), while crucial for accuracy, can become a bottleneck for achieving both comprehensive data completeness and maintaining semantic consistency in vast, evolving KGs.

From a practical standpoint, cost-effectiveness is a major impediment. The extensive manual effort required for data collection, curation, and validation—an inherently continuous process for a comprehensive educational KG—renders thorough quality control prohibitively expensive and often impractical for many institutions. Finally, a significant systemic issue is the lack of standardized metrics and benchmarks. The absence of universally accepted formats or ontologies for educational KGs, unlike in more established domains such as medicine, means there is no common ground to objectively compare the effectiveness of different input control protocols or to seamlessly integrate KGs across diverse educational platforms. This fundamental lack of standardization not only hinders the development of a cohesive educational knowledge ecosystem but also contributes to broader challenges of limited interoperability.

These identified gaps underscore the critical need for standardized protocols for educational KGs. Such a framework should provide clear guidelines for data collection and ingestion, including handling multi-modal content and varied sources; robust ontology and schema design to manage sparsity and ensure semantic consistency; and rigorous knowledge extraction and fusion methodologies to improve completeness and accuracy. Crucially, these protocols must encompass multi-layered validation and fact-checking processes, explicitly integrating mechanisms for external ground-truth verification and for combating issues like LLM hallucination, drawing on advancements such as Knowledge-Aware Validation (Yang R. et al., 2024).

Finally, a comprehensive standard would address scalability considerations from initial design to ongoing maintenance and integrate clear user feedback loops that directly contribute to continuous data refinement and quality enhancement, thereby fostering truly reliable and effective educational KGs.

6 Conclusion

This paper presented a systematic literature review that highlighted key insights from our analysis, particularly regarding the need for high-quality, verifiable knowledge within educational Knowledge Graphs (KGs). Our comprehensive evaluation of current methodologies reveals significant limitations in addressing the unique data challenges inherent in educational contexts. Existing protocols frequently fall short in adequately handling the sheer scale, remarkable diversity, and dynamic nature of educational data.

The full promise of educational KGs is currently constrained by several critical weaknesses, primarily concerning data quality. Key challenges include ensuring the completeness, consistency, and accuracy of information, particularly when managing the extensive and dynamically evolving datasets characteristic of educational environments. KGs frequently contend with sparse and incomplete data, where some learning concepts are exhaustively detailed while others lack any coverage, directly compromising the efficacy of downstream applications such as personalized learning paths. Furthermore, a persistent challenge lies in verifying the truthfulness and semantic accuracy of knowledge extracted from potentially noisy or unvalidated sources.

These pervasive data quality issues, compounded by the computational burden of applying rigorous quality controls to large and dynamic KGs, underscore a pressing need. To fully unlock the transformative potential of KGs in education, the development and adoption of standardized protocols are urgently required. Such protocols would guide the entire lifecycle of educational KGs, ensuring their truthfulness, mitigating the risks of misinformation, enhancing data quality and completeness, and ultimately improving their overall utility and scalability.

Acknowledgements

The research has been co-funded by the University of Rijeka under projects “uniri-drustv- 23-230 3224.

References

Abu-Rasheed, H., Jumbo, C., Al Amin, R., Weber, C., Wiese, V., Obermaisser, R. and Fathi, M. (2025, April). Llm-assisted knowledge graph completion for curriculum and domain modelling in personalized higher education recommendations. In *2025 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1-5). IEEE.

Abu-Salih, B. and Alotaibi, S. (2024). A systematic literature review of knowledge graph construction and application in education. *Heliyon*, *10*(3).

Abu-Salih, B. (2021). Domain-specific knowledge graphs: A survey. *Journal of Network and Computer Applications*, *185*, p.103076.

Agrawal, G., Deng, Y., Park, J., Liu, H. and Chen, Y.C. (2022). Building knowledge graphs from unstructured texts: Applications and impact analyses in cybersecurity education. *Information*, *13*(11), p.526.

Agrawal, G., Kumarage, T., Alghamdi, Z. and Liu, H. (2023). Can knowledge graphs reduce hallucinations in llms?: A survey. *arXiv preprint arXiv:2311.07914*.

Bernasconi, E., Redavid, D. and Ferilli, S. (2025). Enhancing Personalised Learning with a Context-Aware Intelligent Question-Answering System and Automated Frequently Asked Question Generation. *Electronics*, *14*(7), p.1481.

Chen, L., Li, G., Ma, B., Tang, C. and Yamada, M., (2024). A Three-Step Knowledge Graph Approach Using LLMs in Collaborative Problem Solving-Based STEM Education. *International Association for Development of the Information Society*.

Chen, Z., Zhang, Y., Fang, Y., Geng, Y., Guo, L., Chen, X., Li, Q., Zhang, W., Chen, J., Zhu, Y. and Li, J. (2024). Knowledge graphs meet multi-modal learning: A comprehensive survey. *arXiv preprint arXiv:2402.05391*.

Choi, H., Lee, H. and Lee, M. (2023). Optimal knowledge component extracting model for knowledge-concept graph completion in education. *IEEE Access*, *11*, pp.15002-15013.

Deng, Y., Zeng, Z., Jha, K. and Huang, D. (2021). Problem-based cybersecurity lab with knowledge graph as guidance. *Journal of Artificial Intelligence and Technology*.

Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., Metropolitansky, D., Ness, R.O. and Larson, J. (2024). From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*.

Elkaimbillah, Z., Rhanoui, M., Mikram, M. and El Asri, B. (2021, September). Comparative study of knowledge graph models in education domain. In *Proceedings of the Internationa L Conference on Bigdata, Modelling and Machine Learning*.

Fan, X., Sun, Y., Wang, Q. and Wang, X. (2024, December). Research on the Application of Knowledge Graph in Higher Education of Art and Design. In *2024 6th International Conference on Literature, Art and Human Development (ICLAHD 2024)* (pp. 494-501). Atlantis Press.

Fettach, Y., Ghogho, M. and Benatallah, B. (2022). Knowledge graphs in education and employability: A survey on applications and techniques. *IEEE Access*, *10*, pp.80174-80183.

Gligorea, I., Cioca, M., Oancea, R., Gorski, A.T., Gorski, H. and Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. *Education Sciences*, *13*(12), p.1216.

Hou, B., Lin, Y., Li, Y., Fang, C., Li, C. and Wang, X. (2025). KG-PLPPM: A Knowledge Graph-Based Personal Learning Path Planning Method Used in Online Learning. *Electronics*, *14*(2), p.255.

Hu, X., Zhou, Y. and Qi, J. (2023, May). Research on the teaching model of knowledge graph for professional introduction course based on OBE. In *2023 35th Chinese Control and Decision Conference (CCDC)* (pp. 1928-1933). IEEE.

Li, N., Shen, Q., Song, R., Chi, Y. and Xu, H. (2022). MEduKG: A deep-learning-based approach for multi-modal educational knowledge graph construction. *Information*, *13*(2), p.91.

Mzwri, K. and Turcsányi-Szabo, M. (2023). Internet wizard for enhancing open-domain question-answering chatbot knowledge base in education. *Applied Sciences*, *13*(14), p.8114.

Paneque, M., del Mar Roldán-García, M. and García-Nieto, J. (2023). e-LION: Data integration semantic model to enhance predictive analytics in e-Learning. *Expert Systems with Applications*, *213*, p.118892.

Qi, C., Jia, L., Wei, Y., Jiang, Y.H. and Gu, X. (2025). IntelliChain: An Integrated Framework for Enhanced Socratic Method Dialogue with LLMs and Knowledge Graphs. *arXiv preprint arXiv:2502.00010*.

Qin, Y., Cao, H. and Xue, L. (2020, August). Research and Application of Knowledge Graph in Teaching: Take the database course as an example. In *Journal of Physics: Conference Series* (Vol. 1607, No. 1, p. 012127). IOP Publishing.

Qu, K., Li, K.C., Wong, B.T., Wu, M.M. and Liu, M. (2024). A survey of knowledge graph approaches and applications in education. *Electronics*, *13*(13), p.2537.

Tian, L., Zhou, X., Wu, Y.P., Zhou, W.T., Zhang, J.H. and Zhang, T.S. (2022). Knowledge graph and knowledge reasoning: A systematic review. *Journal of Electronic Science and Technology*, *20*(2), p.100159.

Waagmeester, A., Stupp, G., Burgstaller-Muehlbacher, S., Good, B.M., Griffith, M., Griffith, O.L., Hanspers, K., Hermjakob, H., Hudson, T.S., Hybiske, K. and Keating, S.M. (2020). Wikidata as a knowledge graph for the life sciences. *Elife*, *9*, p.e52614.

Weichselbraun, A., Waldvogel, R., Fraefel, A., van Schie, A. and Kuntschik, P. (2022). Building knowledge graphs and recommender systems for suggesting reskilling and upskilling options from the web. *Information*, *13*(11), p.510.

Xue, H. (2023, September). Development History and Frontier Trends of Natural Language Processing Technology in Education: A Knowledge Graph-Based Visualization Analysis. In *Proceedings of the 15th International Conference on Education Technology and Computers* (pp. 436-443).

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y. and Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), pp.90-112.

Yang, R., Yang, B., Ouyang, S., She, T., Feng, A., Jiang, Y., Lecue, F., Lu, J. and Li, I. (2024). Graphusion: Leveraging large language models for scientific knowledge graph fusion and construction in nlp education. *arXiv preprint arXiv:2407.10794*.

Yang, Y. (2024). Research on the digital construction mode of college music teaching curriculum based on knowledge mapping. Applied Mathematics and Nonlinear Sciences. 9. 10.2478/amns-2024-0571.

Yang, Y., Huang, C., Xia, L. and Li, C. (2022, July). Knowledge graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 1434-1443).

Yu, H., Chen, Y., Wang, X., Hu, S. and Chang, L. (2021, May). A capability-oriented method for integration of liberal arts knowledge graph in primary education. In *2021 International Conference on Computers, Information Processing and Advanced Education (CIPAE)* (pp. 342-346). IEEE.