The Hallucination Challenge: Strategies for Ensuring Factuality in AI-Driven Knowledge Management Systems.

Mohd Anas Bin Md Amin, Mohamad Amirul Ain Bin Sakila, Wang Jiamei, Normal Binti Mat Jusoh

Azman Hashim International Business School, Universiti Teknologi Malaysia, Malaysia

DOI: https://doi.org/10.47772/IJRISS.2026.100700320

Received: 18 July 2026; Accepted: 23 July 2026; Published: 31 July 2026

ABSTRACT

This conceptual review sets out to explore and offer a framework in order to tackle the new challenge of AI hallucination in Knowledge Management Systems (KMS). Large Language Models (LLMs) are predicted to affect the creation of knowledge, and they frequently produce content that is not true or even makes up. Raising questions on the integrity and reliability of organizational knowledge, user trust and the reliability of information that is crucial for effective decision making. The paper demonstrates how the existing approaches to establish factual accuracy in the current system are disjoint and distributed between technical and organizational aspects, based on theoretical concepts in artificial intelligence, information systems and knowledge management. Hence in this paper, a hybrid architecture for factuality assurance has been proposed, which does not involve single technical intervention. The proposed framework is aligned with evidence-based retrieval protocols, powerful organizational oversight, explainability protocols, and human-in-the-loop approaches, thus offering a socio-technical approach to AI reliability. This research adds to Knowledge Management theory by moving factuality assurance from being a model-performance competency to a competency of governance. The chapter provides practitioners and policymakers with insights into how to build trust between the knowledge that is produced in organizations and the technology that is rapidly changing, thus enhancing the resilience of knowledge systems powered by artificial intelligence.

Keywords: Artificial Intelligence, Hallucination, Knowledge Management, Retrieval-Augmented Generation, Organizational Governance

INTRODUCTION

The Hallucination Challenge in Knowledge Management

Artificial Intelligence (AI) is a revolutionary technology that is altering the way organizations are treating knowledge, constructing, maintaining and utilizing it. In recent years, due to the development of new technologies like machine learning and Generative AI, KMSs have extended their ability to support the production and synthesis of knowledge, and process decision (Taherdoost & Madanchian, 2023; Sure et al. , 2025). The application of AI based KMS for knowledge discovery is more prevalent than ever in the organizations, helping in automating knowledge processing, organizational learning, and operational decision making (Alsammak et al. 2026; Irfan et al. 2026). It has also driven this evolution to use of knowledge assets to operate more efficiently, innovatively and competitively within organizations.

Large Language Models (LLMs) are the newest generation of AI tools that can generate text that is similar to human writing, answer complex queries, summarize text, and many other knowledge-intensive tasks (Bender et al. , 2021; Ray, 2023). Whereas traditional information retrieval systems discover and present past knowledge, LLMs can generate new content by predicting the linguistic patterns observed during training (Bender et al., 2021). This aspect enables flexible generation of knowledge and conversation; however, it also poses obstacles to factual reliability, as the output is generated through probabilistic prediction and not verified directly against authoritative knowledge sources (Ji et al., 2023; Zhang et al., 2024; Fang et al., 2024). For this reason, these systems can provide flexible, context-appropriate interactions that promote knowledge sharing among

organizations and facilitate the retrieval of relevant information (Taherdoost & Madanchian, 2023; Alsammak et al., 2026).

Generative AI has several advantages for Knowledge Management Systems. However, it has also cast doubt on the reliability and trustworthiness of AI-generated knowledge. The most important of these challenges is that it generates information that superficially appears plausible and coherent, but is factually inaccurate, lacks evidence, contradicts the source information, and/or is completely fabricated, referred to as hallucination (Ji et al., 2023; Zhang et al., 2024; Huang et al., 2025; Fang et al., 2024). A more recent study also highlights that hallucinations arise from failures in knowledge grounding, context reasoning, information retrieval, and model generation in AI models (Liu et al., 2024; Huang et al., 2025), and they remain one of the most persistent constraints on reliably practicing AI at work. Hallucination is a central drawback of existing LLMs, producing inaccurate, incorrect, and misleading output (Fang et al., 2024; Liu et al., 2024). Hallucinations can be a problem for many AI models. However, they can be especially problematic in knowledge-heavy settings where users rely on the AI's output to learn, solve problems, and make decisions.

As concern about hallucinations has grown, they have had a direct impact on broader Knowledge Management goals, rather than just on the model's technical performance. An expected outcome for Knowledge Management systems is that they will deliver information that is accurate, reliable, complete, consistent, and useful to organizational decision-making (Wang, 1996; Mishra & Darade, 2024). Recent studies on data and knowledge quality management also highlight the importance of quality information governance, quality assurance, and information reliability in knowledge processes for organizational value (Addanki & Hullurappa, 2025; Mishra & Darade, 2024). These aspects of quality can be affected by hallucinated outputs that produce false information, different explanations, references that are not always supported, or unsupported conclusions in organizational knowledge repositories (Ji et al., 2023; Zhang et al., 2024). As a result, hallucinations can compromise the quality of knowledge, diminish users' trust in AI-generated knowledge, and limit the effectiveness of AI knowledge initiatives (Alsammak et al., 2026; Cheng et al., 2026).

As AI technologies are introduced into vital business operations, the quality of knowledge has become a growing concern. From this perspective, hallucinations are not merely model errors but are risk factors for knowledge quality that may affect the effectiveness of organizational learning and institutional memory as well as decision making (Taherdoost & Madanchian, 2023; Alsammak et al., 2026 and Irfan et al., 2026). Wang (1996) suggested that the attributes of the quality of information are: accuracy, completeness, consistency and relevance. In addition to access, the trustworthiness and reliability of information are also crucial to its value, and hence its usefulness for an organization, as reported in more recent studies on data and knowledge quality management, (Mishra & Darade, 2024) and (Addanki & Hullurappa, 2025). Hallucinations are a challenge to quality dimensions with artificial intelligence, such as when AI-generated content looks authoritative but is factually incorrect. Therefore, factuality could be an important aspect of the successful use of Knowledge Management Systems (KMSs) with AI.

To address these challenges and improve the accuracy of the knowledge generated by the AI, numerous approaches have been suggested for fact-checking. Some suggestions have been made on how to enhance the factuality and reliability of AI-generated knowledge. One of the most popular approaches is Retrieval-Augmented Generation (RAG), where the LLMs are supplemented with external knowledge sources to enhance their responses and output (Lewis et al., 2020). RAG can be used to generate content grounded in verifiable evidence, to mitigate hallucinations and promote factual consistency. Similarly, Knowledge Graphs and structured knowledge repositories have become a recent trend for supporting semantic reasoning, understanding the context and validating knowledge (Yang et al., 2025). These methods try to reinforce the link between the outputs produced and authoritative sources of information.

A lot of research has been conducted on factuality-evaluation and verification mechanisms and on retrieval and knowledge integration techniques. To assess factual consistency, find unsupported claims and enhance hallucinated content detection, frameworks like QAFactEval and HaluCheck have been developed (Fabbri et al. 2022, Heo et al. 2025). Moreover, researchers have stressed that there should be explainability, governance, accountability, and human oversight regarding the responsible utilization of AI systems (Ribeiro et al., 2016;

Cheng et al., 2026; Li et al., 2026). In recent studies on governance, hallucinations are recognized not only as a technical error but also as an epistemic, organizational, and societal risk that requires appropriate governance mechanisms and accountability systems (Li et al., 2026). These are signs of greater awareness that factuality assurance needs technical and organizational solutions.

While the literature on AI hallucinations has burgeoned in recent years, most of the research is scattered across multiple areas, including machine learning, natural language processing, information systems, and knowledge management. This means that, so far, there is no synthesis on how factuality assurances would be implemented specifically in AI-based Knowledge Management Systems. Given that generative AI is increasingly prevalent in the organizational landscape, understanding these strategies will ensure high knowledge quality, build trust in knowledge, and facilitate human-AI collaboration.

So this paper aims to summarize the hallucination challenge posed by AI-driven Knowledge Management Systems and to investigate recent trends to guarantee the factuality of generated knowledge. It focuses on the reasons and impact behind hallucinations; the evaluation of existing factuality assurance practices, such as Retrieval-Augmented Generation, knowledge-based approaches, verification systems, and governance models; and future research designs for constructing trustworthy AI-powered knowledge ecosystems. By combining contributions from research on concepts in artificial intelligence, knowledge management, and information systems, this study aims to explore how companies can address the trade-off between the potential benefits of generative AI for knowledge management and the need for reliable, trustworthy knowledge management processes. Large Language Models (LLMs) have greatly enhanced the capabilities of artificial intelligence (AI) systems for producing natural-language text and knowledge work. While these systems are robust, they suffer from hallucinations, which pose a considerable challenge to the dependable use of generative AI (Ji et al., 2023; Zhang et al., 2024; Huang et al., 2025).

Understanding Hallucinations in AI-Driven Knowledge Management Systems

Definition and types of hallucinations

Large Language Models (LLMs) have contributed considerably to the emergence of Artificial Intelligence (AI) systems capable of generating natural-language text and performing knowledge tasks. Although these systems can be achieved, they have the drawback of hallucinations, which reduces the potential for using generative AI (Ji et al. 2023; Zhang et al. 2024; Huang et al. 2025). Hallucinations are usually characterized as false or unsubstantiated statements, inconsistent with the original information or even entirely fabricated but seemingly correct (Ji et al., 2023; Zhang et al., 2024; Huang et al., 2025; Fang et al., 2024). Though they may sound grammatically correct, they can be inconsistent with actual knowledge, which poses a major challenge in the reliability of AI for practical applications (Liu et al. 2024; Huang et al. 2025).

Hallucinations are not like typical software bugs that can be easily verified, as the outputs might be coherent, authoritative, and linguistically plausible, yet factually incorrect (Bender et al., 2021; Ji et al., 2023; Fang et al., 2024). This phenomenon has resulted in users' greater willingness to accept false information as knowledge in knowledge-intensive environments (Liu et al. 2024; Cheng et al. 2026). Misinformation can mislead people without them being aware of its factual invalidity (Bender et al., 2021; Liu et al., 2024). This quality is a major problem in Knowledge Management Systems (KMS) because information reliability is critical to the learning process, knowledge sharing, and decision-making within the organization.

There are several classifications of hallucinations in the literature. Ji et al. (2023) divide hallucinations into two categories: Intrinsic and Extrinsic. Intrinsic hallucinations are when information generated does not match what is available or contextual to the source and information; and extrinsic hallucinations are when there is information that is not supported by the available source and contextual information, and cannot be substantiated with available evidence. Huang et al. (2025) also list fact hallucinations, reasoning hallucinations, fabricated references, and misinterpretations of context as frequent occurrences of generation errors. In the same way, Zhang et al. (2024) propose that hallucinations can result from knowledge representation or retrieval, reasoning, or contextual interpretation.

Such classification suggests that hallucinations are not necessarily a single category of error but rather a broad spectrum of biases that involve knowledge representation, contextual interpretations, retrieval, reasoning, and generation behaviors (Ji et al., 2023; Zhang et al., 2024; Fang et al., 2024; Huang et al., 2025). For this reason, to design effective factuality-assurance measures that facilitate knowledge management, we need to discern the nature of hallucination.

Causes and mechanisms of hallucinations

Hallucinations are a complex phenomenon that arises from multiple factors, including the training data used, the model architecture, the language generation processes, and the model's understanding of context. While LLMs are impressive in their language abilities, at their core they make predictions about likely word sequences they observed during training, rather than necessarily checking facts against authoritative knowledge sources (Bender et al., 2021; Zhang et al., 2024).

Limitations in training data are one source of hallucinations. The outputs produced by LLMs can reflect or reinforce inaccuracies found in large-scale corpora that lack the LLM's training, inconsistencies, or inaccuracies in the information that was available when the corpora were created, and biases embedded in the information it contains (Ji et al., 2023; Zhang et al., 2024; Huang et al., 2025). Representativeness and quality of the training data are thus important factors that can affect the reliability of the knowledge generated (Liu et al., 2024). LLMs are trained on vast amounts of text data that are collected from various sources, which may be inaccurate, inconsistent, biased, or contain outdated information (Ji et al., 2023; Huang et al., 2025). Because of the statistical relationships MLMs learn from these datasets, they can replicate inaccuracies or produce answers based on patterns in the training data. In addition, the data used for static training need not continuously reflect real-world knowledge, which poses problems with information freshness and temporal validity.

The self-reliance of language generation is another factor in the formation of hallucinations. In the process of generating text, the language model produces each token one by one, given the tokens already generated. This process allows for coherent and fluent responses, but can also amplify errors via cascading prediction effects. The errors introduced during the initial generation stages can affect later responses, leading to increasingly inaccurate responses (Zhang et al., 2024; Huang et al., 2025).

A second consideration is the limitations of reasoning. Although contemporary LLM systems perform well on a range of language tasks, they are oriented more towards pattern matching than towards semantic understanding or causal reasoning for the task at hand (Bender et al., 2021; Fang et al., 2024). Therefore, when solving complex or domain-specific problems, the models can produce plausible explanations, relationships, or conclusions that are not supported by facts (Zhang et al., 2024; Liu et al., 2024). While the current generation of LLMs can generally perform well on language tasks, they lack real-world understanding of concepts or cause-and-effect relationships. They use rather learned statistical associations to generate responses (Bender et al., 2021). As a result, models can sometimes provide plausible but incorrect explanations, create relationships between concepts that do not exist, or draw conclusions unsupported by the information in a complex query that requires domain knowledge or logic to answer (Fang et al., 2024; Liu et al., 2024).

Another problem with hallucinations is that they are related to poor grounding knowledge. Without relying on external authoritative information sources, language models can use only internal parameterized knowledge, leading them to generate fabricated information and unsupported claims (Ji et al., 2023; Zhang et al., 2024; Huang et al., 2025). This restriction has been a key driver of the emergence of retrieval-based and knowledge-enhanced factuality assurance methods (Yang et al., 2025). Old-fashioned language models produce answers primarily based on knowledge acquired during training and may not have access to newer information sources. However, models tend to continue generating content when they encounter new topics or missing information, thereby increasing the likelihood of producing fabricated outputs (Ji et al., 2023; Zhang et al., 2024). This is especially problematic in the organization environment, where the users seek factual and verified information.

Finally, some contextual ambiguity is possible, to cause hallucination generation. Often times, models can provide details that users do not necessarily mean to provide, such as from ambiguous questions, incomplete instructions and vague prompts (Huang et al. , 2025). Sometimes these speculations can enhance conversation

but at other times can result in unsupported or inaccurate statements. Hence, hallucinations are to be understood as an effect of several technological and contextual factors – and not a model error.

Implications for knowledge quality and organizational decision-making

In the context of Knowledge Management, hallucinations negatively impact the quality of knowledge resources, especially in terms of their accuracy, consistency, completeness, and reliability (Wang, 1996; Mishra & Darade, 2024). With the growing adoption of AI generated content in knowledge-heavy tasks, the correct operation of these quality metrics is crucial for the effective management of knowledge (Taherdoost & Madanchian, 2023; Alsammak et al. , 2026). The quality of knowledge is the degree by which the information is accurate, complete, consistent, relevant and useful to the decision making processes and organizational goal (Wang, 1996). Hallucinations directly impact the quality dimensions by introducing false or unsupported information into AI systems' outputs, thus compromising the trustworthiness of organizational knowledge assets.

The most directly impacted dimension is accuracy, as the information generated by hallucination might take the form of inaccurate facts, false citations, or misleading explanations of facts (Ji et al., 2023; Zhang et al., 2024). However, completeness can also be an issue when outputs are incomplete, or only partially provide a solution to complex issues. Similarly, inconsistencies can arise if AI models give varied responses to similar inquiries or if the content generated by the AI conflicts with the organization's knowledge (Huang et al. 2025; Liu et al. 2024).

The existing problems are quite complex as the use of AI-based Knowledge Management Systems (AIKMS) for knowledge generation, retrieval, exchange, and decision-making processes is becoming more and more common (Taherdoost & Madanchian, 2023; Sure et al. , 2025; Alsammak et al. , 2026; Irfan et al. , 2026). Thus hallucinated information can spread beyond the personal interactions and even into the organizational information repositories, reports and analyses and impact future decision making and organizational learning.

AI systems can produce inaccurate knowledge which can infect other organizational knowledge stores, reports and decision support systems resulting in hazards that reach beyond the immediate interactions with the system. Therefore, hallucinations are not only technical errors, but also threats to the quality of organizational knowledge (Alsammak et al. 2026; Irfan et al. 2026).

Another key factor to consider is trust. Users' trust in the accuracy of AI-generated information is crucial for the success of AI-driven Knowledge Management Systems. Research indicates that hallucination can have a negative impact on cognitive and emotional trust that could, in turn, hinder human-AI partnerships and the use of AI-powered knowledge tools (Cheng et al. , 2026). In this regard, factual reliability shall be taken as the foundation to sustainable knowledge management supported by AI.

Users' trust and belief in the AI-generated content are more likely to lead to the use of AI-powered knowledge tools. Users' trust in AI-generated knowledge is positively associated with their use of AI tools for knowledge generation. Exposure to hallucinated results could lead to a decrease in trust in AI systems and a loss of trust in AI-generated knowledge for decision-making purposes (Cheng et al. , 2026). Especially for knowledge-intensive situations, where organisational performance is dependent on the quality and trustworthiness of the information available.

Recent studies depict it as a delicate balance between promoting automation and ensuring the factual accuracy and quality of knowledge in Knowledge Management Systems (KMS) powered by artificial intelligence (Taherdoost & Madanchian, 2023; Alsammak et al. , 2026). With the increasing adoption of generative AI capabilities by organizations, it is essential to manage hallucinations to ensure trust, informed decision-making, and integrity of organizational knowledge resources.

To address these challenges, there has been a significant amount of research work to develop methods to improve factual correctness and minimize hallucinations. Next are the existing approaches to factuality assurance, including Retrieval-Augmented Generation, knowledge-based approaches, verification approaches, hybrid approaches, etc., for reliable AI-powered Knowledge Management Systems.

Strategies For Ensuring Factuality in AI-Driven Knowledge Management Systems

Retrieval-augmented generation

One of the most impactful methods for mitigating hallucinations and enhancing factual accuracy in Large Language Models (LLMs) is Retrieval-Augmented Generation (RAG) (Lewis et al., 2020; Yang et al., 2025). Such limitations were addressed in this approach, as models might generate plausible yet factually incorrect responses when the necessary information is unavailable or outdated (Ji et al., 2023; Zhang et al., 2024; Huang et al., 2025). Traditional language models can provide answers primarily from information learned during training, which limits access to up-to-date, domain-specific, and organization-specific knowledge. Consequently, they can generate "reasonable" answers even when the information they are asked about is unfamiliar or changing (Ji et al., 2023; Zhang et al., 2024). To address this, RAG introduces information-retrieval components into the generation phase, allowing models to retrieve relevant external knowledge sources prior to generation (Lewis et al. , 2020).

The premise of RAG is to retrieve relevant information, such as documents, records or knowledge assets, from external sources and use them as input to the generation process (Lewis et al. , 2020). RAG helps to provide grounded responses, and reduce the need to rely on internal model memory, which can be more consistent with facts, transparent, and explainable (Yang et al. 2025; Huang et al. 2025; Liu et al. 2024). The system also makes inferences based on the evidence from authoritative sources, in addition to the evidence from internal models (Lewis et al., 2020; Yang et al., 2025). In Knowledge Management Systems (KMS) it is particularly important that data repositories are updated regularly, and that decisions based on knowledge are based on the latest information.

As RAG becomes more popular, it's proving to be a great way of reducing hallucinations. RAG allows users to access reliable information when generating a response, minimizing the need to rely on potentially incomplete or outdated information embedded in the model parameters (Huang et al. , 2025; Yang et al. , 2025). For example, empirical studies have demonstrated that grounding with retrieval provides more factually consistent results, enhances transparency and generates more reliable decisions compared to language models alone (Lewis et al., 2020; Zhang et al., 2024).

RAG is in-line with Knowledge Management objectives of retrieval, reuse and dissemination. RAG helps to connect AI systems to institutional repositories, policy documents, databases, and other knowledge resources, providing relevant and evidence-based information. RAG, on the other hand, connects AI systems with institutional knowledge sources like repositories, policy papers, databases, and more, delivering relevant and evidence-based information. Modern businesses store large volumes of documents, policies, reports, databases, and institutional knowledge. RAG allows AI systems to dynamically retrieve information from these repositories, enabling them to respond based on the organization's knowledge rather than generic knowledge acquired during pretraining (Taherdoost & Madanchian, 2023; Sure et al., 2025). This allows it to make more relevant and reliable knowledge.

Even though it has its merits, RAG does not solve all the challenges of hallucination. Retrieval performance remains highly correlated with the quality of the generated outputs. Although access to accurate information, the selection of relevant documents, and the quality of knowledge repositories can ensure accurate outputs, these factors are not guaranteed. In addition, retrieved information can be incomplete and contain contradictions, so additional mechanisms are needed for validation and reasoning. As such, it is becoming an increasingly important solution for ensuring factuality, apart from RAG.

Knowledge-based methods and knowledge graph integration

Alongside retrieval strategies, researchers started to consider knowledge-based solutions as adjuncts to improve the reliability of fact-based information and decrease hallucinations (Yang et al., 2025; Zhang et al., 2024). The intent of these is to utilize structured knowledge representations (i.e., ontologies, semantic networks, Knowledge Graphs) to improve language models. These knowledge-centric approaches seek to enhance language models with formal knowledge representations such as ontologies, semantic networks, and Knowledge Graphs (Yang et

al., 2025). These techniques explicitly represent entities, concepts, and relationships, enabling consistent reasoning across the domain and validation of the knowledge gained. Recent reviews indicate that structural knowledge representations can overcome the shortcomings of only statistically based language generation by explicitly establishing relationships between entities, concepts, and facts (Yang et al., 2025; Fang et al., 2024). These representations help to keep the semantics consistent and to perform evidence based reasoning processes.

An interesting aspect of Knowledge Graphs is that they present information in a structured manner, relating different entities and their relationships. While unstructured textual information is useful for humans, Knowledge Graphs, which can be machine-processed, can be used for semantic reasoning and relationship verification (Yang et al. , 2025). By using structured knowledge, AI systems can generate more uniform content, reducing the risk of making unwarranted claims.

Sure et al. (2025) and Taherdoost & Madanchian (2023) state that Knowledge Graphs can be used in Knowledge Management Systems to connect knowledge, facilitate information retrieval, and provide context about information coming from different sources of knowledge within the organization. These facilitate the creation of better quality knowledge, and it makes sure the products created are consistent with the organizational structure of knowledge (Alsammak et al. 2026). They facilitate knowledge integration across different sources of information, improve information findability, and add context to organizational knowledge resources (Sure et al. , 2025). These capabilities come in very useful when knowledge is stored in different forms and in different places. Semantic relations help AI systems to link relevant information and make decisions which are more closely aligned with organizational knowledge systems. Explainability is also improved with knowledge based approaches. The concepts are clearly communicated and the relationships between them are explicitly represented, allowing the user to learn more about how the conclusions are developed and what information is used to generate the outputs.

This transparency helps to build trust and accountability, and helps to validate knowledge (Yang et al. , 2025). Hence, Knowledge Graphs are emerging as an important part of Knowledge Management Systems based on Artificial Intelligence techniques that can be trusted.

However, several issues need to be addressed using knowledge-based methods. Building and maintaining Knowledge Graphs is resource consuming, particularly if the organisation's knowledge is continually changing. Moreover, when the representation of knowledge is incomplete and/or outdated, there may be limitations such as those of traditional retrieval systems. Nevertheless, the literature highlights that the knowledge-based approach can form a good base for enhancing factual reliability and knowledge quality as an organization.

Verification and evaluation frameworks

Both retrieval and knowledge integration methods tend to minimize hallucinations during content generation, and verification methods are designed to verify the factual consistency of generated outputs against available evidence. (Fabbri et al., 2022; Heo et al., 2025). These approaches are an important advancement in hallucination research, as they enable consideration of reliability alongside generation quality.

A significant one is the work of QAFactEval, which tests factual consistency by comparing information generated in a given context with reference information using question-answering approaches (Fabbri et al., 2022). The framework does not rely solely on linguistic measures of similarity; it also assesses whether an algorithm-generated claim can be supported by evidence. This is a more useful way to assess factual accuracy and has become the standard in the research on factuality assessment.

More recently, the explainable hallucination detection framework, HaluCheck, detects hallucinations and provides evidence-based explanations for detected inconsistencies (Heo et al., 2025). These methods help to establish factuality and transparency, key elements of trustworthy AI systems (Liu et al., 2024; Cheng et al., 2026). HaluCheck not only helps detect hallucinations but also enhances transparency and user trust through automated verification and clear explanations.

Verification frameworks are important quality assurance tools for Knowledge Management Systems. Like classical information quality control, these mechanisms are used to ensure that knowledge assets meet requirements of accuracy, consistency, and reliability (Wang, 1996; Mishra & Darade, 2024). Verification frameworks are complementary to retrieval and knowledge integration strategies, as they can provide an additional source of fact-checking prior to integrating generated information into organizational knowledge processes. In an organizational setting, verifiers can be considered mechanisms to ensure the quality of knowledge creation and governance and to minimize the likelihood that inaccurate knowledge will be added to knowledge repositories or decision-making processes (Mishra & Darade, 2024; Addanki & Hullurappa, 2025).

Hybrid factuality assurance architectures

It is now becoming clear that given the diversity of the sources of hallucinations, such as limited training data, reasoning failures, retrieval errors, and contextual ambiguities (Ji et al., 2023; Fang et al., 2024; Huang et al., 2025), no single method can be used to remove hallucinations completely. Therefore, researchers have turned their focus to hybrid methods that combine several complementary approaches to FA (Yang et al., 2025). Therefore, hybrid approaches for factuality assurance have attracted attention, integrating multiple mechanisms into a unified system (Yang et al., 2025; Huang et al., 2025).

Hybrid approaches usually include retrieval systems, formal knowledge bases, verification systems, governance measures, and human supervision procedures. Retrieval-Augmented Generation is applied to evidence grounding (Lewis et al. 2020), Knowledge Graphs for semantic reasoning (Yang et al. 2025), and verification systems for checking factual consistency (Heo et al. 2025) and human experts for contextual judgment and domain knowledge. The external knowledge can be accessed via Retrieval-Augmented Generation, the Knowledge Graph can be used to perform semantic reasoning, the verification framework can verify the factual consistency, and the human expert can provide contextual judgment and domain knowledge. All these components together provide several levels of protection against the risk of hallucination.

It should be noted that from the Knowledge Management point of view, hybrid architectures closely resemble the socio-technical aspects of organizational knowledge ecosystems where technology, processes, governance structures, and the expertise of humans play a role in maintaining knowledge quality (Taherdoost & Madanchian, 2023; Alsammak et al., 2026; Irfan et al., 2026). Getting good knowledge management results does not always rely on a single technology. Combinations of repositories, governance structures, validation processes and human expertise, are adopted to ensure knowledge quality (Taherdoost & Madanchian, 2023; Alsammak et al. 2026). The idea of hybridity is then expanded to hybrid factuality assuring systems, where complementary mechanisms are taken into consideration to fight against other types of hallucinations in AI-assisted settings.

In the literature, these so integrated approaches are increasingly cited over the years as a component of future reliable Knowledge Management Systems. To enhance reliability, organizations are more inclined to leverage a more complete solution that integrates retrieval, knowledge, verification, and governance into a unified factuality-assurance ecosystem, as opposed to trying to address hallucinations only with model improvements (Yang et al. , 2025; Heo et al. , 2025). All these developments have a significant impact on the accuracy, reliability, and appropriateness of AI knowledge for organizational decision-making.

Governance, Trust, And Human Oversight in AI-Driven Knowledge Management Systems

Trust and human-AI collaboration

Therefore, AI applications of Knowledge Management Systems (KMS) rely heavily on trust, as KMS users will not trust knowledge produced through AI when it is considered suspect, unverifiable, or conflicting (Cheng et al., 2026; Ray, 2023). Trust has become a pivotal factor in the acceptance of generative AI, the use of knowledge generated by such technology, and the collaborative relationship between humans and AI, particularly as organizations use generative AI for knowledge-intensive tasks (Taherdoost & Madanchian, 2023; Alsammak et al., 2026). Even if AI's performance is good, users are not going to trust knowledge created by AI if they believe it is unreliable or unverifiable. Hallucinations are becoming a significant challenge as Large Language Models

(LLMs) have become commonplace, since they can lead to distrust of information and negatively influence AI-supported knowledge processes (Cheng et al., 2026; Ray, 2023).

Recent studies have shown that hallucinations can have a detrimental effect on cognitive and emotional trust in AI systems (Cheng et al., 2026). Cognitive trust is the extent to which users trust the accuracy and reliability of AI output, and emotional trust is the readiness for reliance on AI systems for cooperation. The trust of both can be undermined if users experience hallucinations, a serious threat that will further dampen their confidence in AI decision-making (Cheng et al., 2026; Liu et al., 2024). Cognitive trust is defined as trust in the accuracy, competence, and reliability of AI-generated information. In contrast, emotional trust is defined as users' willingness to rely on AI systems in collaborative situations (Cheng et al., 2026). If AI systems provide false or fabricated information, both measures of trust can suffer, potentially hindering user adoption and reducing humans' productive use of AI.

AI-generated outputs play an increasingly significant role in knowledge discovery, problem-solving, and decision-making, making this an even more significant issue in the context of Knowledge Management Systems (KMS) (Taherdoost & Madanchian, 2023; Alsammak et al., 2026). Misinformation can also impact organizational learning and knowledge-sharing processes. Thus, mechanisms must be in place to keep people's trust that the knowledge generated is reliable and transparent.

The literature has progressed, and it is increasingly recognized that successful knowledge management must be viewed as a collective project that requires human intervention and the application of Artificial Intelligence (AI) capabilities (Sure et al., 2025; Irfan et al., 2026). AI is intended to augment, not replace, human sensemaking in the process of creating, retrieving, and sharing knowledge, as human-derived evidence and expertise can assist in understanding the context, evaluating critically, and logically constructing conclusions according to that knowledge (Taherdoost & Madanchian, 2023; Cheng et al., 2026). AI systems should never replace human judgment (Sure et al., 2025; Irfan et al., 2026); their purpose is to augment knowledge acquisition processes within an organization and improve retrieval and analysis. From this view, trust is a key determinant in the nature and impact of human-AI collaboration.

Governance and human oversight

Although technological solutions such as Retrieval-Augmented Generation, Knowledge Graph integration, and verification methods have considerable potential to reduce the risk of hallucinations, they are not always accurate (Lewis et al., 2020; Yang et al., 2025; Heo et al., 2025). In that context, governance and human oversight have emerged as important complementary tools for ensuring the reliable application of AI (Li et al., 2026; Cheng et al., 2026). Governance and human oversight become critical elements of the factuality assurance frameworks in this context (Li et al., 2026; Cheng et al., 2026). This is evidenced by Li et al.(2026), who state that AI governance encompasses policies, procedures, standards, and accountability measures for managing the development, deployment, monitoring, and evaluation of AI systems. Hallucinations have been found to be epistemic, organizational and societal risks, which can affect the quality of decision making, trust in the system and adherence to regulatory requirements, in addition to being a technical error (Li et al., 2026). In Knowledge Management Systems (Ray, 2023), information quality and accuracy, and responsible use of the system are all supported by governance frameworks. These are important because hallucinations are not just technical issues, but issues that impact organizational decision making, adherence and trust of stakeholders.
Human participation is still required in governance, as it is the human experts who are able to detect any mistakes in the context, assess the quality of the evidence, and make judgments when automated systems might be inaccurate (Cheng et al., 2026; Ray, 2023). This is particularly important in a knowledge-intensive environment, where information might be created that impacts strategic planning, policy-making, or organizational learning processes (Taherdoost & Madanchian, 2023; Alsammak et al. , 2026; Hameed et al. , 2026). In knowledge-intensive situations, where the result can impact on strategic, policy or operational decisions, there must be human involvement.

Human oversight in Knowledge Management is an important means of ensuring knowledge quality, providing an additional layer of quality control to prevent information from being incorporated into enterprises' knowledge

bases and decision processes (Taherdoost & Madanchian, 2023; Alsammak et al. , 2026). The takeaway is that the adoption of AI isn't credible without the technological and organizational measures.

The use of governance measures introduces accountability by outlining responsibility around the output of AI. With the increasing number of organizations, there's a heightened need for a comprehensive validation process of the knowledge creation process, monitoring system performance, and rectifying errors (Li et al. , 2026). These measures can contribute to minimizing the occurrence of hallucinations and fostering safe application of AI tools in organizations. Furthermore, governance processes can be used to facilitate continuous monitoring and quality assurance systems and to enhance validation, auditing and accountability. Through these mechanisms, organizations can detect potential risks for hallucination and ensure knowledge informed by AI remains accurate and reliable over time (Mishra & Darade, 2024; Addanki & Hullurappa, 2025).

Explainability and accountability

Explainability is also a critical aspect of trustworthy AI, as users are more likely to trust and interact with the AI-generated knowledge when they have an understanding of how the AI reaches its conclusions and the evidence it relied on in the process (Ribeiro et al. 2016; Cheng et al. 2026). In Knowledge Management Systems, explainability helps to validate, transparently, and accountably knowledge based on the users' ability to evaluate the validity of the suggestions and decisions.

Explainability goes hand in hand with hallucination because generated (and unsupported) findings can sound convincing and authoritative (Ji et al. 2023; Fang et al. 2024; Liu et al. 2024). Explainable AI methods allow users to find inconsistencies and comprehend how the model works and how reliable the created content is (Ribeiro et al. , 2016). The information received in the hallucination is often the users thinks is valid and authoritative so they may have trouble telling what is valid and what is not. Users believe that the information in a hallucination is truthful and correct, which is why it is difficult to determine what is real and what is not. For this, the explainable AI techniques try to understand the model's behavior and find factors that influence the model's responses (Ribeiro et al. , 2016). This will enable people to understand better and help them to make better assessments of AI-generated knowledge.

Evidence attribution is a practical approach to explanation, whereby the output generated is associated with evidence documents, knowledge bases, or external sources (Yang et al., 2025; Heo et al., 2025). These safeguards also apply to independently verifiable information, enhancing trust in AI-generated knowledge, especially within organizations that prioritize transparency. AI can also connect information to relevant documents, repositories, or knowledge resources to confirm or verify it, without requiring input or depending on other people, using its own information resources. This ability is particularly important in organizations where knowledge quality and accountability are vital (Sure et al., 2025; Yang et al., 2025).

Explainability can also aid in accountability by enhancing the transparency of processes and providing users and decision makers with evidence. As such, this level of transparency aids governance, quality assurance, and accountability-assignment processes for AI-generated decisions and knowledge generation (Li et al. , 2026; Cheng et al. , 2026).

Invisible reasoning processes and supporting evidence help organizations to compare information generated, identify possible holes in its reasoning, and discover possible loopholes in its reasoning. This transparency allows for more accountability, not only in the creation of the system and the promotion of governance goals, but also in determining accountability for AI-driven decisions and knowledge products (Li et al. , 2026).

In conclusion, the literature review results show that aspects related to trust, governance, human oversight and explainability need not be solved but can be considered augmentations/supplements to factuality assurance. Technical methods improve the factual content while organizational methods increase the credibility and accountability of the AI-generated knowledge, support its integration into the organisation and give it the context to be appropriate. These socio-technical factors will increasingly be relevant, as knowledge management technologies become smarter.

Research Gaps and Future Directions

While the understanding and quantification of AI hallucinations have also significantly progressed, there are identified areas where factuality assurance is not highly effective in KMS that rely on AI. Existing research with valuable insights on mechanisms of hallucination, retrieval-based methods, knowledge integration methods and verification frameworks. However, today much research is fragmented, distributed across fields such as artificial intelligence, natural language processing, and information systems, leading to poor integration between technical and organizational perspectives (Yang et al., 2025; Huang et al., 2025). In this context, several regions still require further research to develop trustworthy knowledge ecosystems with AI.

Integrating factuality assurance with knowledge management theory

There is a lack of integration of Knowledge Management theory in the current literature on hallucination. The majority of research focuses on the technical aspects of hallucinations, including model architecture, retrieval performance, factual accuracy, and task performance (Ji et al., 2023; Zhang et al., 2024; Heo et al., 2025). These contributions have furthered our understanding of mechanisms and strategies for mitigating hallucinations, but there has been less research on the nature of hallucinations as a challenge to OTQRs as organizational knowledge quality risks (Taherdoost & Madanchian, 2023; Alsammak et al., 2026). To date, most studies use technical metrics such as factual consistency, retrieval accuracy, or model performance (Heo et al., 2025; Ji et al., 2023) to evaluate hallucinations. These are obviously vital, but they do not account for the broader organizational issues arising from hallucinated information.

Taherdoost and Madanchian (2023) and Alsammak et al. (2026) highlight knowledge quality, organizational learning, knowledge sharing, and decision support as important aspects of Knowledge Management research. However, relatively few studies focus on factuality dimensions and their long-term impact on knowledge quality management, as well as on the impact of factuality assurance strategies on these dimensions. Hallucination, however, should be studied not only as a technical failure of the model, but as an organizational failure of knowledge, in the future. This could involve studying how hallucination affects knowledge generation, knowledge sharing, organizational learning, and the efficiency of the decisions taken in knowledge environments enabled by AI (Sure et al., 2025; Irfan et al., 2026). A combination of such sets would provide a broader understanding of the implications of AI knowledge management for organizational activities.

Managing dynamic knowledge environments

The other challenge concerns the management of dynamic knowledge environments. An organization's knowledge is continuously evolving with the advent of new technologies, regulatory changes, market changes, and operational changes (Sure et al., 2025; Irfan et al., 2026). While numerous existing methods for factuality assurance still rely on fixed knowledge bases, pre-defined evaluation datasets, or periodically refreshed knowledge stores (Lewis et al., 2020; Yang et al., 2025), there is a lack of methods that can leverage real-time knowledge updates. Organizational knowledge is continually evolving in response to changes in regulations, technology, markets, and operational needs (Sure et al., 2025; Irfan et al., 2026). Many of the approaches to factuality assurance, however, are based on relatively static knowledge repositories or evaluation datasets.

While Retrieval-Augmented Generation enhances the retrieval of external information, the quality, relevance, and timeliness of the retrieved knowledge remain major challenges (Lewis et al., 2020; Yang et al., 2025). In the same way, knowledge-based approaches need to be updated regularly to ensure that knowledge structures reflect the organization's evolving context. A critical element for assuring factuality and reliability in fast-changing knowledge environments are adaptive factuality assurance mechanisms which are essential for better research in the future. Specific focus needs to be given to knowledge freshness, repository governance, dynamic retrieval strategies, and ongoing validation processes to maintain the quality of knowledge in the long term (Mishra & Darade, 2024; Addanki & Hullurappa, 2025). Other factors, such as knowledge freshness, repository governance, and continuous validation processes need to be considered.

Explainability, trust, and human-AI collaboration

While the significance of explainability in building trustworthy AI deployments has been acknowledged (Ribeiro et al., 2016; Cheng et al., 2026), the connection between explainability, trust, and human–AI collaboration is still poorly understood. Existing studies tend to view explainability as a tool for gaining users' trust, but little empirical research has been conducted on the impact of various ways of explaining on the acquisition of knowledge and the quality of the resulting decision. The literature suggests that explainability can improve transparency and increase the level of trust of the user, while there is limited empirical evidence, specifically, on the effectiveness of the type of explainability mechanism in the organizational setting (Ribeiro et al. 2016; Cheng et al. 2026).

Users' perceptions of AI explanations, their assessment of supporting evidence, and responses to uncertainty in AI-generated knowledge are critical questions to consider. Similarly, more research is required to balance the automation and human supervision in knowledge-intensive environments where hallucinations are possible (Cheng et al. 2026; Li et al. 2026). Furthermore, the interplay between risks, hallucinations, and collaboration between humans and AI needs to be further examined. While human participation is highly desirable to avoid factual mistakes, it is not yet known what is the "best" mix between automatic and human. Research into mechanisms for explainability and control to facilitate safe human-AI collaboration in Knowledge Management scenarios is therefore a possibility.

Towards hybrid factuality assurance ecosystems

The literature is progressively hinting that no single method can eliminate hallucinations because hallucinations come from a complex interaction of the following factors: training data limitations, reasoning failures, contextual ambiguity, and insufficient knowledge grounding (Ji et al., 2023; Fang et al., 2024; Huang et al., 2025; Liu et al., 2024). Therefore, integrated factuality-assurance systems with co-occurring mitigations still need to be investigated by researchers. To address various problems at once, the retrieval mechanisms, knowledge-based approach, verification frameworks, and governance controls pose distinct challenges that require attention, yet they all have limitations (Yang et al., 2025; Heo et al., 2025). Therefore, the next step in the research is to create hybrid ecosystems of factuality-assurance combining various complementary approaches. These ecosystems might also involve Retrieval-Augmented Generation, Knowledge Graphs, automated verification frameworks, governance controls, explainability mechanisms, and human oversight processes (Lewis et al. , 2020; Yang et al. , 2025; Heo et al. , 2025; Li et al. , 2026). This orientation has many parallels with Knowledge Management, where technology, processes, governance and people need to be combined to ensure the quality and effectiveness of knowledge in the organization (Taherdoost & Madanchian, 2023; Alsammak et al., 2026). This is also in alignment with Knowledge Management theory, which asserts that meaningful ways of creating and applying knowledge in relation to the domains of technology and processes cannot be produced solely through the interaction of technology and human expertise (Taherdoost & Madanchian, 2023; Alsammak et al., 2026). Future studies should examine whether these integrated approaches can affect knowledge quality, organizational trust, and decision-making performance across the application domains. The literature as a whole demonstrates significant progress in the area of hallucinatory experiences and the reliability of the truth. AI-generated knowledge, however, is a nascent challenge in context, which requires linking up recent developments in the AI and Knowledge Management domains with current research. To create AI-driven Knowledge Management Systems that are innovative, reliable, and trusted by the organization, filling these identified gaps is becoming more critical.

CONCLUSION

The increasing adoption of AI and Large Language Models (LLMs) in Knowledge Management Systems (KMS) has created opportunities to enhance knowledge generation, accessibility, sharing, and decision-making. This is what makes AI-based solutions appealing options that may assist organisations in learning in a systems approach, leverage huge data sets and gain insights that are relevant to their context, and enhance operational efficiencies. Despite this, a key concern surrounding the increased use of generative AI is the accuracy of the content generated by AI. As part of our review, we have explored hallucinations in AI systems and their implications for knowledge management in this area. From previous studies, it is proposed that there are multiple reasons for the emergence of hallucinations in LLM, among them inadequate training data, probabilistic language

production, lack of reasoning capabilities, lack of knowledge grounding, and general ambiguity. Hallucinated information can seem realistic and credible at first glance, but brings inaccurate, unverified and false data into any work that relies on knowledge in an organization. Thus, hallucinations should not only be seen as the shortcomings of technical AI systems but as significant risks of knowledge loss, impacting organizational learning, decision-making, and user trust. Recognizing the above challenges, the review looked into different recent solutions to the problem of factuality assurance. Retrieval-Augmented Generation (RAG), which has gained significant momentum recently, is an approach that uses external knowledge sources to improve factual correctness in the content of the responses generated, while knowledge-based approaches like Knowledge Graphs provide structured representations that can be used for semantic reasoning/knowledge validation. Tools for verification are also useful for the identification of factual inaccuracies and for verification of the output. Also we mention in our literature that these approaches are not suitable to be used as stand-alone solutions. According to the review, governance, human oversight, explainability, and accountability were key dimensions of reliable AI deployment. Although technology can help avoid this issue and reduce the risk of hallucination, organizations must develop appropriate measures to maintain effective knowledge and ensure the appropriate utilization of AI-generated knowledge. Thus, trustworthy Knowledge Management Solutions depend on a trade-off between automated intelligence and human decision-making, backed by a governance infrastructure with mechanisms that promote transparency, validation, and responsibility. Several research gaps remain. More specifically, greater integration of hallucination research into knowledge management theory is warranted better to understand the organizational ramifications of AI knowledge production. Moving forward, we encourage future work on factuality assurance mechanisms that can respond to a dynamic knowledge landscape, support explainability, account for trust relations, and adopt hybrid architectures (a retrieval, knowledge integration, verification, and governance system, in essence). More broadly, these results suggest that factuality is an important concern when constructing AI-based Knowledge Management systems. The reliability, transparency, and trustworthiness of knowledge will be even more important as organizations are increasingly adopting generative AI technologies. This makes integrated factuality assurance strategies for managing knowledge across an organization appropriate, as well as the ability to trust the AI model when making decisions, critical to defeating the hallucination problem.

REFERENCES

1. Addanki, S. & Hullurappa, M. (2025). Sustainable Data Quality Management: AI-Driven Approaches for Reducing Technical Debt in Data Governance. In S. Kulkarni, M. Valeri, & P. William (Eds.), Driving Business Success Through Eco-Friendly Strategies(pp. 419-440). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-9750-3.ch022
2. Cheng Q., Dai Y., Liu X., Peng S. (2026). The trust crisis in artificial intelligence: AI hallucinations and human-AI collaboration, Technology in Society, Volume 86, 103286, ISSN 0160-791X, https://doi.org/10.1016/j.techsoc.2026.103286
3. Chen P. H., Huang Y. M., Wu T. T., Lee H. Y. (2025). Mitigating Artificial Intelligence Hallucinations in Education: A Comparative Study of Retrieval-Augmented Generation (RAG) and Large Language Models, 7th International Conference on Modern Educational Technology (ICMET), Fukuoka, Japan, 2025, pp. 230-236, doi: 10.1109/ICMET67594.2025.11451842
4. Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 610–623. https://doi.org/10.1145/3442188.3445922
5. Fabbri A. R., Wu C. S., Liu W., Xiong C. (2022). QAFactEval: Improved QA-Based Factual Consistency Evaluation for Summarization. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 2587–2601, Seattle, United States. Association for Computational Linguistics. 10.18653/v1/2022.naacl-main.187
6. Fang X., Yuan H., Li H., Kou J., Gu C., Zhang W., Duan X., Fang Y. (2024). Reframing Hallucination in Large Language Models: A Lifecycle-Based, Mechanism-Aligned, and Phenomenon-Consistent Definition, 7th International Conference on Universal Village (UV), Boston, MA, USA, 2024, pp. 1-15, doi: 10.1109/UV63228.2024.11189217

7. Gunasekara, C., Hamel, Z., Joseph, R.B., Du, F. (2026). Bilingual Knowledge Management System for Information Retrieval from Military Policies. In: Fred, A., De Marsico, M., Castrillón-Santana, M. (eds) Pattern Recognition Applications and Methods. ICPRAM 2024. Lecture Notes in Computer Science, vol 15568. Springer, Cham. https://doi.org/10.1007/978-3-032-06007-5_6
8. Heo S., Son S., Park H. (2025). HaluCheck: Explainable and verifiable automation for detecting hallucinations in LLM responses, Expert Systems with Applications, Volume 272, 126712, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2025.126712
9. Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B. (2025). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. ACM Trans. Inf. Syst. 43, 2, Article 42 (March 2025), 55 pages. https://doi.org/10.1145/3703155
10. Irfan, D., X. Tang, A. Khushk, andX. Yi. (2026). AI-Driven Knowledge Management for Resilient Electric Vehicle Manufacturing: Optimizing China's Supply Chains.Knowledge and Process Management1–16. https://doi.org/10.1002/kpm.70073.
11. Ji Z., Lee N., Frieske R., Yu T., Su D., Xu Y., Ishii E., Bang Y. J., Madotto A., Fung P. (2023). Survey of Hallucination in Natural Language Generation. ACM Comput. Surv. 55, 12, Article 248 (December 2023), 38 pages. https://doi.org/10.1145/3571730
12. Kushwah S., Dave N. (2025). AI Hallucination and Strategies to Overcome: Enhancing Human-AI Interaction,2025 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, pp. 1-6, doi: 10.1109/AIMV66517.2025.11203756
13. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks.Advances in neural information processing systems,33, 9459-9474. https://doi.org/10.48550/arXiv.2005.11401
14. Li, Z., Yi, W., & Chen, J. (2026). Accuracy paradox: Addressing epistemic, manipulative, and societal risks of hallucination in AI governance. Computer Law & Security Review, 61, 106311. https://doi.org/10.1016/j.clsr.2026.106311
15. Liu, Y., et al. (2024). Comprehensive evaluation of AI hallucination and novel UV-oriented framework toward safe and trustworthy AI. 7th International Conference on Universal Village (UV), Boston, MA, USA, 2024, pp. 1-136, doi: 10.1109/UV63228.2024.11189137.
16. Mishra V., Darade H. (2024). Advancements in Data Quality Management in the Big Data Era: A Comprehensive Review, Third International Conference on Artificial Intelligence, Computational Electronics and Communication System (AICECS), MANIPAL, India, 2024, pp. 1-6, doi: 10.1109/AICECS63354.2024.10956447
17. Nah F. F., Zheng R., Cai J., Siau K., & Chen L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research,25(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814
18. Rahman, A., Yu, A., & Cho, K. (2026). Game Knowledge Management System: Schema-Governed LLM Pipeline for Executable Narrative Generation in RPGs.Systems,14(2), 175. https://doi.org/10.3390/systems14020175
19. Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems, 3, 121–154. Volume 3, Pages 121-154, ISSN 2667-3452, https://doi.org/10.1016/j.iotcps.2023.04.003
20. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 1135–1144. https://doi.org/10.1145/2939672.2939778
21. Sure T. A. R., Saigurudatta P.V., Kapoor S., Kandula S. T. R., Choudhury A., Devendran P.D., (2025) The Role of Natural Language Processing in Developing Intelligent Knowledge Repositories, IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), Bali, Indonesia, 2025, pp. 785-790, doi: 10.1109/IAICT65714.2025.11101416.
22. Taherdoost, H., & Madanchian, M. (2023). Artificial Intelligence and Knowledge Management: Impacts, Benefits, and Implementation.Computers,12(4), 72. https://doi.org/10.3390/computers12040072

23. Troussas, C., Papakostas, C., Krouska, A., Mylonas, P. and Sgouropoulou, C. (2024). FASTER-AI: A Comprehensive Framework for Enhancing the Trustworthiness of Artificial Intelligence in Web Information Systems. In Proceedings of the 20th International Conference on Web Information Systems and Technologies, pages 385-392 ISBN: 978-989-758-718-4; ISSN: 2184-3252. DOI: 10.5220/0013061100003825
24. Wang, R. Y., Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers.Journal of Management Information Systems,12(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099
25. Xu, N., Chen, X., Luo, J.et al.Knowledge graph–large language model fusion approach for emergency knowledge recommendation in gas tunnels.Sci Rep16, 11438 (2026). https://doi.org/10.1038/s41598-026-39204-0
26. Yang J., Li Y., Huang Z. (2025).ReLoop: "Seeing Twice and Thinking Backward" via Closed-loop Training to Mitigate Hallucinations in Multimodal Understanding. In Findings of the Association for Computational Linguistics: EMNLP 2025, pages 4162–4179, Suzhou, China. Association for Computational Linguistics. 10.18653/v1/2025.findings-emnlp.222
27. Zhang, Wuyang & Zhang, Chenkai & Gu, Chuqiao & Kou, Jieren & Yuan, Hao & Fang, Xinyi & Duan, Xiaoman & Fang, Yajun. (2024). Hallucination in Large Language Models: From Mechanistic Understanding to Novel Control Frameworks. 1-36. 10.1109/UV63228.2024.11189226.