EDUKASIA: Jurnal Pendidikan dan Pembelajaran
Vol. 7, 1 (January-June, 2026), pp. 605-618
ISSN: 2721-1150; EISSN: 2721-1169; DOI: 10.62775/edukasia.v7i1.2141

Mitigating Hallucinations in Large Language Models via Retrieval Augmented Generation: A Systematic Review of n8n-Based Implementations

I Ketut Widhi Adnyana[1], Rosalin Theophilia Tayane[2], Fahmi[3], Freddy Wicaksono[4], Bagja Nugraha[5], Enjang Yusup Ali[6]

[1] Institut Teknologi dan Bisnis STIKOM Bali, Indonesia
[2] Universitas Sains dan Teknologi Jayapura, Indonesia
[3] Universitas Swadaya Gunung Jati, Indonesia
[4] Universitas Muhammadiyah Cirebon, Indonesia
[5] Universitas Singaperbangsa Karawang, Indonesia
[6] Universitas Pendidikan Indonesia, Indonesia

ARTICLE INFO

Keywords:

Large Language Models (LLMs); Natural Language Processing (NLP); Retrieval-Augmented Generation (RAG)

Article history:

Received 2026-02-25
Revised 2026-05-27
Accepted 2026-07-04

ABSTRACT

This study systematically examines hallucination phenomena in Large Language Models (LLMs), focusing on their characteristics, causal factors, and mitigation strategies through Retrieval-Augmented Generation (RAG) and low-code orchestration platforms such as n8n. Using a Systematic Literature Review (SLR) approach based on PRISMA 2020 guidelines, this study analysed 40 peer-reviewed articles published between 2020 and 2025 from major scientific databases. The findings reveal that hallucinations are multidimensional, consisting of factual, semantic, and contextual hallucinations influenced by static training data, probabilistic token prediction, prompt ambiguity, and insufficient validation mechanisms. The review further demonstrates that RAG significantly improves factual accuracy by integrating external retrieval systems with LLM generation processes. Recent innovations such as Hybrid Retrieval and GraphRAG enhance contextual relevance and knowledge representation. A major finding of this study is the identification of "Conflict of Information" between external retrieved data and internal LLM knowledge in automated RAG pipelines. Furthermore, this study proposes a novel conceptual framework and taxonomy for hallucination mitigation in low-code AI environments, integrating retrieval, validation, conflict resolution, and workflow orchestration mechanisms. These findings contribute to the development of more reliable, transparent, and scalable AI systems.

Corresponding Author:
I Ketut Widhi Adnyana
Institut Teknologi dan Bisnis STIKOM Bali, Indonesia; widhi_adnyana@stikom-bali.ac.id

1. INTRODUCTION

The development of Artificial Intelligence (AI) technology in recent years has progressed rapidly, particularly in the field of Natural Language Processing (NLP) with the emergence of Large Language Models (LLMs) (Bansal et al., 2025).. These models are capable of generating coherent and contextual text that closely resembles human communication patterns, leading to their widespread adoption across various sectors such as education, healthcare, customer service, legal services, and data-driven decision-making. This generative capability positions LLMs as a key innovation in digital transformation, particularly in improving efficiency and automating information-based processes (Li et al., 2025). However, despite their advanced capabilities, LLMs still face a fundamental challenge known as hallucination. In this context, hallucination refers to the generation of information that appears plausible and convincing but is factually inaccurate, unverifiable, or inconsistent with the provided context. This issue has become a major concern because it may lead to misinformation, reduce user trust, and create serious consequences in high-risk domains such as healthcare, law, and public policy.

One of the primary causes of hallucination is the limitation of LLMs in accessing dynamic, domain-specific, and up-to-date information (Ji et al.,2024). Most LLMs are trained on large-scale but static datasets that may not reflect recent developments or specialised contextual knowledge. In addition, LLMs optimise probabilistic word prediction rather than factual verification, resulting in outputs that are linguistically coherent but substantively incorrect (Farquhar et al., 2024). To address this limitation, Retrieval-Augmented Generation (RAG) has emerged as one of the most promising approaches (Abdallah & Beltagy, 2025). RAG combines the generative capabilities of LLMs with external retrieval systems, enabling models to access relevant and updated information sources before generating responses (Xu et al., 2025). As a result, RAG improves factual accuracy, transparency, and source traceability in AI-generated outputs (Ding et al., 2025). Previous studies have demonstrated that RAG significantly reduces hallucination in knowledge-intensive tasks; however, its effectiveness varies across application domains depending on data complexity, retrieval structure, and workflow requirements (Riza et al., 2025).

Several studies indicate that the implementation of RAG in the healthcare domain tends to prioritise factual precision, evidence traceability, and real-time validation because inaccurate information may directly affect clinical decision-making. In contrast, studies in the legal domain highlight challenges related to long-context retrieval, document ambiguity, and interpretative consistency across legal documents. These differences suggest that the effectiveness of RAG is highly dependent not only on retrieval quality but also on the orchestration and automation mechanisms used to manage multi-stage AI workflows. Despite the growing literature on RAG architectures, most studies focus primarily on retrieval accuracy and model performance, while limited attention has been given to workflow orchestration systems that enable flexible integration between retrieval pipelines, external databases, vector stores, APIs, and LLM components across different domains.

Recent developments in RAG architecture have introduced significant innovations, including hybrid retrieval mechanisms and knowledge graph integration through approaches such as Document GraphRAG (Gao et al., 2023). These approaches improve contextual understanding by enabling systems to capture relationships between documents in a more structured manner. However, the increasing complexity of RAG pipelines also creates challenges related to workflow coordination, modular integration, monitoring, and real-time automation. In this context, the use of workflow orchestration platforms such as n8n becomes increasingly relevant. Compared with proprietary orchestration platforms such as Zapier and Make, n8n offers several advantages, including open-source architecture, self-hosted deployment, higher workflow customisation, and flexible API integration. Furthermore, unlike more technically intensive orchestration frameworks such as Apache Airflow, n8n provides a low-code visual workflow environment that simplifies the integration of AI services, vector databases, retrieval systems, and real-time automation processes. These characteristics make n8n particularly suitable for adaptive RAG implementation across domains with diverse workflow requirements, including healthcare and legal services, where transparency, modularity, and integration flexibility are essential.

Nevertheless, the implementation of RAG-based systems still faces several challenges, including retrieval optimisation, domain adaptation, workflow scalability, and real-time evaluation of generated responses (Yu, & Cheng, 2025). In addition, conflicts may arise between externally retrieved information and the model's internal parametric knowledge, potentially affecting output consistency and reliability. These challenges demonstrate the importance of automation and orchestration systems capable of managing complex AI workflows in a structured and adaptable manner. Therefore, more systematic studies are required not only to understand the effectiveness of RAG in reducing hallucinations, but also to analyse how orchestration frameworks such as n8n can support more reliable, scalable, and domain-adaptive AI systems (Anh-Hoang et al., 2025).

Against this background, this study aims to conduct a systematic analysis of hallucination phenomena in Large Language Models, including their classification, causes, and impacts. Furthermore, this study examines various mitigation strategies with a particular focus on Retrieval-Augmented Generation and the role of n8n-based workflow orchestration in supporting adaptive and automated AI pipelines. By integrating discussions on domain-specific RAG challenges and workflow automation systems, this study seeks to contribute both theoretically and practically to the development of more accurate, transparent, scalable, and trustworthy AI systems.

2. METHODS

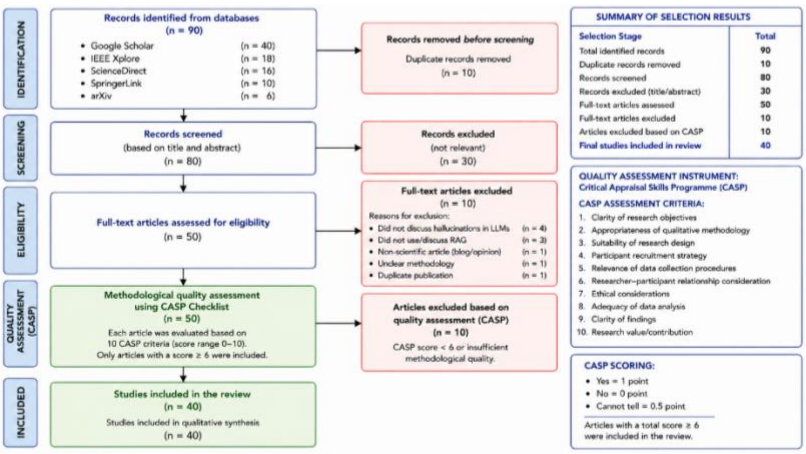
This study employs a Systematic Literature Review (SLR) approach to comprehensively analyse the phenomenon of hallucinations in Large Language Models (LLMs) and mitigation strategies using the Retrieval-Augmented Generation (RAG) approach. SLR is a systematic, structured, and transparent research method used to identify, evaluate, and synthesise findings from studies relevant to a specific research question. This approach was selected because it enables the development of a comprehensive and scientifically accountable synthesis of knowledge while allowing replication by other researchers (Moher, et al., 2021). The implementation of the SLR in this study follows the PRISMA 2020 guidelines established by the PRISMA framework, which emphasises transparency and completeness in systematic review reporting. These guidelines provide a structured framework consisting of four stages: identification, screening, eligibility assessment, and inclusion of relevant studies, thereby enhancing the validity and reliability of the review process (Mckenzie, et al., 2021).

The initial phase of this study involved formulating research questions focusing on the characteristics of hallucinations in LLMs, their underlying causes, and the effectiveness of RAG-based mitigation strategies. Subsequently, a systematic literature search was conducted across several academic databases, including Google Scholar, IEEE Xplore, ScienceDirect, SpringerLink, and arXiv. The search process used combinations of keywords such as "hallucinations in LLMs," "retrieval-augmented generation," "RAG mitigation," "LLM hallucination reduction," and "AI workflow orchestration." The search strategy was designed systematically to ensure comprehensive and relevant coverage of the literature and followed the principles of structured reporting recommended in PRISMA-S as an extension of PRISMA (Pielken et al., 2021).

The article selection process was conducted through several stages following the PRISMA framework. In the identification stage, all records retrieved from the selected databases were collected and duplicate articles were removed. During the screening stage, titles and abstracts were evaluated to exclude studies that were not relevant to the research objectives. Subsequently, the eligibility stage involved full-text assessment of the remaining articles based on predetermined inclusion and exclusion criteria. The inclusion criteria comprised peer-reviewed scientific articles published between 2020 and 2025, studies discussing LLMs, hallucinations, RAG, or AI orchestration systems, and articles available in full-text format. Meanwhile, exclusion criteria included non-scientific publications such as blogs or opinion articles, studies lacking methodological clarity, duplicate publications, and articles unrelated to the focus of this review. Finally, only studies meeting all eligibility criteria were included in the final analysis. The overall selection process is presented in the PRISMA Flowchart, which illustrates the number of records identified, screened, assessed for eligibility, and included in the review.

To ensure the quality and credibility of the selected studies, this review employed the Critical Appraisal Skills Programme (CASP) checklist as a quality assessment instrument. The CASP framework was used to evaluate methodological rigour, clarity of research objectives, appropriateness of research design, validity of data collection procedures, transparency of analysis, and relevance of findings to the review objectives. Each selected article was critically assessed to ensure that only studies with adequate methodological quality were included in the synthesis process. The use of CASP enhanced the reliability and transparency of the review by minimising the inclusion of low-quality or potentially biased studies.

Following the quality assessment stage, eligible studies were analysed using qualitative content analysis techniques. The findings were grouped into several thematic categories, including types of hallucinations, causal factors, mitigation strategies, domain specific implementation challenges, and the role of orchestration systems such as n8n in supporting RAG workflows. The synthesis process was conducted narratively to identify patterns, emerging trends, research gaps, and relationships between studies across different domains such as healthcare and legal services. This approach aligns with SLR principles that emphasise systematic integration of scientific evidence to achieve a deeper understanding of a research phenomenon. By applying a PRISMA guided SLR method combined with CASP based quality assessment, this study is expected to produce a comprehensive, transparent, and methodologically rigorous review of the effectiveness of Retrieval Augmented Generation in mitigating hallucinations in Large Language Models.



Source : Adapted From the PRISMA 2020 Statement (Moher, et al., 2021)

3. FINDINGS AND DISCUSSION

Findings

The findings of this study were obtained through the analysis of studies that met the inclusion criteria using a Systematic Literature Review (SLR) approach. Overall, the review demonstrates that hallucination in Large Language Models (LLMs) is a complex and multidimensional phenomenon influenced by model architecture, data quality, retrieval mechanisms, and human interaction processes. Furthermore, Retrieval-Augmented Generation (RAG) has emerged as one of the most effective approaches for mitigating hallucinations because it integrates generative models with external retrieval systems capable of providing relevant and up-to-date information (Lavrinovics et al., 2025) (Rahman et al., 2026).

Based on the synthesis of the reviewed studies, the findings are categorised into four main aspects: (1) characteristics and types of hallucinations, (2) factors causing hallucinations, (3) effectiveness of

RAG as a mitigation strategy, and (4) evaluation of automation platforms such as n8n in supporting hallucination mitigation workflows.

Table 1 Summary of Literature Review Findings

Key Aspects	Sub-Aspect	Description of Findings	
Characteristics of Hallucinations	Factual Hallucinations	The model generates information that is factually incorrect or unverifiable although it appears convincing	(Ji et al., 2023) dan (Tonmoy et al., 2024)
	Semantic Hallucinations	Responses are linguistically correct but semantically inappropriate or irrelevant	(Ji et al., 2023)
	Contextual Hallucinations	Inconsistency between generated responses and contextual instructions	(Tonmoy et al., 2024)
Causal Factors	Static Data	Training data do not reflect real-time or domain-specific information	(Huang et al., 2025) dan (Qi et al., 2024)
	Probabilistic Mechanism	Token prediction prioritises probability rather than factual verification	(Béchard & Ayala, 2024)
	Prompt Quality	Ambiguous prompts reduce response precision	(Lewis et al., 2020)
	Contextual Inconsistency	Misalignment between prompts and retrieved context	(Holtzman et al., 2020)
	Lack of Verification	Absence of validation mechanisms increases hallucination risk	(Gao et al., 2023)
Effectiveness of RAG	RAG Mechanism	Integration of retrieval systems with LLMs to provide external evidence	(Zhang & Zhang, 2025) ; (Niu et al., 2024)
	Hallucination Reduction	RAG significantly improves factual accuracy and reduces hallucinations	(Purwanto et al., 2026)
	Architectural Innovation	Hybrid Retrieval and GraphRAG improve contextual understanding	(Formal et al., 2022)

	Retrieval Challenge	Retrieval systems may fail to select the most relevant information	(Kovács & Recski, 2025)
	Conflict of Information	Contradictions between external retrieval results and internal LLM knowledge	(Meister et al., 2023)
Automation Platform Evaluation	Workflow Automation	n8n enables automated orchestration between retrieval, validation, and generation modules	(Chen, 2024)
	Real-Time Integration	Automation platforms support dynamic API integration and live data synchronisation	(Yeh et al., 2026)
	Low-Code Flexibility	Low-code architecture simplifies AI workflow development	(Pan et al., 2024).
	Validation Layer	n8n supports additional verification nodes before final output generation	(Pan et al., 2024) dan (Chen, 2024).

The findings indicate that hallucinations in LLMs can be understood through factual, semantic, and contextual dimensions. Factual hallucinations occur when models generate false or unverifiable information, semantic hallucinations arise when outputs are linguistically plausible but conceptually inaccurate, and contextual hallucinations occur when the model fails to maintain consistency with user instructions or retrieved information.

The review also demonstrates that hallucinations are caused by multiple interconnected factors. Static datasets prevent models from accessing recent information, while probabilistic token generation mechanisms prioritise linguistic fluency over factual validity. Furthermore, unclear prompts, contextual mismatches, and the absence of verification systems increase the likelihood of hallucinated outputs. These findings suggest that hallucination is not solely caused by model limitations, but also by weaknesses in data management, interaction design, and retrieval processes.

Regarding mitigation strategies, RAG has shown substantial effectiveness in reducing hallucinations because it enables LLMs to access external evidence during response generation. Architectural innovations such as Hybrid Retrieval and GraphRAG further improve contextual relevance and semantic relationships between information sources. Nevertheless, retrieval optimisation remains challenging because systems may retrieve incomplete or less relevant information.

Literature Evaluation on the Use of n8n-Based Automation Platforms

The literature review further reveals that low-code automation platforms such as n8n play an increasingly important role in operationalising RAG systems. Studies indicate that n8n enables automated orchestration between multiple AI components, including document retrieval, API integration, vector databases, validation layers, and LLM response generation. This automation capability supports real-time workflows and improves the scalability of AI systems.

Several studies also highlight that n8n supports modular workflow architectures where retrieval, filtering, reranking, and validation processes can be implemented sequentially. This enables developers to create multi-stage hallucination mitigation pipelines rather than relying solely on a single LLM generation process. In practice, n8n is frequently used to integrate external APIs, semantic search systems, vector databases, and fact-checking modules within one coordinated workflow.

However, the review also identifies several technical challenges associated with automation-based RAG implementation. One of the most critical issues is "Conflict of Information," where information retrieved from external sources contradicts the internal knowledge embedded in the LLM parameters. This inconsistency may cause the model to produce blended or contradictory responses, particularly when the retrieval source is recent while the LLM relies on outdated pretraining knowledge.

In addition, workflow automation platforms still face challenges related to retrieval latency, API reliability, source prioritisation, and response consistency. These findings suggest that although n8n significantly improves orchestration flexibility and automation efficiency, effective hallucination mitigation still requires robust validation mechanisms, source ranking strategies, and contextual verification processes.

Discussion

The findings of this study indicate that hallucination in Large Language Models (LLMs) is not merely a technical limitation but also an epistemological problem related to how language models construct knowledge probabilistically without genuine factual understanding. The classification of hallucinations into factual, semantic, and contextual categories demonstrates that hallucination occurs at multiple layers of linguistic representation rather than as a singular phenomenon (Bender et al., 2021). These findings reinforce the argument that LLMs fundamentally operate as probabilistic prediction systems rather than systems capable of fully understanding reality or truth (Sun et al., 2025). Consequently, hallucinations emerge because linguistic coherence is often prioritised over factual verification (Bommasani et al., 2022) (Achiam et al., 2024).

The findings also suggest that hallucinations are multifactorial. Static training data limit the model's ability to access updated knowledge, while probabilistic token generation encourages the production of fluent yet potentially inaccurate information (Béchard & Ayala, 2024). Furthermore, ambiguous prompts, contextual inconsistencies, and the absence of systematic verification mechanisms further increase hallucination risks. These findings imply that hallucination mitigation requires not only improvements in model architecture but also stronger retrieval systems, validation layers, and interaction design mechanisms.

Within this context, Retrieval-Augmented Generation (RAG) emerges as a highly relevant solution because it combines generative capabilities with evidence-based retrieval processes (Meister et al., 2023); (Maulana & Abdillah, 2022). By integrating external data retrieval systems, LLMs no longer rely exclusively on internal pretrained knowledge but instead generate responses supported by verifiable evidence (Vishwakarma, 2025). The findings showing reduced hallucination rates through RAG implementation strengthen the argument that retrieval integration significantly improves AI reliability and factual accuracy (Purwanto et al., 2026).

Furthermore, recent developments such as Hybrid Retrieval and GraphRAG indicate a transition from conventional retrieval approaches toward more contextual and structured architectures. Hybrid Retrieval combines keyword-based retrieval with semantic embeddings, enabling more comprehensive relevance matching (Thakur et al., 2021). Meanwhile, GraphRAG enhances contextual reasoning by integrating knowledge graph structures that represent relationships between entities and concepts (Formal et al., 2022). These developments suggest that hallucination mitigation effectiveness depends not only on retrieval existence but also on retrieval quality, contextual relevance, and knowledge representation structures.

Discussion of n8n-Based Automation Platforms

An important finding of this review is the growing role of low-code automation platforms such as n8n in implementing RAG-based hallucination mitigation systems. Unlike conventional AI pipelines that require extensive programming expertise, n8n enables modular orchestration between retrieval systems, APIs, vector databases, validation layers, and LLM interfaces through low-code workflows. This significantly improves implementation flexibility and accelerates the deployment of AI systems in practical environments.

The reviewed studies demonstrate that n8n functions not only as an automation tool but also as an orchestration layer that coordinates information retrieval, validation, reranking, and response generation processes in real time. Through node-based workflows, developers can integrate multiple verification stages before the final response is generated. This architecture supports evidence-based AI generation because external information can be filtered, ranked, and validated before being processed by the LLM.

However, despite these advantages, several technical challenges remain unresolved. The most significant challenge identified across the reviewed studies is the phenomenon of "Conflict of Information." This occurs when external information retrieved through APIs, databases, or web sources contradicts the knowledge already embedded within the LLM during pretraining. In many cases, the LLM may prioritise its internal parametric knowledge even when retrieved external evidence is more recent or factually accurate. As a result, the generated response may contain inconsistencies, mixed factual claims, or partially hallucinated content.

This conflict becomes particularly problematic in real-time systems where external data sources are continuously updated while the LLM's pretrained knowledge remains static. In n8n-based workflows, such conflicts may propagate across multiple automation stages if no validation or source-prioritisation mechanism is implemented. Consequently, the effectiveness of RAG systems depends not only on retrieval accuracy but also on the system's ability to resolve contradictions between retrieved evidence and internal model knowledge.

Additional challenges identified in the literature include retrieval latency, API inconsistency, source reliability, workflow complexity, and output evaluation difficulties. Since low-code workflows integrate multiple external services simultaneously, overall system reliability becomes highly dependent on workflow orchestration quality and validation mechanisms. These findings suggest that automation alone is insufficient for hallucination mitigation unless accompanied by robust conflict-resolution and verification strategies.

Proposed Conceptual Framework for Hallucination Mitigation on Low-Code Platforms

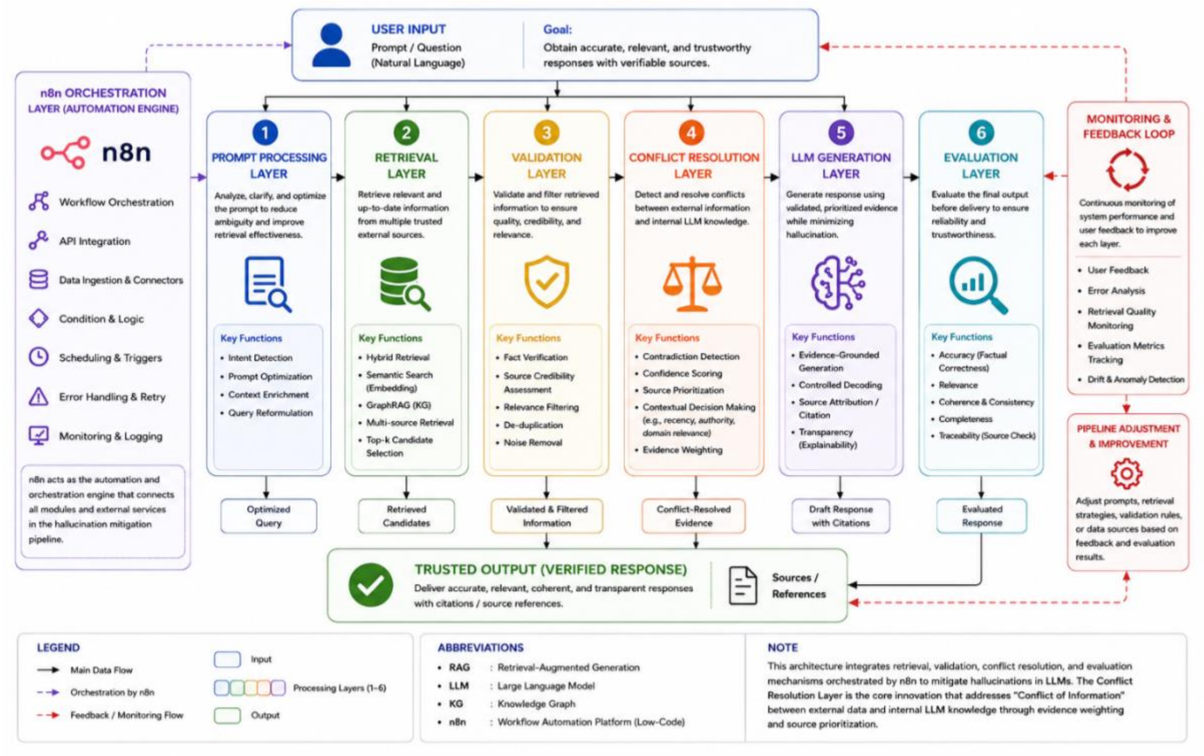


Figure 1 : Conceptual Architecture for Hallucination Mitigation in n8n Based RAG System

The framework demonstrates that hallucination mitigation should not rely solely on retrieval mechanisms, but rather on a multi-layered architecture involving validation, contextual prioritisation, and conflict resolution processes. In this model, n8n acts as an orchestration platform that coordinates interactions between retrieval modules, validation systems, APIs, vector databases, and LLM generation pipelines.

Taxonomy of Hallucination Mitigation Strategies on Low-Code Platforms

Based on the SLR findings, this study also proposes a taxonomy of hallucination mitigation strategies for low-code AI environments.

Table 2. Taxonomy of Hallucination Mitigation Strategies

Taxonomy Category	Main Function	Role in Mitigation	Implementation Example
Prompt-Based Mitigation	Optimising prompts and instructions	Reduces ambiguity and improves contextual clarity	Prompt engineering and instruction refinement
Retrieval-Based Mitigation	Retrieving external information	Improves factual accuracy using updated sources	RAG, Hybrid Retrieval, GraphRAG
Validation-Based Mitigation	Verifying retrieved information	Prevents false or irrelevant outputs	Fact-checking and source validation

Conflict-Resolution Mitigation	Resolving contradictions between sources	Minimises inconsistencies between external and internal knowledge	Confidence scoring and source prioritisation
Workflow-Orchestration Mitigation	Coordinating AI pipelines	Enables automated and modular mitigation workflows	n8n automation pipelines
Human-in-the-Loop Mitigation	Incorporating human supervision	Enhances reliability and ethical oversight	Expert review and manual verification

The taxonomy indicates that hallucination mitigation requires complementary strategies operating across multiple layers of the AI workflow. Consequently, effective mitigation cannot depend on a single model improvement technique but instead requires integrated orchestration between retrieval, validation, automation, and human oversight.

Scientific Contribution and Novelty

This study contributes theoretically and practically to the development of hallucination mitigation research in LLM systems. Unlike previous studies that primarily focused on retrieval accuracy or standalone RAG architectures, this study proposes a workflow-oriented conceptual framework integrating retrieval systems, validation layers, conflict-resolution mechanisms, and low-code orchestration using n8n for hallucination mitigation.

Furthermore, this study introduces a taxonomy of hallucination mitigation strategies specifically designed for low-code AI environments. This contribution extends existing RAG discussions by emphasising the importance of workflow orchestration, real-time validation, and information conflict management in AI automation systems.

Practically, the proposed framework provides guidance for AI developers, automation engineers, and NLP researchers in designing more transparent, reliable, and scalable AI systems. The findings also highlight that future hallucination mitigation research should focus not only on model improvement but also on orchestration architectures capable of integrating retrieval, validation, and adaptive conflict-resolution mechanisms within automated AI ecosystems.

Overall, this study confirms that mitigating hallucinations in LLMs requires a multidimensional approach involving retrieval optimisation, evidence validation, contextual reasoning, and workflow orchestration. Therefore, future AI systems should prioritise transparency, factual grounding, and adaptive verification mechanisms to improve the trustworthiness and reliability of AI-generated outputs.

4. CONCLUSION

Based on the research findings and discussion, it can be concluded that the phenomenon of hallucinations in Large Language Models (LLMs) is a complex and multidimensional issue, encompassing factual, semantic and contextual aspects. Hallucinations are not only caused by the technical limitations of the model, but are also influenced by the quality of training data, probability-based prediction mechanisms, and user interaction via prompts. Consequently, this problem is multifactorial and requires a comprehensive approach to its management. The Retrieval-Augmented Generation (RAG) approach has proven to be one of the most effective strategies for mitigating hallucinations, as it integrates the model's generative capabilities with access to relevant and up-to-date external information sources. The implementation of RAG, particularly through developments such as hybrid retrieval and GraphRAG, demonstrates improvements in the accuracy, relevance, and validity of the generated information. Furthermore, the use of platforms such as n8n strengthens the implementation aspect by enabling more automated, flexible, and real-time system integration. Nevertheless, RAG still faces various challenges, such as optimising the retrieval process, potential information conflicts, and the complexity of evaluating output quality. Therefore, further development is required, encompassing improvements in data quality, refinements to the system architecture, and the strengthening of more holistic evaluation methods. Overall, this research confirms that mitigating hallucinations in LLMs cannot be achieved through a single approach, but rather requires the integration of various complementary strategies. RAG presents a promising solution, yet continuous development remains necessary to realise a more accurate, transparent, and trustworthy AI system.

Acknowledgments: The authors would like to express their sincere gratitude to all parties who have contributed to the completion of this study. Special thanks are extended to colleagues and academic peers for their valuable insights, constructive feedback, and continuous support throughout the research process. The authors also appreciate the institutional support provided, including access to relevant resources and facilities that enabled the successful completion of this research.

REFERENCES

Abdallah, M., & El - Beltagy, S. (2025). HalluSearch at SemEval-2025 Task 3: A Search-Enhanced RAG Pipeline for Hallucination Detection. *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*, 1436–1441. https://aclanthology.org/2025.semeval-1.189/

Anh-Hoang, D., Tran, V., & Nguyen, L. M. (2025). Survey and analysis of hallucinations in large language models: attribution to prompting strategies or model behavior. *Frontiers in Artificial Intelligence*, *8*(September), 1–21. https://doi.org/10.3389/frai.2025.1622292

Bansal, R., Reena Chandra, & Karan Lulla. (2025). Understanding and Mitigating Strategies for Large Language Model (LLMs) Hallucinations in HR Chatbots. *International Journal of Computational and Experimental Science and Engineering*, *11*(3), 4126–4137. https://doi.org/10.22399/ijcesen.2471

Béchard, P., & Ayala, O. M. (2024). Reducing hallucination in structured outputs via Retrieval-Augmented Generation. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2024*, *6*, 228–238. https://doi.org/10.18653/v1/2024.naacl-industry.19

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, *April*, 610–623. https://doi.org/10.1145/3442188.3445922

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2022). *On the Opportunities and Risks of Foundation Models*. 1–214.

Chen, H. (2024). Optimal Rate of Convergence for Vector-valued Wiener-Itô Integral. *Alea (Rio de Janeiro)*, *21*(11731009), 179–214. https://doi.org/10.30757/ALEA.v21-08

Ding, H., Pang, L., Wei, Z., Shen, H., & Cheng, X. (2025). Rowen: Adaptive Retrieval-Augmented Generation for Hallucination Mitigation in LLMs. In *SIGIR-AP 2025 - Proceedings of the 2025 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region* (Vol. 1, Issue 1). arXiv. https://doi.org/10.1145/3767695.3769500

Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, *632*, 1–10. https://doi.org/https://doi.org/10.1038/s41586-024-07421-4

Formal, T., Lassance, C., Piwowarski, B., & Clinchant, S. (2022). From Distillation to Hard Negative Sampling: Making Sparse Neural IR Models More Effective. In *SIGIR 2022 - Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Vol. 1, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3477495.3531857

Gao, Y., et al. (2023). RAG vs Fine-tuning: Pipelines, Tradeoffs, and a Case Study on Agriculture. *ArXiv*

Preprint ArXiv. https://doi.org/https://doi.org/10.48550/arXiv.2401.08406

Gao, L., Dai, Z., Pasupat, P., Chen, A., Chaganty, A. T., Fan, Y., Zhao, V. Y., Lao, N., Lee, H., Juan, D. C., & Guu, K. (2023). RARR: Researching and Revising What Language Models Say, Using Language Models. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, *1*, 16477–16508. https://doi.org/10.18653/v1/2023.acl-long.910

Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2020). THE CURIOUS CASE OF NEURAL TEXT DeGENERATION. *8th International Conference on Learning Representations, ICLR 2020*.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2025). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Transactions on Information Systems*, *43*(2), 1–58. https://doi.org/10.1145/3703155

Ji, Z., et al. (2024). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, *55*(12), 1–38. https://doi.org/https://doi.org/10.1145/3571730

Ji, Z., Yu, T., Xu, Y., Lee, N., Ishii, E., & Fung, P. (2023). Towards Mitigating Hallucination in Large Language Models via Self-Reflection. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 1827–1843. https://doi.org/10.18653/v1/2023.findings-emnlp.123

Kovács, Á., & Recski, G. (2025). *LettuceDetect: A Hallucination Detection Framework for RAG Applications*. http://arxiv.org/abs/2502.17125

Lavrinovics, E., Biswas, R., Bjerva, J., & Hose, K. (2025). Knowledge Graphs, Large Language Models, and Hallucinations: An NLP Perspective. *Journal of Web Semantics*, *85*, 100844. https://doi.org/https://doi.org/10.1016/j.websem.2024.100844

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*, *2020-Decem*.

Li, Y., Fu, X., Verma, G., Buitelaar, P., & Liu, M. (2025). Mitigating Hallucination in Large Language Models (LLMs): An Application-Oriented Survey on RAG, Reasoning, and Agentic Systems. *Arxiv*, 1–25. https://doi.org/https://doi.org/10.48550/arXiv.2510.24476

Maulana, T. I., & Abdillah, A. R. (2022). Pemanfaatan sistem temu kembali informasi dalam pencarian dokumen menggunakan vektor space model. *SINTESIA: Jurnal Sistem Dan Teknologi Informasi Indonesia*, *1*(2), 89–95.

Meister, C., Pimentel, T., Wiher, G., & Cotterell, R. (2023). Locally Typical Sampling. *Transactions of the Association for Computational Linguistics*, *11*, 102–121. https://doi.org/10.1162/tacl_a_00536

Niu, C., Wu, Y., Zhu, J., Xu, S., Shum, K., Zhong, R., Song, J., & Zhang, T. (2024). RAGTruth: A Hallucination Corpus for Developing Trustworthy Retrieval-Augmented Language Models. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, *1*, 10862–10878. https://doi.org/10.18653/v1/2024.acl-long.585

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2024). *GPT-4 Technical Report*. *4*, 1–100.

Page, M. J., Mckenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-wilson, E., Mcdonald, S., … Moher, D. (2021). The PRISMA 2020 statement : an updated guideline for reporting systematic reviews. *BMJ*. https://doi.org/10.1136/bmj.n71

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., Mcdonald, S., … Mckenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *The BMJ*, *372*. https://doi.org/10.1136/bmj.n160

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying Large Language Models and

Knowledge Graphs: A Roadmap. *IEEE Transactions on Knowledge and Data Engineering*, *36*(7), 3580–3599. https://doi.org/10.1109/TKDE.2024.3352100

Pielken, H. J., Urbanitz, D., Koch, P., & van de Loo, J. (2021). Immunological monitoring in remission acute myeloid leukemia during maintenance therapy. *Haematology and Blood Transfusion*, *30*, 385–386. https://doi.org/10.1007/978-3-642-71213-5_65

Purwanto, Wijaya, F., Bernadisman, D., Sutar, & Amrullah, M. (2026). Aplikasi Interaktif Berbasis R Studio Untuk Prediksi Nilai Ujian Mahasiswa Menggunakan Algoritma Random Forest Dengan Evaluasi Multi-Variabel. *Jurnal Sistem Informasi Dan Teknologi (SINTEK)*, *6*, 127–135. https://doi.org/10.56995/sintek.v6i1.255

Qi, F., Hou, Y., Lin, N., Bao, S., & Xu, N. (2024). A Survey of Testing Techniques Based on Large Language Models. *ACM International Conference Proceeding Series*, *0*, 280–284. https://doi.org/10.1145/3675249.3675298

Rahman, S. S., Islam, M. A., Alam, M. M., Zeba, M., Rahman, M. A., Chowa, S. S., Raiaan, M. A. K., & Azam, S. (2026). Hallucination to truth: a review of fact-checking and factuality evaluation in large language models. *Artificial Intelligence Review*, *59*(2). https://doi.org/10.1007/s10462-025-11454-w

Riza, F., Jamal Al Din, S., Yusuf Al Afghani, D., Setiabudi, R., & Teknologi Budi Utomo, I. (2025). LLM-based self-related local ai agent design through n8n orchestration for conversational memory on rag. *Journal of Information Technology and Computer Science (INTECOMS)*, *8*(3), 2025.

Sun, Z., Zang, X., Zheng, K., Xu, J., Zhang, X., Yu, W., Song, Y., & Li, H. (2025). Redeep: Detecting Hallucination in Retrieval-Augmented Generation Via Mechanistic Interpretability. *13th International Conference on Learning Representations, ICLR 2025*, 102578–102607.

Thakur, N., Reimers, N., Rücklé, A., Srivastava, A., & Gurevych, I. (2021). BEIR: A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models. *Advances in Neural Information Processing Systems*.

Tonmoy, S. M. T. I., Zaman, S. M. M., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. (2024). *A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models*.

Vishwakarma, V. K. (2025). Designing Agent-Native Automation in n8n: A Scalable Framework Integrating AI Agents, Multi-Agent Systems, and Retrieval-Augmented Generation. *International Journal for Research in Applied Science and Engineering Technology*, *13*(11), 1044–1054. https://doi.org/10.22214/ijraset.2025.75231

Xu, S., Yan, Z., Dai, C., & Wu, F. (2025). MEGA-RAG: a retrieval-augmented generation framework with multi-evidence guided answer refinement for mitigating hallucinations of LLMs in public health. *Frontiers in Public Health*, *13*. https://doi.org/10.3389/fpubh.2025.1635381

Yeh, S., Li, S., & Mallick, T. (2026). LUMINA: Detecting Hallucinations in RAG System with Context-Knowledge Signals. *Arxiv*, 1–19. https://doi.org/https://doi.org/10.48550/arXiv.2509.21875

Yu, W., Yu, X., Zhang, Y., Li, X., & Cheng, N. (2025). REPLUG: Retrieval-augmented black-box language models. *IEEE Transactions on Knowledge and Data Engineering*. https://doi.org/https://doi.org/10.1109/TKDE.2025.3356789

Zhang, W., & Zhang, J. (2025). Hallucination Mitigation for Retrieval-Augmented Large Language Models: A Review. *Mathematics*, *13*, 856. https://doi.org/10.3390/math13050856