International Journal of Combinatorial Optimization Problems and Informatics, 16(4), Sep-Dec 2025, 345-363. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2025.v16i4.1003

www.editada.org

Application of Retrieval-Augmented Generation Systems in Software Engineering Education: An Approach Based on Generative AI and DevOps

Yazmin Valeria Morales-Martínez[1], *Blanca Dina Valenzuela-Robles*[1], *René Santaolaya-Salgado*[1], *Gabriel González-Serna*[1], *Noé Alejandro Castro-Sánchez*[1], *Mirna Muñoz*[2]
[1] TecNM, Centro Nacional de Investigación y Desarrollo Tecnológico, México
[2] Centro de Investigación en Matemáticas, México
{d21ce063, blanca.vr, rene.ss, gabriel.gs, noe.cs}@cenidet.tecnm.mx, mirna.munoz@cimat.mx

Abstract. This paper presents a systematic literature review of the application of retrieval-augmented generation (RAG) systems in educational settings, with a focus on teaching software engineering and related computing disciplines. Drawing on case studies, academic experiments, and surveys of teachers and students, it provides an overview of the current landscape, highlighting perceptions, reported effectiveness, and the technology's impact in academia. Based on an analysis of 71 selected scientific papers, the review synthesises evidence on the extent to which RAG systems mitigate hallucinations and improve human–AI interaction. In addition, it suggests that many approaches discussed across studies could be strategically aligned with the integration of DevOps practices and RAG, enhancing their use through automation, continuous improvement, and the agile adoption of technologies within educational processes.

Keywords: Academia, DevOps, Generative AI, Retrieval-Augmented Generation, Software Engineering.

Article Info
Received February 25, 2025
Accepted July 6, 2025

1 Introduction

In recent years, advances in artificial intelligence, particularly in Generative AI (GAI or GenAI), have significantly impacted daily life. These technologies have become increasingly accessible and widely adopted (Zhu, 2024; Jawalkar et al., 2024). GenAI enables the creation of content such as text, images, audio, and video through pre-trained models (Uysal, 2025). Among these developments, Large Language Models (LLMs) have emerged as key tools across various domains, including communication, healthcare, education, and entertainment (Ardimento et al., 2024; Zúñiga Sánchez, 2024). However, due to their reliance on static datasets, LLMs often generate outdated or inaccurate responses, leading to hallucinations (Cooper & Klymkowsky, 2024; Jeong, 2024).

To overcome these limitations, Retrieval-Augmented Generation (RAG) systems have been developed. RAG integrates information retrieval with generative capabilities, enabling real-time verification and reducing hallucinations (Jeong, 2024; Singh et al., 2025; Zhang et al., 2025). These systems significantly enhance the performance of LLMs. Studies report improvements in accuracy, from 75% to 86%, when domain-specific tuning is applied (Balaguer et al., 2024). Moreover, by addressing the limitations of static knowledge, RAG systems enable more dynamic, context-aware responses, provided they rely on high-quality data sources (Jawalkar et al., 2024).

Large language models such as ChatGPT and Llama-2 were launched in 2023. In 2024, the era of RAG systems, chatbots, and AI agents became consolidated. Chatbots are designed for conversational interaction via text or voice, while AI agents are capable of making decisions and learning from their environment. Both have gained popularity comparable to RAG systems by enhancing the user experience (Deng et al., 2024; Jeong, 2024; Ciolacu et al., 2024).

RAG systems are also being explored in education to optimize learning and automate pedagogical processes. These technologies support the adoption of innovative teaching practices (Rajeshwari et al., 2024; Ciucu et al., 2019). Their use is increasingly common in the software development industry, particularly within the DevOps model. Since its consolidation in 2009, DevOps has promoted practices that integrate development and operations through automation (Jabbari et al., 2016). These practices align with current industry demands and are especially attractive to students (Mota et al., 2024).

DevOps promotes continuous improvement and agile technology deployment. As such, it provides a valuable framework for modernizing education and preparing students for a dynamic labor market. An experiment involving 30 students who participated in a role-play-based teaching model for DevOps reported promising outcomes. Approximately 90% of participants found the approach more engaging than traditional classes, and 93.3% agreed that it significantly contributed to their learning. Qualitative feedback also highlighted positive perceptions regarding collaborative learning and the integration of real-world tools (Mota et al., 2024).

This study focuses on the growing interest in Software Engineering Education (SEE), which increasingly demands innovative methods to enhance learning experiences and prepare students for industry challenges (Yabaku et al., 2024). In response, this review explores the state of the art in the application of RAG systems in education and related fields of computing and technology. The analysis of 71 scientific articles reveals how the integration of GenAI, particularly RAG systems, along with DevOps practices, can improve academic training and the development of software projects.

2 Research Method

This work presents a systematic literature review (SLR) in the field of Software Engineering (SE), developed based on two strategies for article retrieval: (i) exploration of specialized digital databases following the guidelines established by Kitchenham et al. (2009) for conducting SLRs, and (ii) the snowballing technique (Wohlin, 2014). The exploration technique in specialized databases involves rigorously identifying, evaluating, and synthesizing existing evidence on a specific topic (Kitchenham et al., 2009). The snowballing technique expands the study corpus by identifying relevant studies from the references of already selected publications, thus enabling the discovery of additional material of interest (Wohlin, 2014).

The objective of this study is to analyze the impact of RAG systems in educational environments, with a focus on Software Engineering Education, through a systematic literature review. The study aims to identify their effectiveness in mitigating hallucinations, improving Human-AI interaction, and their potential integration with DevOps practices for teaching and learning.

To achieve the objectives set for this study, the following research questions were defined: *RQ1: What is the impact of using and integrating RAG systems within academic environments in general, and specifically in software engineering? RQ2: How is the mitigation of hallucinations generated by LLMs through the use of RAG reported in education? RQ3: How can the synergy between humans and AI in software development be improved to maximize productivity? RQ4: In which phases of the software lifecycle has the use of RAG systems been documented, and how could they be integrated into DevOps stages?*

In conducting this study, the search strategy combined two complementary approaches: specialized database exploration (Kitchenham et al., 2009) and the snowballing technique (Wohlin, 2014). The first strategy involved searches in the databases described below: ACM Digital Library, IEEE Xplore, Springer Link, Scopus, Wiley Online Library, and arXiv. These platforms were chosen for their inclusion of studies published in leading conferences and journals in the field of education. To construct the search query, the central themes of the research were SEE, SE, teaching, Human-AI, DevOps, RAG systems and GenAI were combined using related terms. Minor adjustments were made to the search string to maximize both the quality and quantity of the retrieved works. Zotero (Zotero, 2025) was used for citation and reference generation.

The final search query was: *((course OR teaching OR education OR academia) AND ("software engineering")) AND (DevOps OR "continuous integration" OR "continuous delivery" OR CI OR CD) AND (RAG OR "retrieval-augmented generation") AND ((GAI OR GenAI OR “AI Generative”) OR (Human-AI OR H-AI))*

The second strategy applied a bidirectional and recursive snowballing process. In the backward snowballing phase, references cited in the initially identified primary studies were reviewed. In the forward snowballing phase, newer studies that cited those primary works were examined.

To ensure the relevance, rigor, and alignment of the selected literature with the objectives of this study, a set of selection, exclusion, and quality assessment criteria was established. These criteria guided the identification of primary studies and maintained a consistent evaluation process. Tables 1, 2, and 3 summarize each set of criteria applied throughout the review.

Table 1. Selection criteria for primary studies

No.	Criterion
1	Scientific studies published in the last five years (2020–2025) to ensure relevance and up-to-date information.
2	Case studies addressing RAG in education or software engineering (SE).
3	Studies published in journals and conferences indexed in recognized academic databases.
4	Documents available in English or Spanish.
5	Full-text, open-access studies available without paywall restrictions.

Table 2. Exclusion criteria

No.	Criterion
1	Studies with a purely technical focus and no relation to educational contexts.
2	Publications with outdated information (older than five years, except for relevant exceptions).
3	Studies that do not provide a clear contribution to this study or present ambiguities in their study or results.
4	Opinion pieces, blogs, or papers that have not undergone peer review.
5	Documents in languages inaccessible for the study (e.g., Chinese or Russian, without an available translation).

Table 3. Quality assessment criteria

No.	Criterion
1	Studies must clearly state their objectives and research questions.
2	Studies must describe their methodology in sufficient detail to allow evaluation and potential replication.
3	Studies must provide a significant contribution to this study or explicitly address at least one of the defined research questions.
4	Preference was given to studies that provided access to data, tools, or frameworks relevant to the implementation of RAG systems.
5	When available, citation count and relevance to current educational or technological trends were considered, as well as the article's contribution to the academic discourse on AI in education and SE.
6	Peer Review and Blind Review Status: Preference was also given to studies published in peer-reviewed venues, particularly those subjected to double-blind review processes.

3 Criteria for Classifying Primary Studies

A classification based on study type was applied, as it was considered relevant to addressing the research questions posed in this study (Table 4). Six categories were established. This distribution highlights a strong emphasis on applied and pedagogical research rather than purely qualitative methodologies. The next step of the review involved the reading and selection of primary studies for the systematic literature review (SLR), following the criteria defined in Section 2.4. Table 5 summarizes the process applied, while the consulted databases included ACM Digital Library, IEEE Xplore, Springer Link, Scopus, Wiley Online Library, and ArXiv.

Table 4. Categorization of primary studies

Category	Description	Quantity
Interviews	Interviews collect qualitative information about participants' experiences and perceptions.	1
Specific case studies	Analyze in depth a particular situation, company, project, or implementation.	6
Experimental/empirical	Include controlled trials, lab simulations, and computational experiments designed to validate hypotheses and assess system behavior.	27
Reviews	Encompass bibliographic studies, systematic mappings, literature reviews, and exploratory analyses to synthesize existing knowledge.	9
Theoretical/ conceptual	Develop new models and theoretical frameworks without direct empirical validation.	15
Education/ methodologies/ teaching	Examine pedagogical methods, instructional strategies, and training approaches.	22

The selection procedure in each source followed a defined sequence: initial screening, removal of duplicate entries, application of selection criteria based on titles and abstracts, and finally, a full-text review. As a result, 71 primary studies were selected: 44 from the digital libraries mentioned above and 27 through the snowballing technique. Finally, the answers to the research questions were derived from the selected primary studies and are presented in the results, based on the methodology previously described.

Table 5. Search stages for selecting primary studies

Source of the studies	Initial selection	Duplicate removal	Application of criteria (Titles and Abstracts)	Application of criteria (Full Reading)
Digital Libraries	5654	2451	62	44
Snowballing backward	70	57	14	10
Snowballing forward	75	8	30	17
Total	5799	2516	106	71

4 Results

This section presents the findings derived from the analysis of the 71 primary studies included in this literature review. The analysis of the year of publication reveals clear trends in research activity. Fig. 1 presents the annual distribution of the selected studies. A significant increase is observed in 2024, with more than 41 publications. This reflects a growing academic interest in GenAI and RAG systems. The upward trend suggests that research in this field will continue expanding, with even greater output expected in 2025.

The rise is especially notable in studies focused on RAG systems, chatbots, and AI agents applied to education. Their increasing adoption indicates a shift in academic practices toward intelligent and interactive technologies. This scenario offers valuable opportunities for further contributions, particularly through innovative approaches that integrate RAG, GenAI, and DevOps into SEE.

4.1 Literature Classification

As explained in Section 3, the selected studies were initially classified by study type into six categories: interviews, case studies, experimental or empirical studies, reviews, theoretical or conceptual works, and education-focused studies (Table 4). Subsequently, the studies were also classified by topic, resulting in 16 specific subcategories that support thematic organization. This classification encompasses various applications of artificial intelligence (AI) and RAG systems in SE, education, and DevOps (Table 6).

Fig. 2 presents the thematic categorization. The highest proportion corresponds to RAG applied to SE (RAG-SE, 25.4%), followed by RAG use in general education (RAG-EDU, 19.7%). These trends help identify dominant research areas and point to others that have been less attended to in the literature.

This classification helps identify predominant thematic areas and those that require further attention for future research. In this section, the research questions established at the beginning of this review are addressed. Each question is examined in detail, starting with an analysis of the impact of integrating RAG systems into academic environments. This includes key aspects such as challenges, opportunities, advantages, disadvantages, risks, and future directions.

The study then explores how RAG systems contribute to mitigating hallucinations generated by large language models (LLMs) in educational contexts, identifying guidelines to improve and regulate automated content generation. Subsequently, the synergy between humans and artificial intelligence in software development is analyzed, with a focus on strategies that enhance productivity and improve the quality of outcomes.

Finally, the incorporation of RAG systems into the DevOps lifecycle is examined, offering a comprehensive view of their applicability in both industrial and academic environments. Collectively, these findings provide a broad understanding of the impact and transformative potential of RAG systems across various fields of knowledge and professional practice.

To answer the question RQ1: *What is the impact of using and integrating RAG systems within academic environments in general, and specifically in software engineering?* we conducted a comprehensive analysis of the impact of RAG systems in academic environments, examining six key dimensions: advantages, opportunities, disadvantages, challenges, risks, and the experiences of both students and educators. This multidimensional approach allows for the identification of benefits as well as the barriers that influence their adoption and effectiveness across educational contexts.

RAG systems vary depending on the data sources integrated, the AI models employed, and their intended objectives. Nonetheless, reviewing past experiences in educational settings provides valuable insights into their capabilities, limitations, and real-world impact. Drawing from this evidence, the study offers a reference framework to better understand the role of RAG systems in education, their evolution, and their potential to support both academic and professional development. The results related to RQ1 are organized according to the six dimensions outlined above.

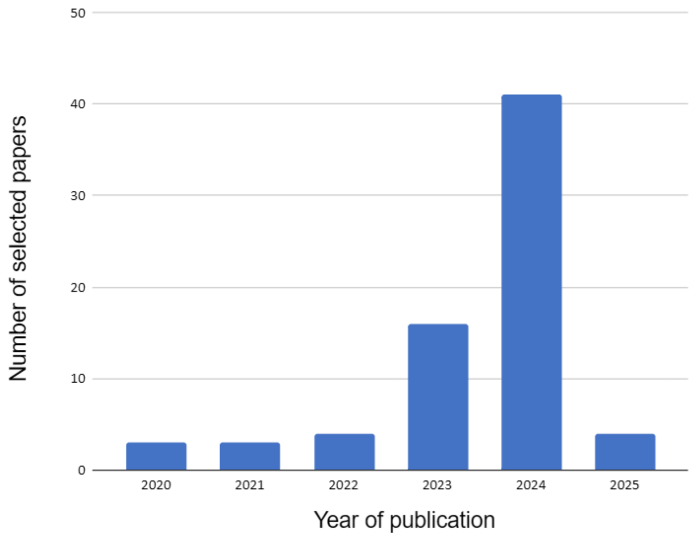

Fig. 1. Annual distribution of the 71 primary studies analyzed.

Table 6. Classification of ai-related studies by category, subcategory, and application context

Main Category	Category and Identifier	Description	Total studies
GAI (Generative AI)	GAI-SE	Applications in software engineering	3
	GAI-SEE	Educational uses within software engineering	3
	GAI-EDU	General education	3
	GAI-DevOps	Integration into DevOps contexts	1
RAG	RAG-SE	Software engineering	2
	RAG-SEE	Educational uses in software engineering	18
	RAG- EDU	General education	14
	RAG-DevOps	Integration with DevOps	0
AI (General AI)	AI-SE	Software lifecycle	8
	AI-SEE	Educational impact in software engineering	5
	AI-EDU	General educational applications	2
	AI-DevOps	AI use in DevOps	2
DevOps	DevOps-SE	Software engineering	0
	DevOps-SEE	Education in software engineering	2
	DevOps-EDU	General education	6
H-AI	H-IA	Human–AI interaction-focused studies	2

Furthermore, various tools and techniques have been explored for the implementation of RAG systems, reflecting a diverse set of methodological approaches. Examples include the integration of AI agents with RAG, the use of RAG pipelines in combination with large language models (LLMs) for academic queries, and the deployment of platforms such as VERTEX AI.

Additional advancements include KG-RAG, which incorporates knowledge graphs; translation-enhanced embeddings for multilingual applications; and RAG pipelines developed with LANGCHAIN. Other notable innovations include Graph-enhanced RAG, RAG-Reward systems based on reinforcement learning, and Dartboard for optimized information retrieval. Systems like QUIM-RAG enhance retrieval quality, while RAG FOUNDRY offers an end-to-end framework for data generation, model training, and performance evaluation.

Use of RAG Systems in Various Academic and Educational Contexts. RAG has emerged as a powerful tool in education, particularly in course design and the enhancement of teaching methodologies. Studies show that RAG significantly improves the accuracy and relevance of educational content compared to traditional methods (Shnaider et al., 2024). For example, its application in structuring university courses has demonstrated how advanced prompting techniques can enhance learning outcomes by improving the precision of information retrieval.

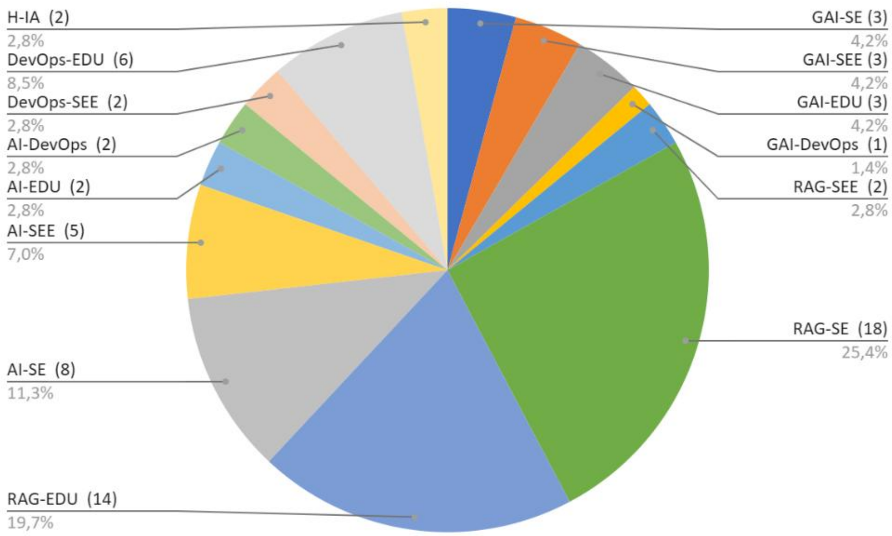

Fig. 2. Distribution of AI-related research studies by subcategory.

In virtual teaching environments, RAG-enhanced models have been effectively adopted to assist both students and educators. These systems provide reliable, interactive support by delivering responses grounded in textbook content, while also encouraging active learning. Students can use AI tutors to compare concepts, organize materials, and generate personalized study guides (Németh et al., 2023). Data from these implementations indicate that approximately 15% of student interactions reflect critical thinking, while another 15% demonstrate active learning through inquiry and comparative analysis (Németh et al., 2023). A synthesis of key aspects is provided in Tables 7 to 11, including advantages, opportunities, disadvantages, challenges, and risks associated with the implementation of RAG systems in educational contexts. Redundant or overlapping ideas have been consolidated, based on the analysis of 14 studies classified under the RAG-EDU category.

Use of RAG Systems in Educational Software Engineering and Computing. Tables 12 to 16 offer a focused overview of the advantages, opportunities, disadvantages, risks, and challenges observed in the application of RAG systems within software engineering, spanning both educational and technical domains. This examination is grounded in 20 studies categorized under RAG-SEE and RAG-SE.

Perspectives of Students, Teachers, and Evaluators on RAG Tools. Table 17 presents a range of perspectives from students, teachers, and evaluators regarding the use of RAG systems. While students highlight the benefits of immediate access to information and personalized support, educators raise concerns about academic integrity and the depth of learning. Evaluators, in turn, offer critical insights into the implementation and pedagogical effectiveness of these tools.

To answer the question *RQ2: How is the mitigation of hallucination generated by LLMs reported with the use of RAG in education?* we analyzed how educational studies address the reliability of language models when augmented with retrieval mechanisms. Hallucinations refer to responses that seem plausible but are actually incorrect or unsupported by verifiable evidence (Morić et al., 2024). Their mitigation in educational settings has been demonstrated through the use of RAG systems. For example, (Jeong, 2024) reports that fine-tuning GPT-4 with domain-specific data improves accuracy from 75% to 81%, and when combined with RAG, accuracy can increase further to 86%. This suggests that RAG systems are less prone to hallucinations than traditional GenAI systems, as they provide source-grounded responses that enhance trust and reliability (Darshan et al., 2024; Dakshit, 2024; Dong et al., 2023; Cooper & Klymkowsky, 2024; Fleischer et al., 2024).

Additionally, (Morić et al., 2024) shows that hallucinations occur in 12.72% of responses generated using the RAG method. Only 2 out of 74 queries showed hallucinations in relevant contexts, compared to 33% in responses with irrelevant context and approximately 30.3% in enumerative answers. These findings emphasize the importance of incorporating structured context and relevant knowledge to mitigate hallucinations, since performance declines significantly without them (Kuratomi et al., 2025).

Table 7. Advantages of using rag systems in various academic and educational contexts

Aspect	Details	Studies
Improved response quality	Increases accuracy and reduces hallucinations; enhances context awareness; maintains factual consistency in complex responses.	(Dieu et al., 2024; Hemmat et al., 2024; Morić et al., 2024; Thüs et al., 2024; Saha & Saha, 2024; Iscan et al., 2024; Alshammary et al., 2024; Cooper & Klymkowsky, 2024; Jiao et al., 2024; Guettala et al., 2024; Kuratomi et al., 2025; Jeong, 2024).
Efficiency and performance optimization	Reduces latency; automates response evaluation; enables rapid creation of tutoring systems using course materials.	(Dieu et al., 2024; Darshan et al., 2024; Dong et al., 2023; Cooper & Klymkowsky, 2024).
Scalability and knowledge management	Improves retrieval from diverse sources; adapts to dynamic knowledge bases without retraining; uses authoritative external sources.	(Morić et al., 2024; Dong et al., 2023; Alshammary et al., 2024; Guettala et al., 2024; Kuratomi et al., 2025)
Academic and teaching applications	Answers academic questions; acts as teaching assistant; enhances educational experiences; improves translation with human-like strategies.	(Dakshit, 2024; Iscan et al., 2024).
Personalization and contextual alignment	Integrates domain-specific knowledge; aligns responses with institutional policies; enables semantic search aligned with user intent.	(Dieu et al., 2024; Hemmat et al., 2024; Morić et al., 2024; Alshammary et al., 2024; Jiao et al., 2024; Kuratomi et al., 2025).
Proven systems and use cases	Systems like OwlMentor and DUETRAG show improvements; outperforms traditional LLMs; multi-agent systems show higher performance and satisfaction.	(Thüs et al., 2024; Dong et al., 2023; Saha & Saha, 2024; Iscan et al., 2024; Cooper & Klymkowsky, 2024; Jiao et al., 2024; Guettala et al., 2024).

Table 8. Opportunities for using rag systems in various academic and educational contexts

Aspect	Details	Studies
Personalized student support	Provides tailored academic assistance to university students; integrates with academic databases; enables RAG-powered chatbots for academic/cultural support; supports equitable access to AI tools.	(Dieu et al., 2024; Hemmat et al., 2024; Morić et al., 2024; Thüs et al., 2024; Saha & Saha, 2024; Guettala et al., 2024).
Quality of interaction and assessment	Delivers responses with high fidelity and clarity; allows fast, context-aware assessments; helps students build arguments; improves knowledge application through chatbots.	(Hemmat et al., 2024; Darshan et al., 2024; Morić et al., 2024; Thüs et al., 2024; Cooper & Klymkowsky, 2024; Guettala et al., 2024).
Scalability and system design	Scales in specialized educational settings; combines generation and retrieval; supports continuous learning; promotes innovation through open chatbot repositories.	(Darshan et al., 2024; Morić et al., 2024; Cooper & Klymkowsky, 2024; Guettala et al., 2024).
Instructional enhancement and teaching support	Enhances instruction via AI-assisted tutoring; functions as virtual TA and content platform; supports assignment design and evaluation; faculty feedback highlights.	(Dakshit, 2024; Cooper & Klymkowsky, 2024).
Technology integration and advancement	Incorporates deep learning; uses advanced NLP; applies RAG to multilingual contexts; enhances LLMs for tasks without retraining.	(Dakshit, 2024; Saha & Saha, 2024; Iscan et al., 2024).
Evaluation and improvement	Performance backed by qualitative acceptability; student feedback refines platforms; cross-disciplinary evaluation; future models aim for stronger retrieval.	(Hemmat et al., 2024; Thüs et al., 2024).

In information retrieval settings, Hybrid RAG (An approach that combines multiple retrieval techniques to enhance language model responses) has shown notable improvements. It achieved over 40% gain on the q2a-100 dataset, more than 60% on cmds-100, and above 70% in Recall on abbr-100 (Shi et al., 2024). Verification mechanisms and response inspection have also helped reduce hallucinations, especially when meaningful context is available (Fleischer et al., 2024; Morić et al., 2024).

Lastly, connecting AI models with external knowledge, as implemented in ChatGPT using RAG, aligns AI tutor responses and reduces hallucinated answers to approximately 19.5% (Ma et al., 2024). However, hallucinations may still persist to a lesser extent (Morić et al., 2024; Thüs et al., 2024; Guettala et al., 2024; Kuratomi et al., 2025).

To answer the question *RQ3: How can the synergy between humans and AI be improved in software development to maximize productivity?* we explored strategies that promote collaborative interaction between developers and AI systems. This synergy can be enhanced by fostering complementary roles, clearly defined responsibilities, and continuous feedback mechanisms. AI tools such as OwlMentor, KG-RAG, and educational chatbots support the understanding of UML diagrams and scientific texts (Thüs et al., 2024; Dong et al., 2023; Saha & Saha, 2024; Ardimento et al., 2024). GenAI also contributes to engagement, creativity, and rapid feedback in learning environments (Yabaku et al., 2024; Kumar et al., 2024; Kumar et al., 2023; Kim et al., 2023; Kästner & Kang, 2020). These tools help students prepare for professional practice through innovative methods (Yabaku et al., 2024), while enabling individualized learning experiences and addressing student diversity (Vetriselvi et al., 2024; Johnson, 2024). However, human supervision combined with well-defined guidelines remains essential to ensure the quality and integrity of learning processes (Yabaku et al., 2024; Bull & Kharrufa, 2024; Ciolacu et al., 2024; Sauvola et al., 2024; Daniel, 2023).

Table 9. Disadvantages of using rag systems in various academic and educational contexts

Aspect	Details	Studies
Contextual and retrieval limitations	May retrieve irrelevant or biased content; failures in retrieval affect response quality; static or rule-based libraries reduce adaptability; limited knowledge bases constrain effectiveness.	(Darshan et al., 2024; Dakshit, 2024; Guettala et al., 2024).
Information accuracy and reliability	Hallucinations may still occur; contextual learning problems can lead to misinformation; ambiguous language or terms may cause misinterpretations.	(Morić et al., 2024; Dakshit, 2024; Thüs et al., 2024; Alshammary et al., 2024; Guettala et al., 2024; Kuratomi et al., 2025).
Ethical and pedagogical concerns	Raises plagiarism and policy issues; may hinder critical thinking and lead students to depend on AI-generated answers rather than engaging in their own reasoning; direct answers can disrupt authentic learning; feedback may be ignored by students.	(Dakshit, 2024; Thüs et al., 2024; Cooper & Klymkowsky, 2024; Barnett et al., 2024).
Complexity of implementation and maintenance	Technical complexity in integration; challenge in balancing efficiency and accuracy; needs expert supervision; dependence on external sources may cause inconsistency.	(Darshan et al., 2024; Morić et al., 2024; Dakshit, 2024).
Explainability and comprehension barriers	Fragmented or incoherent responses due to limited contextual grasp; low transparency complicates validation; cultural or linguistic nuances may reduce clarity.	(Dong et al., 2023; Saha & Saha, 2024).
Scalability and responsiveness issues	System performance may not scale well across settings; some queries may lack sufficient data for quality responses.	(Dakshit, 2024; Barnett et al., 2024).

Table 10. Challenges of using rag systems in various academic and educational contexts

Aspect	Details	Studies
Technical limitations and performance constraints	High computational requirements; dependence on content quality; risks of catastrophic forgetting; weak performance without context; limited multimodal support; static data inefficiency; integration challenges with external sources.	(Dieu et al., 2024; Hemmat et al., 2024; Darshan et al., 2024; Morić et al., 2024; Dakshit, 2024; Alshammary et al., 2024; Jiao et al., 2024; Guettala et al., 2024; Kuratomi et al., 2025).
Accuracy and reliability issues	Hallucinations in complex responses; overfitting and bias; context misinterpretation; fragmented or incomplete outputs; difficulty with basic math and reasoning explanation.	(Darshan et al., 2024; Morić et al., 2024; Dakshit, 2024; Dong et al., 2023; Alshammary et al., 2024; Cooper & Klymkowsky, 2024; Jiao et al., 2024; Kuratomi et al., 2025).
Integration and usability challenges	Integration into academic systems is complex; relies heavily on text-based formats; lacks adaptive feedback; requires expert supervision; struggles with nuanced tasks.	(Dieu et al., 2024; Hemmat et al., 2024; Morić et al., 2024; Dakshit, 2024; Cooper & Klymkowsky, 2024; Jiao et al., 2024).
User perception and adoption barriers	Mismatch between student and expert evaluations; technical quality doesn't ensure engagement; personal perceptions affect adoption.	(Thüs et al., 2024).

To improve productivity, RAG and AI systems help reduce grading time, exam costs and unnecessary resource consumption while maintaining objectivity in evaluation (Darshan et al., 2024; Johnson, 2024; Kästner & Kang, 2020). These systems automate a portion of software development activities, with reported support of up to 50 percent of repetitive tasks (Vetriselvi et al., 2024; Kumar et al., 2024; Kumar et al., 2023). Additionally, they contribute to daily operational efficiency and accelerate development workflows (Vetriselvi et al., 2024; Zhang, 2023; Iyer et al., 2024). Users have reported increased levels of satisfaction, collaboration and performance (Coutinho et al., 2024).

Table 11. Risks of using rag systems in various academic and educational contexts

Aspect	Details	Studies
Knowledge and retrieval limitations	Struggles with novel or ambiguous queries due to outdated data; fails to access localized or niche content; weak multilingual and domain-specific language support; ineffective retrieval affects quality; static libraries limit adaptability.	(Dieu et al., 2024; Hemmat et al., 2024; Morić et al., 2024; Iscan et al., 2024; Alshammary et al., 2024; Guettala et al., 2024; Kuratomi et al., 2025).
System integration and scalability risks	Complex retrieval-generation integration; challenges in real-time scalability due to high demands; difficulty processing structured content; limited by text-only inputs.	(Hemmat et al., 2024; Darshan et al., 2024; Morić et al., 2024; Dakshit, 2024; Guettala et al., 2024).
Educational and pedagogical concerns	Excessive dependence on AI may reduce opportunities for critical thinking; automation may marginalize educators; current systems lack adaptability to diverse learning needs; they also struggle with supporting scientific comprehension; may encourage solution-focused behavior rather than deeper understanding. surface-level understanding.	(Darshan et al., 2024; Thüs et al., 2024; Dong et al., 2023; Cooper & Klymkowsky, 2024).
Ethical, equity, and cultural issues	Training data bias affects fairness; concerns around plagiarism and policy; difficulty adapting culturally; limited support for multilingual and diverse learners.	(Darshan et al., 2024; Morić et al., 2024; Dakshit, 2024; Saha & Saha, 2024; Iscan et al., 2024).
Explainability and validation barriers	Lack of transparency in outputs causes unreliable responses; hard to automate domain validation; KG-RAG systems still need expert oversight.	(Dong et al., 2023).
Research gaps and systemic implementation risks	Few studies on student interaction with RAG; lack of systemic educational strategies; ongoing expert evaluation and database expansion are needed.	(Cooper & Klymkowsky, 2024; Kuratomi et al., 2025).

Table 12. Advantages of using rag systems in educational software engineering and computing

Aspect	Details	Studies
Improved learning and pedagogical support	Enhances feedback accuracy (e.g., UML); fosters safe learning; supports understanding-based learning; achieves high response accuracy in tutoring.	(Ardimento et al., 2024; Ma et al., 2024).
Accuracy, credibility, and hallucination reduction	Uses external knowledge for accuracy; mitigates hallucinations; custom datasets reduce misinformation; refined retrieval improves accuracy.	(Ardimento et al., 2024; Simon et al., 2024; Zhang et al., 2025; Saha et al., 2024; National Technical University of Ukraine & O, 2023; Ahmed et al., 2024; Bernardi et al., 2024; Shi et al., 2024; Vetriselvi et al., 2024; Wang et al., 2024; Jeong, 2024; He et al., 2025; Barochiya et al., 2024).
Enhanced LLM and system capabilities	Combines retrieval and generation for personalized responses; boosts LLMs in knowledge-intensive and coding tasks; mixes retrieval types for effectiveness.	(Ardimento et al., 2024; Simon et al., 2024; Zhang et al., 2025; Saha et al., 2024; National Technical University of Ukraine & O, 2023; Fleischer et al., 2024; Chaubey et al., 2024; Ahmed et al., 2024; Shi et al., 2024; Vetriselvi et al., 2024; Wang et al., 2024; Jeong, 2024; He et al., 2025; Barochiya et al., 2024; Li et al., 2022)
Software engineering and testing applications	Improves unit test generation and coverage; API-level RAG enhances code testing; identifies untested lines.	(Shin et al., 2024).
Development efficiency and prototyping	Speeds up development by indexing unstructured data; avoids heavy annotation; enables fast prototyping and efficient workflows.	(Barnett et al., 2024; Fleischer et al., 2024).
Advanced retrieval and evaluation strategies	QUIM-RAG and hybrids show state-of-the-art performance; API-level and structured evaluation improve results; balances efficiency and effectiveness.	(Simon et al., 2024; Saha et al., 2024; Chaubey et al., 2024; Shi et al., 2024; Wang et al., 2024; He et al., 2025; Li et al., 2022).

The responsible use of LLMs and RAG systems is essential to address ethical and societal risks (Fleischer et al., 2024; Vetriselvi et al., 2024; Virvou & Tsihrintzis, 2023; Bommasani et al., 2022). Base models may introduce disparities that disproportionately affect marginalized users (Bommasani et al., 2022). Therefore, ethical design should consider aspects such as privacy, fairness, transparency and Human-Centered Artificial Intelligence principles (Saha & Saha, 2024; Vetriselvi et al., 2024; Kästner & Kang, 2020). AI should be positioned to assist human reasoning without displacing human judgment (Ciolacu et al., 2024; Sauvola et al., 2024).

Table 13. Opportunities for using rag systems in educational software engineering and computing

Aspect	Details	Studies
Enhancing learning and assessment	Improves feedback and understanding in UML/software tasks; enables execution monitoring; AI tutors assist with homework; supports fair evaluation.	(Ardimento et al., 2024; Ma et al., 2024; Simon et al., 2024).
Advancing system capabilities	Improves semantic search and source-linked responses; enhances LLMs in specific domains; enables real-time Q&A and chatbot use.	(Barnett et al., 2024; Simon et al., 2024; National Technical University of Ukraine & O, 2023; Chaubey et al., 2024; Ahmed et al., 2024; Barochiya et al., 2024).
Supporting research and innovation	Provides structured evaluation for RAG; supports rapid prototyping (e.g., RAG FOUNDRY); expands research with fine-tuning and new variants.	(Simon et al., 2024; Fleischer et al., 2024; Chaubey et al., 2024; Ahmed et al., 2024; Shi et al., 2024).
Integrating with broader AI ecosystems	Combines GenAI and RAG for SOPs, synthesis, and info management; supports data-driven decisions; frameworks like LANGGRAPH improve GenAI.	(Chaubey et al., 2024; Bernardi et al., 2024; Vetriselvi et al., 2024; Jeong, 2024).
Ethics, fairness, and scalability	Promotes ethical AI in education; encourages scalable low-code and LLM solutions.	(National Technical University of Ukraine & O, 2023; Chaubey et al., 2024).
Expanding use cases in natural language processing and beyond	Supports dialogue generation, translation; drives interdisciplinary academic research.	(Barochiya et al., 2024; Li et al., 2022).

Table 14. Disadvantages of using rag systems in educational software engineering and computing

Aspect	Details	Studies
Data dependency and input limitations	Requires large, high-quality datasets; sensitive to input variations; domain-specific data needed to ensure generalization and accuracy; limited test cases reduce reliability.	(Ardimento et al., 2024; Shin et al., 2024; Zhang et al., 2025; Fleischer et al., 2024).
Accuracy, hallucinations, and bias	May hallucinate or misinterpret context; reinforces training data biases; accuracy may degrade due to compression/retrieval trade-offs.	(Ardimento et al., 2024; Barnett et al., 2024; Saha et al., 2024; Vetriselvi et al., 2024; Jeong, 2024; He et al., 2025).
Performance and evaluation challenges	Complexity increases overfitting risk; lacks traceable evaluation; runtime-only analysis; simplistic evaluation metrics; retrieval-generation coupling complicates measurement.	(Barnett et al., 2024; Simon et al., 2024; Fleischer et al., 2024; Chaubey et al., 2024; Pickett et al., 2025; Wang et al., 2024; Jeong, 2024; Li et al., 2022).
Computational and resource constraints	High cost and execution time; expensive token usage; quadratic scaling in some execution models; unsuitable for low-resource settings.	(Shin et al., 2024; Chaubey et al., 2024; Pickett et al., 2025; Wang et al., 2024; Li et al., 2022).
Functional limitations in software engineering tasks	No improvement in test syntactic/dynamic correctness; weak with semantic gaps; higher perplexity; latency in real-time use; misses domain-specific terms.	(Shin et al., 2024; Chaubey et al., 2024; Vetriselvi et al., 2024; Wang et al., 2024; He et al., 2025).
System rigidity and generalization issues	Struggles to generalize across tasks; cannot integrate real-time data after deployment; depending on a single retrieved passage limits system flexibility.	(Ardimento et al., 2024; Fleischer et al., 2024; Jeong, 2024; Li et al., 2022).

Human roles continue to be relevant. Professionals such as security analysts and educators contribute in areas such as incident handling, documentation and architectural decisions (Bernardi et al., 2024; Sauvola et al., 2024; Daniel, 2023). Training users helps avoid misapplication and decreases error rates (Kim et al., 2023). Human evaluation of AI-generated content ensures clarity, precision and contextual relevance (Barochiya et al., 2024).

Challenges in human-AI collaboration still exist. AI tutors may offer slow or inconsistent responses that impact perceived reliability (Hemmat et al., 2024; Morić et al., 2024; Ma et al., 2024; Zhang et al., 2025; Fleischer et al., 2024). To mitigate these issues, continuous monitoring and updated datasets are needed (Zhang et al., 2025; Bernardi et al., 2024). Furthermore, system alignment with user expectations and contextual needs is critical (Thüs et al., 2024; Zhang et al., 2025; Zhang, 2023).

The broader impact of AI is reshaping professional roles. It influences how software engineers work and how responsibilities are distributed across teams (Vetriselvi et al., 2024; Sauvola et al., 2024; Johnson, 2024; Chu & Lim, 2023; Vierhauser et al., 2024; Alenezi et al., 2022; Daniel, 2023). Integrating AI trends into curricula contributes to preparing future professionals (Vierhauser et al., 2024; Nguyen et al., 2024; Zúñiga Sánchez, 2024). In addition, AI systems increasingly affect other domains such as healthcare, agriculture, defense and public policy (Gong, 2021; Virvou & Tsihrintzis, 2023).

Table 15. Risks of using rag systems in educational software engineering and computing

Aspect	Details	Studies
Reliability and validity of research findings	Fast LLM evolution risks outdated findings; AI tutor studies lack causal proof of performance gains; RAG validation success may not be unique or reproducible.	(Ardimento et al., 2024; Ma et al., 2024; Simon et al., 2024).
Content gaps and misinformation	Missing answers can lead to poor outputs; hallucinations persist without verification; static knowledge bases reduce freshness.	(Barnett et al., 2024; Fleischer et al., 2024; Chaubey et al., 2024; Ahmed et al., 2024; Bernardi et al., 2024; Vetriselvi et al., 2024; Wang et al., 2024; He et al., 2025; Li et al., 2022).
Bias and subjectivity in evaluation	Manual evaluation introduces bias; reward models neglect reasoning/security; misalignment with traditional assessments impacts evaluation quality.	(Simon et al., 2024; Shin et al., 2024; Zhang et al., 2025).
Educational impact and ethical risks	AI tutors may short-circuit learning; GenAI raises integrity and cognitive concerns; lack of human-centered design leads to ethical/regulatory issues.	(Ma et al., 2024; Vetriselvi et al., 2024).
Interpretability and trust	Generated content may be inaccurate; privacy and transparency concerns arise; depending solely on one retrieved passage may reduce the depth of responses; clear governance is needed.	(Fleischer et al., 2024; Vetriselvi et al., 2024; Li et al., 2022).

Table 16. Challenges and issues in using rag systems in educational software engineering and computing

Aspect	Details	Studies
Implementation complexity and technical barriers	Requires deep understanding of data and design; constant updates needed due to model evolution; libraries often need customization; real-time data needs extra validation.	(Ardimento et al., 2024; Simon et al., 2024; Shin et al., 2024; Fleischer et al., 2024; Ahmed et al., 2024; Pickett et al., 2025; Bernardi et al., 2024; Shi et al., 2024; Vetriselvi et al., 2024; Jeong, 2024; Barochiya et al., 2024; Li et al., 2022).
Data limitations and information gaps	Hallucinations persist when documents are missing; document exclusion reduces accuracy; engineers struggle to find info in large organizations; lack of documentation causes failures.	(Barnett et al., 2024; Simon et al., 2024; Shin et al., 2024; Saha et al., 2024; Shi et al., 2024; Vetriselvi et al., 2024; Wang et al., 2024; Jeong, 2024; He et al., 2025; Barochiya et al., 2024; Li et al., 2022).
Evaluation, metrics, and reward model challenges	Metrics like line coverage may misrepresent effectiveness; reward models underperform in RAG; existing scores may not match educational utility; preprocessing may distort evaluations.	(Shin et al., 2024; Zhang et al., 2025; Saha et al., 2024; Fleischer et al., 2024; Ahmed et al., 2024; Vetriselvi et al., 2024; He et al., 2025; Barochiya et al., 2024).
User experience and educational impact	Students dissatisfied with slow, shallow AI responses; GenAI may harm academic integrity and critical thinking.	(Ma et al., 2024; National Technical University of Ukraine & O, 2023; Vetriselvi et al., 2024).
Bias and inconsistency	Training bias and manual labeling skew results; non-determinism complicates repeatability; output varies across domains.	(Ardimento et al., 2024; Simon et al., 2024; Shin et al., 2024; Zhang et al., 2025; Li et al., 2022).
Knowledge retention and productivity gaps	LLMs struggle with information retention; RAG productivity gains limited with poor-quality sources.	(Shi et al., 2024; Wang et al., 2024; Li et al., 2022).

To answer the question *RQ4: In which phases of the software lifecycle has the use of RAG systems been documented, and how could they be integrated into DevOps stages?* we reviewed studies classified under the RAG-SE and RAG-SEE categories. These studies examine the application of RAG systems across specific phases of the software development lifecycle. The analysis was structured according to the DevOps workflow, including the stages Plan, Code, Build, Test, Release, Deploy, Operate, and Monitor, in order to explore the potential integration of RAG systems at each phase.

Table 17. Perspectives of students, teachers, and evaluators on rag tools

Perspective	Details	Studies
Student experiences and perceptions	Students found tools like OwlMentor helpful; appreciated AI tutors for clear guidance; liked non-intrusive monitoring and personalized learning; noted issues like slow responses and reduced usefulness over time; equitable access is important.	(Hemmat et al., 2024; Thüs et al., 2024; Dong et al., 2023; Cooper & Klymkowsky, 2024; Ardimento et al., 2024; Ma et al., 2024; Vetriselvi et al., 2024; Fernandes et al., 2022).
Faculty and teaching perspectives	Faculty highlighted workload reduction and consistent assessment; saw RAG as effective teaching aids and safe environments; useful for supporting large student cohorts.	(Darshan et al., 2024; Dakshit, 2024; Dong et al., 2023; Cooper & Klymkowsky, 2024; Ardimento et al., 2024; Ma et al., 2024).
Evaluator and assessment insights	Evaluators valued clarity and coherence; AI grading seen as more objective; custom materials integration improves relevance.	(Hemmat et al., 2024; Darshan et al., 2024; Vetriselvi et al., 2024).
System feedback and adaptation	User feedback improves response quality; feedback loops drive tool adaptation; GenAI adoption highlights future-focused integration; current systems need continued refinement.	(Morić et al., 2024; Cooper & Klymkowsky, 2024; Ardimento et al., 2024; Fernandes et al., 2022).

The reviewed literature shows that RAG systems have primarily been integrated into selected lifecycle phases, indicating their adaptability to different development contexts. Regarding the Plan phase, study (Ardimento et al., 2024) introduces a novel approach to teaching UML during the planning and design stage. In the Code phase, study (National Technical University of Ukraine & O, 2023) presents a RAG system to support developers in low-code environments, (Ahmed et al., 2024) explores the integration of large language models (LLMs) within RAG contexts, and (He et al., 2025) highlights the benefits of external knowledge bases for improving coding tasks. In the Test phase, study (Shin et al., 2024) evaluates the effectiveness of RAG in generating unit tests. For the Operate phase, study (Barochiya et al., 2024) identifies challenges in adopting multimodal large language models (MM-LLMs), particularly due to outdated internal knowledge and hallucinations. Finally, in the Monitor phase, study (Bernardi et al., 2024) introduces a RAG-based framework for security reporting and monitoring.

5 Discussion

The discussion and results section present the interpretations and analysis of each question. Each sub-theme presented below reflects the interpretation of the findings obtained.

5.1 Impact of RAG Systems on Education and Educational Software Engineering

This sub-theme addresses research question RQ1. The analysis draws on 34 primary studies, which include interviews, experimental studies, case analyses, and theoretical reviews. These studies explore the advantages, opportunities, disadvantages, challenges, and risks associated with the use of RAG systems in education, SE, and computing. While diverse methodologies were employed to cover a wide range of scenarios, the possibility of bias in the findings is acknowledged. As shown in Fig. 3, most studies fall under the RAG-SEE and RAG-EDU categories, which reflects the emphasis on applied educational contexts. This distribution supports the focus of the present discussion on teaching practices, assessment tools, and feedback mechanisms.

The results indicate that RAG systems improve both the accuracy and relevance of information. They contribute to optimizing teaching practices, personalizing learning, and reducing teacher workload by automating assessments and feedback processes (Darshan et al., 2024; Dakshit, 2024; Thüs et al., 2024; Dong et al., 2023; Saha & Saha, 2024; Alshammary et al., 2024; Cooper & Klymkowsky, 2024; Guettala et al., 2024; Kuratomi et al., 2025; Ardimento et al., 2024; Ma et al., 2024). Furthermore, their integration with virtual assistants reinforces academic tutoring.

In higher education and lifelong learning, RAG systems support adaptive learning models that benefit both students and instructors. In fields such as SE and computer science, these systems have been used to optimize unit test generation, code evaluation, and technical information retrieval. They also enhance personalized tutoring and enable interactive simulations (Dakshit, 2024; Thüs et al., 2024; Dong et al., 2023; Cooper & Klymkowsky, 2024; Ardimento et al., 2024; Ma et al., 2024; Shin et al., 2024; National Technical University of Ukraine & O, 2023; Fleischer et al., 2024; Bernardi et al., 2024; He et al., 2025).

However, their implementation also introduces challenges and risks. The quality and accuracy of RAG outputs may be compromised by biases in training data or retrieval errors. There is a concern that excessive reliance on these systems may diminish students' critical thinking abilities. Other notable challenges include high computational demands, which may hinder adoption in institutions with limited technological infrastructure. Additional risks involve academic integrity, potential plagiarism, and the limited explainability of some generative outputs (Darshan et al., 2024; Morić et al., 2024; Alshammary et al., 2024; Cooper & Klymkowsky, 2024; Guettala et al., 2024; Ma et al., 2024; Shin et al., 2024; Saha et al., 2024; Vetriselvi et al., 2024).

To ensure effective implementation, the literature recommends developing hybrid strategies that combine automated generation with human validation. It is also important to provide training for educators and students in the appropriate use of these tools and to establish protocols to evaluate their impact on learning outcomes. Finally, ongoing research should focus on reducing bias, minimizing hallucinations, and improving overall accuracy (Hemmat et al., 2024; Thüs et al., 2024; Dong et al., 2023; Alshammary et al., 2024; Jiao et al., 2024; Ardimento et al., 2024; Simon et al., 2024; Zhang et al., 2025; Fleischer et al., 2024; Wang et al., 2024; Jeong, 2024).

5.2 Mitigation of LLM-Generated Hallucination with the Use of RAG in Teaching

In relation to RQ2, the studies indicate that integrating RAG systems into teaching significantly reduces hallucinations in large language models (LLMs). Refining information in GPT-4 improves accuracy from 75% to 81%, and when combined with RAG, it reaches 86% (Jeong, 2024). By incorporating verifiable sources and relevant context, RAG reduces hallucination rates to 12.72% in context-appropriate responses and to 19.5% when retrieving information from external sources such as ChatGPT (Ma et al., 2024).

Although some errors persist (Morić et al., 2024; Thüs et al., 2024; Guettala et al., 2024; Kuratomi et al., 2025), the literature confirms that RAG can effectively mitigate hallucinations and biases, provided that the underlying database and configuration are properly implemented (Saha & Saha, 2024). Despite methodological differences among studies, the findings consistently support the effectiveness of RAG in generating more accurate and trustworthy responses (Darshan et al., 2024; Dakshit, 2024; Dong et al., 2023; Cooper & Klymkowsky, 2024; Fleischer et al., 2024).

5.3 Synergy Between Humans and AI in Software Development to Maximize Productivity

The literature reviewed for RQ3 shows that combining RAG systems with GenAI in software development can substantially enhance productivity. Benefits include high user acceptance and improved user experience (Hemmat et al., 2024; Thüs et al., 2024; Dong et al., 2023), reduced bias and time savings in evaluations and report generation (Darshan et al., 2024; Zhang et al., 2025), and more efficient resource allocation (Darshan et al., 2024).

Nevertheless, several challenges are identified. These include the risk of hallucinated or incorrect responses that can compromise information reliability (Morić et al., 2024; Fleischer et al., 2024), as well as the need for continuous human oversight to ensure process accuracy and security (Ciolacu et al., 2024; Sauvola et al., 2024). These findings align with the reported benefits in Table 12 and Table 13, particularly regarding the reduction of workload and the ability to support routine development tasks with AI assistance.

Maximizing productivity requires a balance between AI automation and human intervention. This involves establishing ethical guidelines (Dakshit, 2024; Saha & Saha, 2024; Chaubey et al., 2024; Vetriselvi et al., 2024; Kästner & Kang, 2020; Daniel, 2023; Bommasani et al., 2022), and implementing ongoing training strategies to mitigate risks and improve integration into development workflows (Kim et al., 2023).

5.4 The Synergy Between RAG and DevOps in SEE

The analysis for RQ4 reveals that RAG systems have been integrated into specific DevOps stages, including Plan, Code, Test, Operate, and Monitor. However, they have not yet been applied in the Build, Release, or Deploy stages. This gap presents a valuable opportunity to optimize the entire software development and delivery lifecycle.

Beyond software engineering, Artificial Intelligence has shown significant importance in the broader educational context by enabling personalized learning, automating administrative processes, and providing equitable access to resources (Grote & Bogner, 2023; Bhandari et al., 2023; Pan et al., 2023; Cico et al., 2023; Dong & Jia, 2020; Kästner & Kang, 2020; Heck & Schouten, 2021; Chu & Lim, 2023; Vierhauser et al., 2024; Gong, 2021; Slimi, 2023). The literature suggests that the synergy between RAG and DevOps in SE education can enhance current practices. As shown in Table 6, though, the RAG-DevOps category includes no studies, which indicates that this area remains largely unexplored. Some studies highlight that Artificial Intelligence for Software Engineering (AI4SE) has the potential to redefine development practices (Lo, 2023; Kumar et al., 2023). It also promotes collaboration among professionals, a core principle of DevOps that encourages integration between development and operations. In addition, AI can shorten development cycles, reduce costs, and increase testing productivity (Kumar et al., 2023). It also enhances creativity and supports DevOps objectives related to automation and performance.

Compared to general education, where RAG systems are mostly used for tutoring, personalized learning, or content generation, their role in SE Education is more technical. In SEE, RAG tools are applied to support practical tasks, such as those in DevOps workflows. However, the limited presence of studies in the RAG-DevOps category (Table 6 and Fig. 2) suggests that this area has received relatively little attention. This observation underscores the potential value of further exploring how RAG applications could be integrated into the DevOps pipeline. This includes helping students work with automation, continuous integration, and real-world development practices. These applications go beyond teaching concepts and aim to improve hands-on skills that align with industry needs.

Study (Bernardi et al., 2024) emphasizes that DevOps improves learning efficiency within the framework of Education 4.0. Meanwhile, studies (Mota et al., 2024) and (Iyer et al., 2024) consider its integration an innovative alternative to traditional methods. Lastly, the growing demand for DevOps professionals has led to adaptations in SE curricula, with greater emphasis on continuous delivery, implementation pipelines, and real-world industry practices (Ferino et al., 2023; Ferino et al., 2021; Fernandes et al., 2020; Rocha et al., 2023;). These insights directly address research question RQ4, which seeks to identify the phases of the software lifecycle where RAG systems have been applied and explore their possible integration into DevOps workflows.

5.5 Critical synthesis and knowledge gaps

This study presents a thematic classification of RAG applications in education and software engineering. However, several tensions and gaps in the literature can be identified. One tension lies in the contrast between the ability of RAG systems to offer personalized learning and feedback, and the risk that such automation may lead to reduced critical thinking when students rely too much on AI-generated responses (Dakshit, 2024; Thüs et al., 2024; Cooper & Klymkowsky, 2024).

Another tension appears between the reported benefits of AI tutors in teaching contexts and the practical challenges related to computational requirements and system integration. These barriers may limit adoption, particularly in institutions with restricted technological resources (Darshan et al., 2024; Morić et al., 2024; Shin et al., 2024; Chaubey et al., 2024; Wang et al., 2024).

In addition, the review shows that although there is increasing interest in combining RAG with DevOps, no primary studies have addressed this connection directly (RAG-DevOps = 0, Table 6). This absence highlights a gap that may be relevant for future research on both technical and educational fronts.

Lastly, there is limited evidence on the long-term use of RAG systems in education. Most studies evaluate short-term outcomes or tool performance, but few examine their sustained impact on student learning or curriculum design. This suggests the need for research that includes longitudinal approaches and broader institutional analysis.

5.6 Contribution of the classification

The classification proposed in Table 6 offers a structured view of how RAG systems and related technologies have been applied across different domains, including software engineering, general education, DevOps, and human–AI interaction. This framework supports the identification of patterns in existing literature and reveals areas that have received limited attention. For instance, while the categories RAG-SEE and RAG-EDU show a relatively high number of studies, there is a complete absence of work categorized as RAG-DevOps.

By organizing the reviewed literature into sixteen subcategories grouped by application domain, the framework contributes to mapping the intersection between RAG and educational practices in software-related fields. Unlike traditional reviews that focus only on benefits or limitations, this structure highlights how research is distributed across contexts. This contribution may assist future studies in selecting research directions and identifying underrepresented areas where further exploration is needed, especially in aligning RAG systems with DevOps principles and curricular design.

6 Conclusions and Future Work

The integration of RAG systems into SEE has demonstrated a positive and multifaceted impact. This SLR offers a consolidated view of the state of the art, showing how these systems enhance information accuracy, optimize learning, personalize instruction, and automate tasks such as assessment and feedback. Additionally, the SLR highlights how RAG helps mitigate hallucinations and biases in language models, provided that high-quality data sources and well-designed configurations are used.

This review contributes by providing a detailed thematic classification of existing studies and by identifying areas that have been less attended to in the literature, such as the integration of RAG into specific stages of DevOps. In this way, the SLR not only systematizes existing knowledge but also outlines a research agenda for future studies, particularly in contexts where RAG can deliver both educational and technical value.

In practical terms, the findings are relevant to curriculum designers, educators, and researchers, as they offer empirical and conceptual evidence to guide the informed adoption of RAG-based technologies in higher education and SE. Despite its benefits, the review also identifies challenges, including the need for human oversight, technical limitations, and ethical risks, all of which must be addressed to ensure effective and responsible implementation.

Moreover, the integration of RAG systems in SEE presents a unique opportunity to bridge theoretical knowledge with real-world practices. These systems can be configured to deliver highly personalized learning experiences, adapting content and difficulty levels to individual student needs. RAG-based platforms also serve as intelligent tutoring systems, offering context-aware assistance, answering programming-related queries, and guiding students through complex problem-solving tasks. Additionally, RAG contributes to the mitigation of hallucinations and biases commonly found in large language models by grounding responses in verified sources, thereby increasing the reliability of educational content. Their ability to reinforce source-based reasoning also improves the trustworthiness of AI-generated content in academic environments.

In software development contexts, the combination of RAG with generative AI has been shown to boost productivity, streamline evaluations, and reduce the time and cognitive effort required for coding and documentation tasks. By supporting activities such as automated testing, code evaluation, and continuous integration within DevOps workflows, RAG not only enhances student engagement but also aligns academic training with current industry demands. For curriculum designers, this synergy opens possibilities to design hands-on learning experiences that reflect real-world practices. This synergy between human expertise and AI-driven tools encourages the development of both technical competencies and critical thinking skills. As educational institutions move toward Education 4.0 paradigms, incorporating RAG into teaching strategies will be essential to cultivating adaptive, future-ready software professionals. For educators, RAG systems can serve as complementary tools that help reduce repetitive tasks and enhance instructional delivery.

The results obtained indicate great potential in the integration of RAG systems and DevOps practices to transform SE education by providing personalized learning, access to up-to-date information and assistance in problem solving. As part of future work, it is proposed to conduct a comparative study on techniques and trends in the development and implementation of RAG systems in academia and industry. Such a study could offer insights into the adaptability and impact of these systems across educational and industrial domains. Also, given the rapid and dizzying evolution of the subject, it is proposed to continue updating the state of the art on the performance of these systems in education, with a specific focus on SE, and to continue reporting relevant findings in the literature. Additionally, future studies may examine long-term effects on student engagement, the evolving role of educators in hybrid AI–human environments, and ethical considerations in deploying RAG systems in classrooms.

Acknowledgement: This work was supported by the National Council for Humanities, Science, and Technology (SECIHTI) with scholarship No. CVU 1147172 and by the National Technological Institute of Mexico (TecNM) at the National Center for Research and Technological Development (CENIDET).

References

Ahmed, M., Dorrah, M., Ashraf, A., Adel, Y., Elatrozy, A., Mohamed, B. E., & Gomaa, W. (2024). CodeQA: Advanced programming question-answering using LLM agent and RAG. In *2024 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES)* (pp. 494–499). IEEE. https://doi.org/10.1109/NILES63360.2024.10753267.

Alenezi, M., Zarour, M., & Akour, M. (2022). *¿Can artificial intelligence transform DevOps?* (No. arXiv:2206.00225). arXiv. http://arxiv.org/abs/2206.00225.

Alshammary, M., Uddin, M. N., & Khan, L. (2024). RFPG: Question-answering from low-resource language (Arabic) texts using factually aware RAG. In *2024 IEEE 10th International Conference on Collaboration and Internet Computing (CIC)* (pp. 107–116). IEEE. https://doi.org/10.1109/CIC62241.2024.00023.

Ardimento, P., Bernardi, M. L., & Cimitile, M. (2024). Teaching UML using a RAG-based LLM. In *2024 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE. https://doi.org/10.1109/IJCNN60899.2024.10651492.

Balaguer, A., Benara, V., Cunha, R. L. de F., Filho, R. de M. E., Hendry, T., Holstein, D., Marsman, J., Mecklenburg, N., Malvar, S., Nunes, L. O., Padilha, R., Sharp, M., Silva, B., Sharma, S., Aski, V., & Chandra, R. (2024). RAG vs fine-tuning: Pipelines, tradeoffs, and a case study on agriculture, *In Computation and Language (cs.CL)* (No. arXiv:2401.08406). arXiv. https://doi.org/10.48550/arXiv.2401.08406.

Barnett, S., Kurniawan, S., Thudumu, S., Brannelly, Z., & Abdelrazek, M. (2024). *Seven failure points when engineering a retrieval augmented generation system* (No. arXiv:2401.05856). arXiv. https://doi.org/10.48550/arXiv.2401.05856.

Barochiya, M., Makhijani, P., Patel, H. N., Goel, P., & Patel, B. (2024). Evaluating RAG pipeline in multimodal LLM-based question answering systems. In *2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS)* (pp. 69–75). IEEE. https://doi.org/10.1109/ICACRS62842.2024.10841620.

Bernardi, M. L., Cimitile, M., & Pecori, R. (2024). Automatic job safety report generation using RAG-based LLMs. In *2024 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE. https://doi.org/10.1109/IJCNN60899.2024.10651320.

Bhandari, K., Kumar, K., & Sangal, A. L. (2023). Artificial intelligence in software engineering: Perspectives and challenges. In *2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC)* (pp. 133–137). IEEE. https://doi.org/10.1109/ICSCCC58608.2023.10176437.

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* (No. arXiv:2108.07258). arXiv. https://doi.org/10.48550/arXiv.2108.07258.

Bull, C. N., & Kharrufa, A. (2024). Generative AI assistants in software development education: A vision for integrating generative AI into educational practice, not instinctively defending against it. *IEEE Software*, 41(2), 52–59. https://doi.org/10.1109/MS.2023.3300574.

Chaubey, H. K., Tripathi, G., Ranjan, R., & Gopalaiyengar, S. K. (2024). Comparative analysis of RAG, fine-tuning, and prompt engineering in chatbot development. In *2024 International Conference on Future Technologies for Smart Society (ICFTSS)* (pp. 169–172). IEEE. https://doi.org/10.1109/ICFTSS61109.2024.10691338.

Chu, R., & Lim, S. C. J. (2023). Education and training for future engineering teachers in the age of artificial intelligence: A bibliometric analysis. In *2023 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 416–420). IEEE. https://doi.org/10.1109/IEEM58616.2023.10406630.

Cico, O., Cico, B., & Cico, A. (2023). AI-assisted software engineering: A tertiary study. In *2023 12th Mediterranean Conference on Embedded Computing (MECO)* (pp. 1–6). IEEE. https://doi.org/10.1109/MECO58584.2023.10154972.

Ciolacu, M. I., Marghescu, C., Mihailescu, B., & Svasta, P. (2024). Does Industry 5.0 need an Engineering Education 5.0? Exploring potentials and challenges in the age of generative AI. In *2024 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1–10). IEEE. https://doi.org/10.1109/EDUCON60312.2024.10578712.

Ciucu, R., Adochiei, F. C., Adochiei, I.-R., Argatu, F., Serițan, G. C., Enache, B., Grigorescu, S., & Argatu, V. V. (2019). Innovative DevOps for artificial intelligence. *The Scientific Bulletin of Electrical Engineering Faculty*, 19(1), Article 1. https://doi.org/10.1515/sbeef-2019-0011.

Cooper, M. M., & Klymkowsky, M. W. (2024). Let us not squander the affordances of LLMs for the sake of expedience: Using retrieval augmented generative AI chatbots to support and evaluate student reasoning. *Journal of Chemical Education*, 101(11), 4847–4856. https://doi.org/10.1021/acs.jchemed.4c00765.

Coutinho, M., Marques, L., Santos, A., Dahia, M., França, C., & De Souza Santos, R. (2024). The role of generative AI in software development productivity: A pilot case study. In *Proceedings of the 1st ACM International Conference on AI-Powered Software* (pp. 131–138). ACM. https://doi.org/10.1145/3664646.3664773.

Dakshit, S. (2024). *Faculty perspectives on the potential of RAG in computer science higher education* (No. arXiv:2408.01462). arXiv. https://doi.org/10.48550/arXiv.2408.01461.

Daniel, S., Olaoye, G., & Joseph, S. (2023). Using AI-powered techniques in DevSecOps to provide sturdy cloud security. *ResearchGate*. https://www.researchgate.net/publication/376682260.

Darshan, P., Mandahas, N., Mp, P. R., Rajesh, R. N., & N, D. (2024). Leveraging LLM and RAG for automated answer script evaluation. In *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)* (pp. 1–5). IEEE. https://doi.org/10.1109/CSITSS64042.2024.10817016

Deng, Z., Guo, Y., Han, C., Ma, W., Xiong, J., Wen, S., & Xiang, Y. (2024). *AI agents under threat: A survey of key security challenges and future pathways* (No. arXiv:2406.02630). arXiv. https://doi.org/10.48550/arXiv.2406.02630

Dieu, A. N. T., Nguyen, H. T., & Cong, C. T. D. (2024). The enhanced context for AI-generated learning advisors with advanced RAG. In *2024 18th International Conference on Advanced Computing and Analytics (ACOMPA)* (pp. 94–101). IEEE. https://doi.org/10.1109/ACOMPA64883.2024.00021

Dong, C., Yuan, Y., Chen, K., Cheng, S., & Wen, C. (2023). *How to build an AI tutor that can adapt to any course using knowledge graph-enhanced retrieval-augmented generation (KG-RAG)* [Preprint]. arXiv. https://arxiv.org/abs/2311.17696

Dong, X., & Jia, J. (2020). Teaching reform of software engineering course based on computational thinking. In *2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE)* (pp. 399–402). IEEE. https://doi.org/10.1109/ICAICE51518.2020.00084

Ferino, S., Fernandes, M., Cirilo, E., Agnez, L., Batista, B., Kulesza, U., Aranha, E., & Treude, C. (2023). Overcoming challenges in DevOps education through teaching method. In *2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)* (pp. 166–178). IEEE. https://doi.org/10.1109/ICSE-SEET58685.2023.00022

Ferino, S., Fernandes, M., Fernandes, A., Kulesza, U., Aranha, E., & Treude, C. (2021). Analyzing DevOps teaching strategies: An initial study. In *Brazilian Symposium on Software Engineering* (pp. 180–185). ACM. https://doi.org/10.1145/3474624.3477071

Fernandes, M., Ferino, S., Fernandes, A., Kulesza, U., Aranha, E., & Treude, C. (2022). DevOps education: An interview study of challenges and recommendations. In *Proceedings of the ACM/IEEE 44th International Conference on Software Engineering: Software Engineering Education and Training* (pp. 90–101). IEEE. https://doi.org/10.1145/3510456.3514152

Fernandes, M., Ferino, S., Kulesza, U., & Aranha, E. (2020). Challenges and recommendations in DevOps education: A systematic literature review. In *Proceedings of the XXXIV Brazilian Symposium on Software Engineering* (pp. 648–657). ACM. https://doi.org/10.1145/3422392.3422496

Fleischer, D., Berchansky, M., Wasserblat, M., & Izsak, P. (2024). *RAG Foundry: A framework for enhancing LLMs for retrieval augmented generation* (No. arXiv:2408.02545). arXiv. https://doi.org/10.48550/arXiv.2408.02545

Gong, X. (2021). Educational artificial intelligence (EAI) connotation, key technology and application trend: Interpretation and analysis of the two reports entitled "Preparing for the Future of Artificial Intelligence" and "The National Artificial Intelligence Research and Development Strategic Plan". In *2021 International Conference on Intelligent Computing, Automation and Applications (ICAA)* (pp. 219–223). IEEE. https://doi.org/10.1109/ICAA53760.2021.00046

Grote, M., & Bogner, J. (2023). A case study on AI engineering practices: Developing an autonomous stock trading system. In *2023 IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN)* (pp. 145–157). IEEE. https://doi.org/10.1109/CAIN58948.2023.00032

Guettala, M., Bourekkache, S., Kazar, O., & Harous, S. (2024). Building advanced RAG Q&A with multiple data sources using Langchain: A multi-search agent RAG application in ubiquitous learning. In *2024 2nd International Conference on Computing and Data Analytics (ICCDA)* (pp. 1–7). IEEE. https://doi.org/10.1109/ICCDA64887.2024.10867361

He, P., Wang, S., Chowdhury, S., & Chen, T.-H. (2025). *Evaluating the effectiveness and efficiency of demonstration retrievers in RAG for coding tasks* (No. arXiv:2410.09662). arXiv. https://doi.org/10.48550/arXiv.2410.09662

Heck, P., & Schouten, G. (2021). Lessons learned from educating AI engineers. In *2021 IEEE/ACM 1st Workshop on AI Engineering – Software Engineering for AI (WAIN)* (pp. 1–4). IEEE. https://doi.org/10.1109/WAIN52551.2021.00013

Hemmat, A., Vadaei, K., Heydari, M. H., & Fatemi, A. (2024). *Leveraging retrieval-augmented generation for Persian university knowledge retrieval* (No. arXiv:2411.06237). arXiv. https://doi.org/10.48550/arXiv.2411.06237

Iscan, C., Ozara, M. F., & Akbulut, A. (2024). Enhancing RAG pipeline performance with translation-based embedding strategies for non-English documents. In *2024 Innovations in Intelligent Systems and Applications Conference (ASYU)* (pp. 1–6). IEEE. https://doi.org/10.1109/ASYU62119.2024.10756977

Iyer, G. N., Yisheng, A. G., Er Metilda, C. H., Choong, W. X., & Koh, S. W. (2024). A web-based IDE for DevOps learning in software engineering higher education. In *2024 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)* (pp. 1–8). IEEE. https://doi.org/10.1109/TALE62452.2024.10834361

Jabbari, R., Bin Ali, N., Petersen, K., & Tanveer, B. (2016). What is DevOps? A systematic mapping study on definitions and practices. *In Proceedings of the Scientific Workshop Proceedings of XP2016* (pp. 1–11). ACM. https://doi.org/10.1145/2962695.2962707

Jawalkar, A. A., Gothane, S., & Bruno, A. (2024). Generative AI: A structured review, techniques, application and future prospects. *International Journal of Research in Advent Technology*, *12*(4), 14–20. https://doi.org/10.32622/ijrat.124202408

Jeong, C. (2024). *A study on the implementation method of an agent-based advanced RAG system using graph* (Version V3). arXiv. https://doi.org/10.48550/arXiv.2407.19994

Jiao, D., Cai, L., Huang, J., Zhang, W., Tang, S., & Zhuang, Y. (2024). *DuetRAG: Collaborative retrieval-augmented generation* (No. arXiv:2405.13002). arXiv. https://doi.org/10.48550/arXiv.2405.13002

Johnson, M. (2024). Generative AI and CS education. *Communications of the ACM*, 67(4), Article 4. https://doi.org/10.1145/3632523

Kästner, C., & Kang, E. (2020). Teaching software engineering for AI-enabled systems. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering Education and Training* (pp. 45–48). ACM. https://doi.org/10.1145/3377814.3381714

Kim, D. J., Locke, S., Chen, T.-H. P., Toma, A., Sporea, S., Weinkam, L., & Sajedi, S. (2023). Challenges in adopting artificial intelligence based user input verification framework in reporting software systems. In *2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)* (pp. 99–109). IEEE. https://doi.org/10.1109/ICSE-SEIP58684.2023.00014

Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering: A systematic literature review. *Information and Software Technology*, 51(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009

Kumar, R., Naveen, V., Kumar Illa, P., Pachar, S., & Patil, P. (2023). The current state of software engineering employing methods derived from artificial intelligence and outstanding challenges. In *2023 1st International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP)* (pp. 105–108). IEEE. https://doi.org/10.1109/IHCSP56702.2023.10127112

Kumar, T., Garg, V., Lalar, S., & Kumar, R. (2024). Measuring impact of generative AI in software development and innovation. In B. Shukla, B. K. Murthy, N. Hasteer, H. Kaur, & J.-P. Van Belle (Eds.), *Intelligent IT solutions for sustainability in Industry 5.0 paradigm* (Vol. 1185, pp. 57–67). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-1682-1_6

Kuratomi, G., Pirozelli, P., Cozman, F. G., & Peres, S. M. (2025). *A RAG-based institutional assistant* (No. arXiv:2501.13880). arXiv. https://doi.org/10.48550/arXiv.2501.13880

Li, H., Su, Y., Cai, D., Wang, Y., & Liu, L. (2022). *A survey on retrieval-augmented text generation* (No. arXiv:2202.01110). arXiv. https://doi.org/10.48550/arXiv.2202.01110

Lo, D. (2023). Trustworthy and synergistic artificial intelligence for software engineering: Vision and roadmaps. In *2023 IEEE/ACM International Conference on Software Engineering: Future of Software Engineering (ICSE-FoSE)* (pp. 69–85). IEEE. https://doi.org/10.1109/ICSE-FoSE59343.2023.00010

Ma, I., Martins, A. K., & Lopes, C. V. (2024). *Integrating AI tutors in a programming course* (No. arXiv:2407.15718). arXiv. https://doi.org/10.48550/arXiv.2407.15718

Mekić, E. S., Jovanović, M. N., Kuk, K. V., Prlinčević, B. P., & Savić, A. M. (2024). Enhancing educational efficiency: Generative AI chatbots and DevOps in Education 4.0. *Computer Applications in Engineering Education*, 32(6), e22804. https://doi.org/10.1002/cae.22804

Morić, Z., Mršić, L., Filjak, M., & Đambić, G. (2024). Integrating a virtual assistant by using the RAG method and VERTEX AI framework at Algebra University. *Applied Sciences*, 14(22), 10748. https://doi.org/10.3390/app142210747

Mota, L. L., Santos, R. P. D., Fontão, A., & Araújo, A. A. (2024). From theory to interpreting the practice: Exploring the role-play to teach DevOps. In *Anais do XXXVIII Simpósio Brasileiro de Engenharia de Software (SBES 2024)* (pp. 444–454). SBC. https://doi.org/10.5753/sbes.2024.3519

Nguyen, T., Chin, P., & Tai, Y.-W. (2024). *Reward-RAG: Enhancing RAG with reward driven supervision* (No. arXiv:2410.03780). arXiv. https://doi.org/10.48550/arXiv.2410.03780

Pan, C., You, J., & Gao, Y. (2023). Survey on reliability engineering for AI software systems: An extension based on the IEEE 1633 standard. In *2023 3rd International Symposium on Artificial Intelligence and Intelligent Manufacturing (AIIM)* (pp. 116–121). IEEE. https://doi.org/10.1109/AIIM60438.2023.10441228

Pickett, M., Hartman, J., Bhowmick, A. K., Alam, R., & Vempaty, A. (2025). *Better RAG using relevant information gain* (No. arXiv:2407.12101). arXiv. https://doi.org/10.48550/arXiv.2407.12101

Rajeshwari, S. B., Mane, P., Sanjeetha, R., Kallimani, J. S., Balajee, R. M., & Kale, L. (2024). Innovative pedagogical approaches for teaching cloud-native DevOps: Integrating theory and practice. *Nanotechnology Perceptions, 20*(S6), 760–768. http://www.nano-ntp.com

Rocha, C., Alves, I., Pérez-Martínez, J. E., & Diaz, J. (2023). *Learning DevOps with an industrial perspective in the software engineering curricula* [Preprint]. TechRxiv. https://doi.org/10.36227/techrxiv.22096535.v1

Saha, B., & Saha, U. (2024). Enhancing international graduate student experience through AI-driven support systems: A LLM and RAG-based approach. In *2024 International Conference on Data Science and Its Applications (ICoDSA)* (pp. 300–304). IEEE. https://doi.org/10.1109/ICoDSA62899.2024.10651944

Saha, B., Saha, U., & Malik, M. Z. (2024). Advancing retrieval-augmented generation with inverted question matching for enhanced QA performance. *IEEE Access*, 1–1. https://doi.org/10.1109/ACCESS.2024.3513155

Sauvola, J., Tarkoma, S., Klemettinen, M., Riekki, J., & Doermann, D. (2024). Future of software development with generative AI. *Automated Software Engineering, 31*(1), Article 1. https://doi.org/10.1007/s10515-024-00426-z

Shi, L., Kazda, M., Sears, B., Shropshire, N., & Puri, R. (2024). Ask-EDA: A design assistant empowered by LLM, hybrid RAG and abbreviation de-hallucination. In *2024 IEEE LLM Aided Design Workshop (LAD)* (pp. 1–5). IEEE. https://doi.org/10.1109/LAD62341.2024.10691824

Shin, J., Aleithan, R., Hemmati, H., & Wang, S. (2024). *Retrieval-augmented test generation: How far are we?* (No. arXiv:2409.12682). arXiv. https://doi.org/10.48550/arXiv.2409.12682

Simon, S., Mailach, A., Dorn, J., & Siegmund, N. (2024). *A methodology for evaluating RAG systems: A case study on configuration dependency validation* (No. arXiv:2410.08801). arXiv. https://doi.org/10.48550/arXiv.2410.08801

Singh, A., Ehtesham, A., Kumar, S., & Khoei, T. T. (2025). Agentic retrieval-augmented generation: A survey on agentic RAG (arXiv:2501.09136). arXiv. https://doi.org/10.48550/arXiv.2501.09136

Slimi, Z. (2023). The impact of artificial intelligence on higher education: An empirical study. *European Journal of Educational Sciences, 10*(1). https://doi.org/10.19044/ejes.v10no1a17

Thüs, D., Malone, S., & Brünken, R. (2024). Exploring generative AI in higher education: A RAG system to enhance student engagement with scientific literature. *Frontiers in Psychology, 15*, 1474892. https://doi.org/10.3389/fpsyg.2024.1474892

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", & O., N. (2023). Using retrieval-augmented generation to elevate low-code developer skills. *Artificial Intelligence, 28*(3), 126–130. https://doi.org/10.15407/jai2023.03.126

Uysal, D. M. (2025). *A comprehensive review of generative AI: Concepts, leading products, and performance comparison* (Vol. 7). https://doi.org/10.36948/ijfmr.2025.v07i01.35199

Vetriselvi, T., Mathur, M., & Bhuvaneswari, M. (2024). Applying generative AI to create SOP, reducing API costs through prompt compression and evaluating LLM responses with Tonic Validate RAG metrics. In *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)* (pp. 10–18). IEEE. https://doi.org/10.1109/ICUIS64676.2024.10867024

Vierhauser, M., Groher, I., Antensteiner, T., & Sauerwein, C. (2024). Towards integrating emerging AI applications in SE education. In *2024 36th International Conference on Software Engineering Education and Training (CSEE&T)* (pp. 1–5). IEEE. https://doi.org/10.1109/CSEET62301.2024.10663045

Virvou, M., & Tsihrintzis, G. A. (2023). Pre-made empowering artificial intelligence and ChatGPT: The growing importance of human AI-experts. In *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1–8). IEEE. https://doi.org/10.1109/IISA59645.2023.10345881

Wang, W., Ma, J., Zhang, P., Hu, Z., Jiang, Q., & Liu, Y. (2024). Application of multi-way recall fusion reranking based on tensor and ColBERT in RAG. In *2024 IEEE 7th International Conference on Information Systems and Computer Aided Education (ICISCAE)* (pp. 138–141). IEEE. https://doi.org/10.1109/ICISCAE62304.2024.10761558

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *In Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* (pp. 1–10). https://doi.org/10.1145/2601248.2601268

Yabaku, M., Pombo, N., & Ouhbi, S. (2024). Exploring the potential use of generative AI in software engineering education. In *2024 IEEE 18th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1–7). IEEE. https://doi.org/10.1109/AICT61888.2024.10740416

Zhang, H., Song, J., Zhu, J., Wu, Y., Zhang, T., & Niu, C. (2025). *RAG-Reward: Optimizing RAG with reward modeling and RLHF* (No. arXiv:2501.13264). arXiv. https://doi.org/10.48550/arXiv.2501.13264

Zhang, Q., Chen, S., Bei, Y., Yuan, Z., Zhou, H., Hong, Z., Dong, J., Chen, H., Chang, Y., & Huang, X. (2025). *A survey of graph retrieval-augmented generation for customized large language models* (No. arXiv:2501.13958). *arXiv*. https://doi.org/10.48550/arXiv.2501.13958

Zhang, Y. (2023). Construction of computer software development system based on artificial intelligence. In *2023 International Conference on Mechatronics, IoT and Industrial Informatics (ICMIII)* (pp. 597–601). IEEE. https://doi.org/10.1109/ICMIII58949.2023.00125

Zhu, C. (2024, July 12). Artificial intelligence applications in everyday life. *In Proceedings of the 2024 7th International Conference on Signal Processing and Machine Learning (SPML 2024)*, Qingdao, China. https://doi.org/10.54254/2755-2721/121/2025.20074

Zotero. (2025). *Zotero | Your personal research assistant*. Retrieved February 25, 2025, from https://www.zotero.org/

Zúñiga Sánchez, O. (2024). El impacto de ChatGPT en la formación y producción académica: Que no cunda el pánico. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo, 14*(28). https://doi.org/10.23913/ride.v14i28.1867.