INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Generative Artificial Intelligence for Education: The Present and the Future

Teresa Vaz Reynaud

Master in Computer Science and Business Management

Supervisor:
PhD Rúben Filipe de Sousa Pereira, Assistant Professor,
ISCTE - University Institute of Lisbon

September, 2024

iscte

TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technology

Generative Artificial Intelligence for Education: The Present and the Future

Teresa Vaz Reynaud

Master in Computer Science and Business Management

Supervisor:
PhD Rúben Filipe de Sousa Pereira, Assistant Professor,
ISCTE - University Institute of Lisbon

September, 2024

“Change, Achievement, Self-Growing”

- Pedro Ameixa

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, PhD Rúben Filipe de Sousa Pereira, for his constant support and guidance throughout my master's journey. His belief in my abilities and exceptional mentorship allowed me to explore my research interests freely. His encouragement and dedication gave me the tools and confidence to articulate our findings over the past year. Thank you for being an incredible professor and an inspiring project manager.

I would also like to extend my heartfelt thanks to my family. Each member, in their unique way, has contributed to shaping the person I am today, enabling me to undertake and accomplish this investigation. Your love and support have been invaluable.

Finally, a special thanks to ISCTE-IUL for being my academic home for the past five years. ISCTE-IUL has provided me with numerous opportunities for growth and learning. Thank you for everything you have offered me during this incredible journey.

Resumo

O crescimento da inteligência artificial (IA) generativa, como o ChatGPT, DALL-E, Bing AI e Copilot, tem demonstrado um impacto significativo em diversos setores, incluindo a educação. Estas ferramentas capazes de gerar conteúdo semelhante ao produzido por humanos, apresentam tanto oportunidades quanto desafios para a educação. Este estudo realiza uma revisão sistemática da literatura, com um conjunto de 121 artigos publicados entre 2018 e 2023, oferecendo uma visão abrangente da IA generativa na educação. Explora as aplicações e benefícios, as limitações, riscos e preocupações, os níveis de educação abordados e recomendações futuras, assim como as ferramentas mencionadas.

Os resultados revelam que a IA generativa tem o potencial de transformar a educação, possibilitando uma aprendizagem personalizada, feedback em tempo real e suporte para os educadores no planeamento de aulas e criação de material. No entanto, surgiram preocupações como o risco de desonestidade académica, a dependência excessiva nestas ferramentas e um possível declínio nas competências de pensamento crítico e resolução de problemas. Questões como informação falsa, conteúdo tendencioso e considerações éticas reforçam a necessidade de uma integração cautelosa da IA.

São fornecidas recomendações para que as instituições de ensino integrem estas ferramentas, ao mesmo tempo que mitigam os riscos associados. Isto inclui a incorporação de IA no currículo, a inovação dos métodos de avaliação e a promoção de um uso equilibrado da tecnologia que complemente as práticas de ensino tradicionais. Esta investigação oferece uma perspetiva valiosa sobre a IA generativa na educação, proporcionando insights para educadores, instituições e investigadores.

Palavras-chave: Inteligência Artificial, IA Generativa, Educação, IA na Educação, Revisão Sistemática da Literatura

Abstract

The advancements in generative artificial intelligence (genAI) technologies, such as ChatGPT, DALL-E, Bing AI, and Copilot, have significantly disrupted various sectors, including education. These tools are capable of generating human-like content and present opportunities and challenges for education. This study conducted a systematic literature review (SLR) of 121 peer-reviewed articles published between 2018 and 2023, offering a comprehensive overview of the current research on genAI in education. It explores the applications and benefits, limitations, risks and concerns, the educational levels addressed, future recommendations, and the genAI tools mentioned in the literature.

The findings reveal that genAI has the potential to transform education by offering personalised learning experiences, real-time feedback, and support for educators in lesson planning and content generation. On the other hand, significant concerns have emerged, including the risk of academic dishonesty, over-reliance on AI tools, and a potential decline in students' critical thinking and problem-solving skills. Additionally, issues related to misinformation, biased content, and ethical considerations emphasise the need for a cautious and well-regulated integration of genAI in educational settings.

Recommendations are provided for institutions to integrate these tools while mitigating associated risks effectively. This includes incorporating AI into the curriculum, updating assessment methods to prevent academic misconduct, and encouraging a balanced use of technology that complements traditional teaching practices. This research offers a valuable perspective on the evolving role of genAI in education, providing insights for educators, institutions, and researchers.

Keywords: Artificial Intelligence, Generative AI, Education, AI in Education, Systematic Literature Review

Index

Glossary

AGI - Artificial General Intelligence

AI - Artificial Intelligence

ANI - Artificial Narrow Intelligence

ASI - Artificial Super Intelligence

GenAI - Generative Artificial Intelligence

LLM - Large Language Model

ML - Machine Learning

NLP - Natural Language Processing

SLR - Systematic Literature Review

List of Figures

List of Tables

CHAPTER 1

Introduction

With the rapid advancements in technology, artificial intelligence (AI) has experienced exponential growth, particularly in the field of natural language processing (NLP) [1]. Generative AI (genAI) tools, such as ChatGPT, Bard [2], or GitHub Copilot [3], have the ability to generate human-like responses given a prompt [2], thereby disrupting today's society [4]. These technological advancements have transformed numerous sectors [1], leading to the need to comprehend both the applications and limitations of genAI tools in industries [5], including healthcare [6], journalism and media [7], and education [8].

In education, the impact of genAI has become a highly debated topic [9], particularly since November 2022 [10], ChatGPT has seen a rapid increase in users, reaching a million within days of its launch. [5]. ChatGPT is a large language model (LLM) chatbot trained on a massive dataset to generate human-like responses [10] in real-time conversations [11]. These responses can be used for completing school homework or to answer exam-type questions [12], offering various benefits to both students and educators.

Students can receive personalised feedback on different topics, while teachers can receive helpful suggestions [13] to conduct classes, as these tools function as virtual tutors [11]. Additionally, genAI tools can contribute to time management [4], by facilitating brainstorming [10] and offering immediate feedback on specific topics or issues [14].

However, concerns arise with the widespread use of these tools [15]. Ethical responsibility [2] is crucial, particularly in education, where issues such as plagiarism [13] and academic dishonesty [16] are prevalent. A content analysis study [17] on the impact of ChatGPT in higher education revealed that most articles discussed plagiarism and academic dishonesty as the main concerns with implementing this tool in universities.

Furthermore, since these tools can generate responses and conduct human-like conversations, students might get used to this feature and start using it without thinking for themselves [18]. This reliance on genAI might decrease critical thinking skills, as students may use these tools without engaging in independent thought. Someone who uses critical thinking will be better prepared to identify the inaccuracies in these tools [11], as some authors have found that genAI sometimes provides fake references [19], false information [2], [20], and biased content [21], [22], due to limitations in the quality of its training dataset [13].

The integration of genAI in education has initiated discussions about how universities should respond to the technology [4]. Recent studies have suggested different approaches to address these concerns, such as training students to use this technology ethically [7], assignments focusing on particular skills, such as critical thinking [12], and adapting teaching methods to this new era [23].

Given that genAI is an emerging technology, there needs to be more understanding of its use in education. Sullivan et al. [17] focus on how ChatGPT is disrupting higher education in specific geographic areas, while others examine the perceptions of scholars and students regarding what ChatGPT means for universities [5]. The proposed research will offer a perspective on all academic levels, including higher education.

This research aims to provide valuable insights into integrating genAI in education through a systematic literature review (SLR). It will explore the potential *advantages* and limitations of genAI in education and provide recommendations for educational institutions, educators, and students. Additionally, it will assess the genAI tools discussed in the literature.

To gain a better understanding of the various forms of AI, including genAI, Chapter 2 will begin with a theoretical background defining important concepts that serve this investigation. Chapter 3 will present the methodology that was used, followed by a description of the motivation and review protocol for this research. Chapter 5 will cover the selected studies for this review and include a data extraction analysis. Chapter 6 will report the SLR, where the data from the relevant literature will be assessed. Finally, Chapter 7 will present the reflections and insights derived from the findings of this study, offering a critical analysis of the reported outcomes and Chapter 8 will conclude this work.

CHAPTER 2

Theoretical Background

When exploring the complex challenge of defining AI, Wang's study [24] provides insight into the various interpretations associated with this term. McCarthy, a pioneering figure in the modern AI research field, characterises AI as the science and engineering dedicated to crafting intelligent machines, particularly in the form of sophisticated computer programs [25].

AI can be categorised into three distinct types [26]: Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI), and Artificial Super Intelligence (ASI). The distinction between them lies in their scope and capabilities. ANI is designed to perform single tasks, lacking the ability to adapt to tasks outside its predefined domain [27]. It emulates human-like intelligence within specific conditions and predefined contexts [28]. While ANI is confined to structured examples and specific domains, AGI can transfer knowledge through them [29] without that confinement. AGI is a type of AI that equates to human intelligence [26], which can learn by itself [29]. Concerning ASI, it is the AI that would exceed human intelligence [26]. AGI and ASI are theoretical concepts yet to be achieved [30]. All the NLP sophisticated tools, such as Apple Siri or Google Translate, fall into the ANI category [27].

Machine learning (ML), a subfield of AI, is specifically designed to leverage data and algorithms to mimic human learning [31], improving its accuracy through experience [32]. ML has a subfield named deep learning [26], which involves neural networks with multiple layers [33]. It is designed to behave as the human brain does by recognising patterns and having the ability to generate outputs such as insights or predictions [34]. An example of a technology using deep neural networks is ChatGPT [35], which is a type of AI that falls into the category of LLMs. An LLM is trained on a vast amount of data to be able to generate human-like responses with the utmost accuracy [36]. While LLMs are related to generating natural text, others generate code (e.g., GitHub Copilot [3]), or images (e.g., DALL-E [10]).

Therefore, the concept of genAI embraces all these AI models that generate content regardless of the type [37]. The output generated by these models depends on the instructions they are provided in the prompt, and they can be improved with new ones until the output generated is the expected one [38]. This process, known as prompt engineering [39], involves designing and testing inputs to achieve the desired result.

CHAPTER 3

Methodology

An SLR is a structured approach that aims to identify, evaluate, and synthesise all relevant studies and findings to provide a comprehensive overview of current knowledge of a chosen topic [40]. An SLR encompasses only formal literature, also known as white literature, which consists of high-credibility sources such as journal articles, conference proceedings, and published books [41] as specified in Table 3.1.

Table 3.1. Spectrum of the White, Grey, and Black Literature [41].

White Literature	Grey Literature	Black Literature
Published journal papers	Preprints	Ideas
Conference Proceedings	e-Prints	Concepts
Books	Technical reports	Thoughts
	Lectures	
	Data sets	
	Audio-Video media	
	Blogs	

This research follows the guidelines outlined by Kitchenham [40], which can be broken down into three phases. Figure 3.1 illustrates these phases, which include identifying the need for a review, conducting the review, and reporting the findings. Each phase corresponds to specific tasks and objectives that ensure the comprehensiveness and rigour of the SLR process.

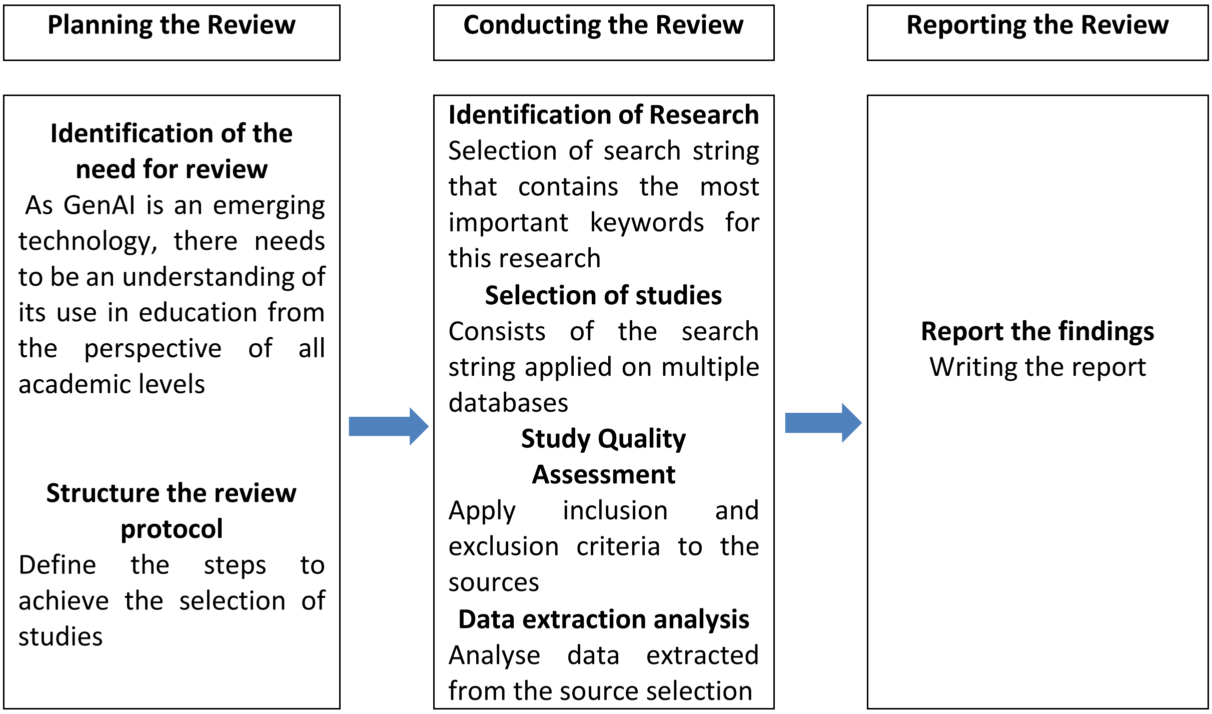

Figure 3.1. Phases Adopted in this Research.

The initial phase involves describing the plan for this review, which introduces the motivation for this work and the structured plan to conduct it. The second phase focuses on executing the review by designing the search string to be applied in selecting databases to retrieve the article dataset. This selection will correspond to the foundation of this review and meet specific standards of quality and relevance. Additionally, a thorough analysis of the gathered data will be presented.

Finally, the third phase corresponds to the findings report. By following these structured phases, this research aims to deliver a thorough review of the current knowledge regarding the use of genAI in education. The subsequent chapters will explore the specific methodologies employed and the findings derived from this systematic approach.

CHAPTER 4

Planning the Review

This section corresponds to the initial phase of the SLR process, outlining the motivation for this research and the intended protocol to obtain the final document set used in the subsequent phases.

4.1. Need for a review

The motivation for this work is to provide a comprehensive understanding of the use of genAI in education, including its potential benefits and limitations. As genAI is an emerging technology, there is a need to explore its applications and implications in the field of education [42]. Therefore, this study aims to provide valuable insights to educators, students, and institutions by conducting an SLR of relevant literature, enabling informed decisions about using this technology in education.

4.2. Review Protocol

This section outlines the systematic approach used to conduct and report the review of relevant literature. The protocol is designed to ensure a comprehensive and unbiased selection of studies and is structured as follows:

- Design a search string.
- Apply the search string to databases.
- Apply inclusion and exclusion criteria.
- Review abstracts.
- Identify and remove duplicate entries.
- Conduct a detailed full-text assessment.
- Snowballing to identify additional studies.
- Final document set of selected studies.

The initial step involves designing a search string using relevant keywords that encapsulate the main themes of the research topic. This search string was applied to four major databases: Scopus, Web of Science, IEEE, and ACM. The choice of these databases ensures a comprehensive coverage of high-quality, peer-reviewed literature.

To refine the initial set of retrieved documents, specific inclusion and exclusion criteria were established. Inclusion criteria involve selecting studies based on factors such as publication date, language (English), and mentions of the search string. Exclusion criteria are used to remove studies that do not meet these standards, such as those with unidentified authors, those not related to the topic or those published in languages other than English. The criteria applied are outlined in the next chapter.

Following this, the abstracts of the remaining documents were reviewed to determine their relevance. Additionally, duplicate entries, which can occur when multiple databases are searched, were identified and removed to avoid redundancy and maintain a unique set of documents.

A detailed evaluation of each study's entire content was performed to judge its relevance, quality, and contribution to the research topic. To enhance the comprehensiveness of the literature review, a snowballing technique was employed. This involved examining the references cited in the selected studies and identifying additional relevant studies that have cited these documents. This method helped discover further literature that may not have been captured in the initial search.

The final document set includes studies rigorously selected through the above steps. This set represents the literature that forms the basis of the review. The documents that are most relevant and valuable to the research topic will be analysed in depth in the subsequent chapters.

CHAPTER 5

Conducting the Review

This section refers to the second phase of the SLR, where the review protocol defined in the previous chapter will be followed to achieve a final selection of studies. A search string will be identified and applied to the databases to obtain the array of white literature. The resulting documents will undergo multiple filters to narrow the set to the most relevant ones. The final selection results will be transformed into data for further analysis.

5.1. Identification of Research

In June 2023, a comprehensive search was conducted across Scopus, Web of Science, IEEE, and ACM to select studies for this research. A specific search string incorporating critical keywords was carefully designed to ensure the retrieval of the most relevant documents. Given that this study aims to enhance the understanding of genAI in education, the search string was designed to include keywords relevant to both areas. This approach led to the creation of a search string outlined in Table 5.1.

Table 5.1. Search String Applied in the Dataset Engines.

Search String	(generative-ai OR genAI OR chatgpt OR bard OR bing-ai) AND (education OR academics OR university)

5.2. Selection of Studies

Selecting the studies for this research involved multiple steps to ensure the most relevant ones were chosen. After designing the search string, the first step for this process corresponds to the first filter in Table 5.2. Applying this string to the chosen databases results in an initial set of 255,719 documents.

Table 5.2. Filters Used in the SLR Protocol.

Database	**Filter 1**	**Filter 2**	**Filter 3**	**Filter 4**	**Filter 5**	**Filter 6**	**Snowballing**
Scopus	238,585	335	179	59	41	29	3
Web Of Science	14,522	192	130	120	81	63	
IEEE	200	34	32	32	29	18	
ACM	2,412	15	14	14	13	8	
Total	255,719	576	355	225	164	118	**121**

Legend: Filter 1 - Query All Fields; Filter 2 - Query Abstract; Filter 3 - Apply Inclusion/ Exclusion Criteria; Filter 4- Remove Duplicates; Filter 5 - After Abstract Screened; Filter 6 - Full-text Document Assess;

A second filter was then applied, reducing the number to 576 documents by excluding each abstract that did not mention the keywords in the search string. Following this, a third filter based on the inclusion and exclusion criteria outlined in Table 5.3 was applied, resulting in 221 articles being removed.

Further narrowing down the set, 130 duplicate documents were identified and removed, leaving 225 documents. An abstract screen was then performed to validate the topic of each article, resulting in sixty-one being removed. The next step involved a full-text assessment to confirm the relevance of the selected articles to the research topic. This final filter led to the removal of forty-six articles. During the full-document assessment, relevant articles identified were subsequently included in the dataset. This iterative process, referred to as snowballing, added three more studies. The final selection comprised 121 studies.

5.3. Study Quality Assessment

Establishing clear inclusion and exclusion criteria ensures that the literature reviewed is relevant, high-quality, and directly applicable to the topic in question [40]. The inclusion criteria ensure that only studies meeting specific standards of relevance and quality are considered. These criteria are as follows:

- Publication Date: To ensure the research is up-to-date and relevant, we include only studies published from 2018 onwards.
- Relevance to Search String: Studies must explicitly mention or address the keywords used in the search string to ensure they are relevant to the research topic.
- Access: Studies must be free and accessible.

The exclusion criteria help filter out studies that do not meet the necessary standards or are irrelevant to the research topic. These criteria are as follows:

- Unidentified Authors: Studies with unidentified or anonymous authors are excluded.
- Non-English Publications: Studies published in languages other than English are excluded.
- Lack of publication date: Studies published without a publication date are excluded.

These criteria are summarised in Table 5.3, and they were systematically applied to filter the initial set of documents, corresponding to the third filter in Table 5.2.

Table 5.3. Inclusion and Exclusion Criteria Applied in this Research.

Inclusion Criteria	**Exclusion Criteria**
Published 2018 onwards	No Publication Date
Mentions Of Search String	Not Written in English
Full-Text Accessible	Unidentified Author

5.4. Data Extraction Analysis

Upon completion of the selection process, a final set of white literature was obtained. This section presents the data extraction analysis for a total of 121 publications. It begins by examining the distribution of the final set by illustrating the number of publications contained in each database. Following this, Figure 5.2 demonstrates a distribution based on the type of document, and then, to conclude this analysis, Figure 5.3 explores the distribution according to the document's publication year.

5.4.1. Distribution of Studies per Data Engine

The pie chart in Figure 5.1 illustrates the distribution of the final set of documents across different data engines. Figure 5.1 reveals that Web of Science contains the highest number of publications, accounting for 63 out of the 121 documents. This dominance suggests that this database is a crucial source for high-quality research in the field of genAI and education, highlighting its importance in this review. The remaining are distributed as follows: twenty-nine from Scopus, eighteen from IEEE, and eight from ACM.

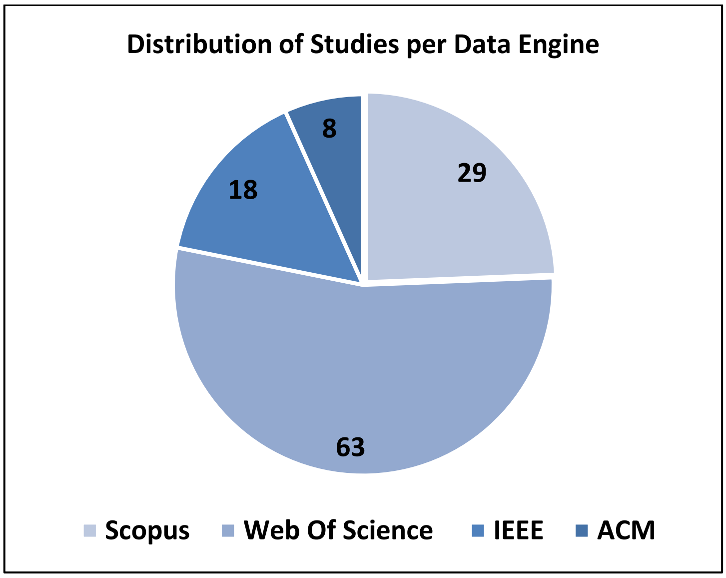

Figure 5.1. Distribution of the Final Set of Publications per Database.

5.4.2. Distribution of Studies per Type of Document

Another aspect to analyse is the categories of documents in the final set and their distribution. For the current study, Figure 5.2 displays the distribution of the final set of publications by document type. Most of the articles are journal articles (99), followed by 21 conference papers and 1 report.

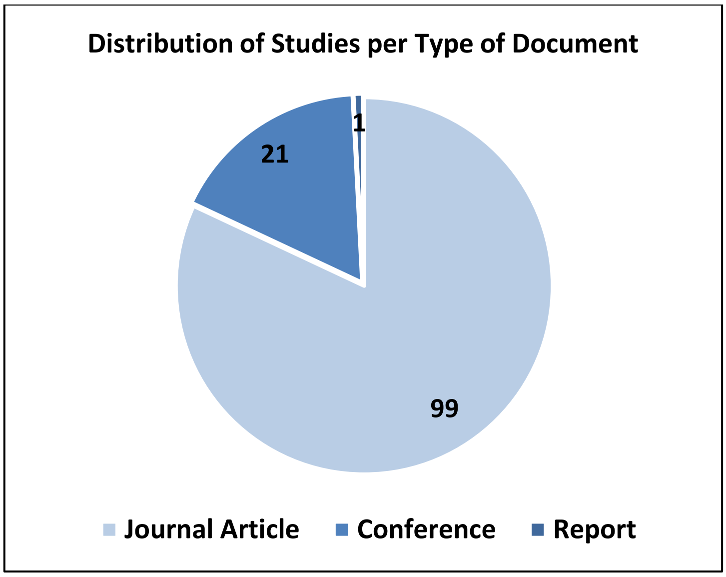

Figure 5.2. Distribution of the Final Set of Publications per Document Type.

5.4.3. Distribution of Studies per Year

As described in section 5.1, the period chosen for this research was from 2018 onwards. However, as demonstrated in Figure 5.3, the final selected studies contain six articles from the year 2022 and 115 for 2023, resulting in an increase of one hundred and nine articles from one year to the next. This data supports the need to research genAI in education since it is a recent topic, as Strzelecki [42] mentioned.

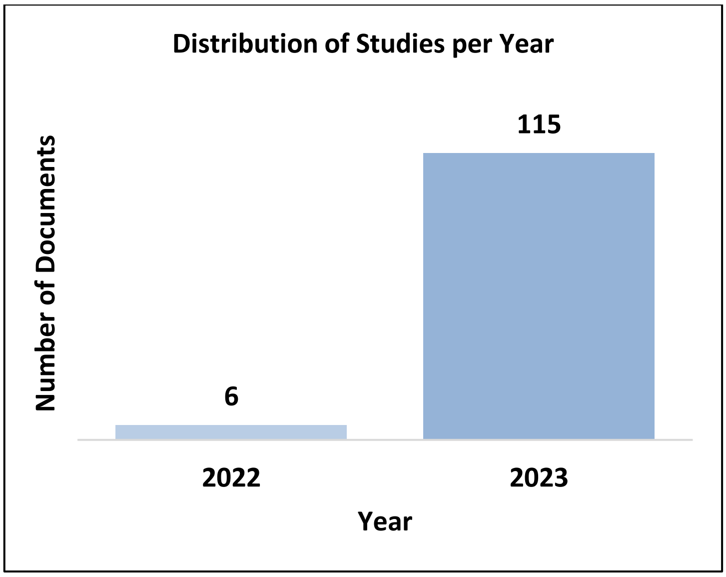

Figure 5.3. Distribution of the Final Set of Publications per Year.

Given the need to examine the applications and concerns of these tools [43] in education, the subsequent chapter will present the findings derived from the analysis of the final selection of studies, which have been filtered according to the criteria outlined Table 5.2.

CHAPTER 6

Reporting the Review

This section represents the final phase of the SLR, where the analysis of selected studies and the corresponding findings are presented. Through thoroughly examining relevant articles, five major areas have been identified as crucial for understanding the current literature surrounding genAI and its applications in education.

The five key areas are Applications and Benefits, Limitations, Risks and Concerns, Future Suggestions, Educational Level and GenAI Tools. Each of these main themes will be expanded upon in the following subsections, providing detailed insights into the review findings from each perspective. To offer a comprehensive understanding, each area is associated with several subcategories. The following conceptual map in Figure 6.1 provides a visual representation of the main areas identified and exemplifies some of each area's categories.

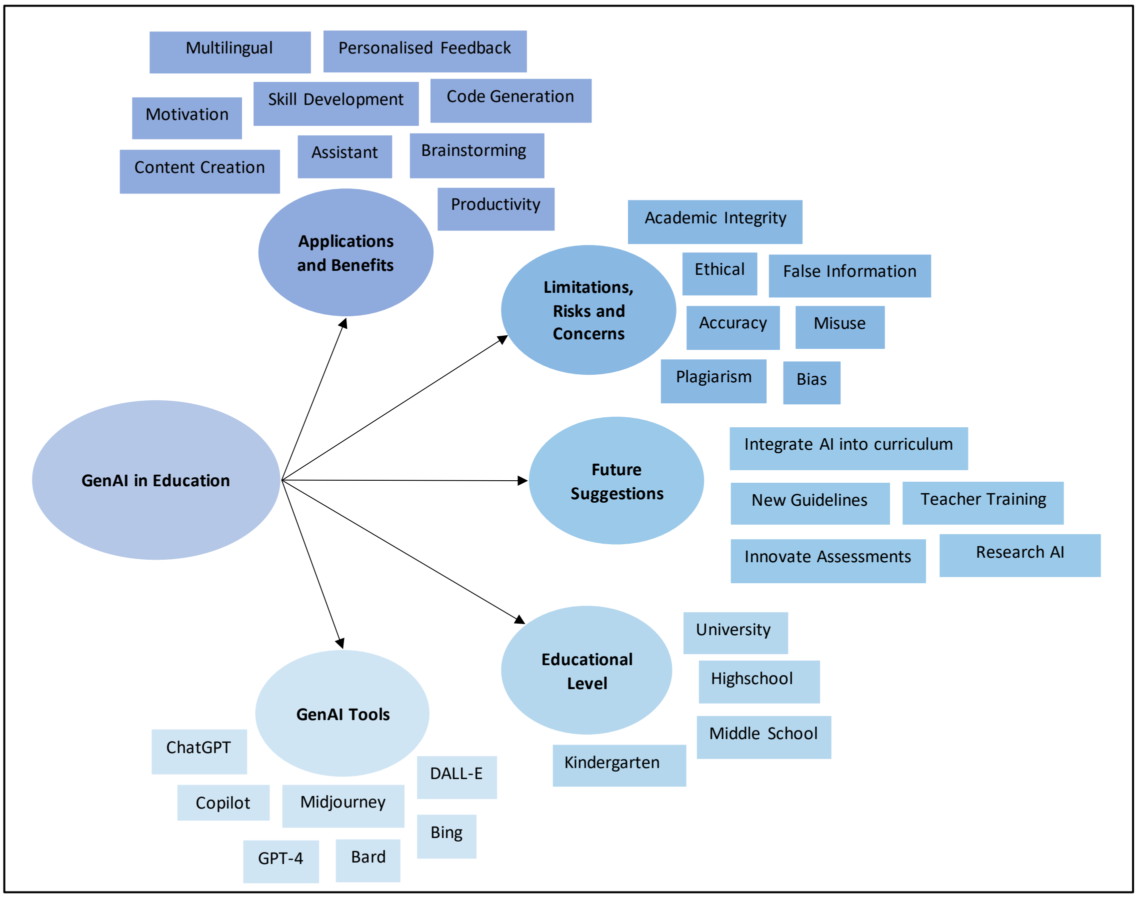

Figure 6.1. Conceptual Map of the Key Areas.

The research follows the distribution proposed by Webster & Watson [44] using a concept-centric approach. First, the analysis vectors for this research were identified and separated into main areas or themes, which represent the main areas in Figure 6.1. Following the identification of these areas, a thorough full-text analysis was conducted for each article in the dataset, categorising the content within the identified areas. Each category in the corresponding area includes the relevant articles mentioning that category.

This approach allowed for a systematic organisation of the information retrieved from the research, facilitating a more comprehensible presentation of the findings. To organise this information, tables were designed to display the distribution of the selected studies by category, along with the corresponding references. This tabulated information is displayed in the corresponding subchapters and, in certain cases, in the Appendix section.

The purpose of this chapter is to report the literature review findings by presenting a clear and organised overview of the current state of research on the use of genAI in education. This structured approach helps educational stakeholders better understand the potential applications, benefits, limitations, and future directions for genAI in educational settings, thereby contributing valuable insights to the academic community. The following subchapters will report each identified area and their corresponding categories.

6.1. Applications and Benefits of GenAI in Education

This section explores the diverse applications and benefits of integrating genAI tools into education. These benefits, derived from the analysis of selected studies, encompass a wide array of aspects, including providing continuous feedback to students, enhancing learning experiences, improving educational process efficiency, and positively impacting various educational stakeholders.

Through the literature analysis, a total of twenty-eight categories were identified under this section, reflecting the broad range of genAI's applications and the positive outcomes associated with its use in education. These categories have been grouped into seven main themes to enhance clarity and organisation. They aggregate related categories and provide a structured understanding of the areas most highlighted in the literature. The seven themes are:

- Automated Processes and Productivity
- Content Creation and Enhancement
- Learning Support and Enhancement
- Personalized Assistance and Feedback
- Language and Communication
- Student Engagement and Assessment
- Equity and Accessibility

The following subsections will explore each of these themes, with accompanying tables detailing the corresponding categories, the number of articles that mentioned each category, and the relevant references. These themes are not ranked by their significance or frequency of mentions but are intended to provide a comprehensive and organised view of the various applications and benefits genAI offers within educational contexts.

Subsequent subsections will present these themes along with the findings from the literature review. Each theme will provide a deeper exploration of the associated categories, illustrating the impact genAI has on enhancing teaching and learning in education.

6.1.1. Automated Processes and Productivity

This section reports on the applications and benefits associated with automating specific tasks through the integration of genAI and the resultant enhancement of productivity within the educational sector. Five categories were extracted from the literature review, as detailed in Table 6.1.

Table 6.1. Frequency of Mentions of Categories in Automated Processes and Productivity.

Category	Count	Article References
Productivity Enhancement	16	[10] [11] [23] [36] [45] [46] [57] [67] [68] [71] [73] [75] [78] [90] [98] [102]
Workload Reduction for Educators	15	[23] [56] [59] [61] [68] [73] [75] [76] [84] [87] [90] [95] [105] [107] [109]
Time Management for Students	15	[4] [13] [23] [45] [60] [95] [99] [106] [110] [112] [115] [123] [127] [132] [134]
Automated Grading	13	[56] [68] [73] [75] [76] [84] [99] [100] [109] [110] [115] [120] [125]
Automated Low-Cost/Zero-Cost Content Creation	2	[15] [57]

The most mentioned category in this theme was *Productivity Enhancement*, cited in sixteen articles. The literature indicates that integrating genAI tools will significantly increase productivity in educational settings [45]. Although the exact methods to achieve this boost are not always specified, several authors suggest that integrating these tools will significantly increase productivity [36] in education. For instance, Johinke et al. [98] highlight that AI can facilitate communication processes, contributing to overall productivity improvements.

The second category was *Workload Reduction for Educators*, cited in fifteen articles. Educators can leverage this technology to automate time-consuming tasks, such as grading students' assessments [56]. This automation allows them to save time previously spent on these tasks, dedicating more time to valuable activities such as providing support and assistance to students and preparing classwork [68], [75]. By reducing administrative burdens, educators can focus more on instructional quality and student engagement [73], while simultaneously reducing their workload [1].

The third category was *Time Management*, cited in fifteen articles. The rapid access to information provided by these models allows users to quickly obtain the information needed [99]. This approach conserves the time and energy that would have been expended by looking through various sources, such as Google [10]. The saved time can be reallocated to other tasks [95], which is particularly beneficial when students are approaching deadlines and experiencing anxiety [110]. These models assist not only students but also educators, enabling them to complete their tasks more efficiently [112].

The literature provides several examples of how genAI can be used for time management. For instance, genAI can aid in the research process, improving efficiency and making it easier to create a final draft [1]. Additionally, genAI can assist in coding project developments by generating code and solutions for specific coding problems, helping developers debug and find solutions more quickly [1].

The fourth one was *Automated Grading*, which was mentioned in thirteen articles. These tools can assist teachers by automating tasks like the grading of assignments, which significantly reduces their workload [68]. Moreover, automated grading can improve the consistency and fairness of assessments by eliminating human biases and errors. This ensures a more objective evaluation of student performance and saves time and resources [76], [110]. The use of AI in grading improves the assessment process and allows educators to allocate more time to providing feedback to students, enhancing the overall learning experience.

Finally, *Automated Low-Cost/Zero-Cost Content Creation* was mentioned in two articles. GenAI tools offer significant economic advantages in educational contexts by enabling the creation of content at low or zero cost [57]. Their integration does not require expensive technological infrastructure [15], which is particularly beneficial for institutions with limited budgets and for educators and students who lack access to costly technologies. This capability democratises access to high-quality educational resources and supports inclusive education initiatives.

The automation of tasks and the reduction in educators' workloads, as well as the time saved through this automation [73], opens new possibilities for educational dynamics, including the teacher-student relationship and the nature of teachers' work. This signifies an enhancement in productivity [11] when integrating these tools into education.

6.1.2. Content Creation and Enhancement

The second theme addresses the creation and enhancement of content using genAI tools. This section reports on the five major categories identified in the SLR, as demonstrated in Table 6.2.

Table 6.2. Frequency of Mentions of Categories for Content Creation and Enhancement.

Category	**Count**	**Article References**
High-Quality Text Generation	44	[2] [7] [12] [14] [18] [29] [36] [42] [43] [56] [57] [59] [62] [65] [66] [69] [70] [73] [75] [76] [77] [79] [81] [84] [85] [86] [88] [89] [91] [96] [98] [102] [106] [110] [112] [114] [117] [118] [124] [125] [127] [128] [130] [135]
Content Creation for Educators	29	[8] [10] [11] [14] [38] [42] [56] [59] [68] [73] [74] [81] [82] [84] [86] [87] [89] [90] [91] [94] [96] [97] [99] [104] [105] [109] [115] [116] [122]
Suggestion Provision	22	[2] [3] [8] [11] [10] [13] [14] [45] [48] [50] [54] [64] [71] [75] [85] [104] [106] [108] [110] [114] [115] [116]
Code Generation	18	[1] [3] [14] [29] [39] [56] [58] [62] [66] [67] [70] [74] [75] [76] [77] [80] [81] [115]
Brainstorming	11	[8] [11] [57] [60] [71] [76] [102] [112] [113] [128] [134]

The second category most frequently mentioned in the literature and the first in this section is *High-Quality Text Generation*, with forty-four mentions in total. The content generated by these tools is consistently described in the literature as high-quality and versatile [127]. Such tools create text [65] that is often indistinguishable from human-generated content [36], enabling them to serve as writing assistants [42]. This capability allows users to produce quality content [43], enhancing the writing process for various educational stakeholders.

For instance, faculty members can rely on genAI to generate content for multiple purposes [89], including drafting guidelines, social media posts, website content, event presentations, statements, and emails [81]. This versatility extends to creating polished and professional documents quickly [127], assisting in both administrative and academic tasks. These tools also support students in writing essays [43], [61], [66], [75], and engaging in creative writing [81] across multiple languages [77]. Furthermore, they can also be used to complete various assignments, from writing poetry and blog texts [89] to draft articles, reports, and code solutions [76].

The second category is *Content Creation for Educators,* which was mentioned in twenty-nine publications. According to the literature, educators can use genAI to reduce workload in numerous activities, including content generation [8], for both classroom and extracurricular use. The ability to generate educational content [116] opens up a new role in genAI for education, as these tools can act as teacher assistants [99], improving the educator's work by helping plan the classwork [59], organise course planning, prepare presentations [61], and innovate assessments to integrate into the classroom

[8], [14], [68], [84], [91], [96], [115]. This includes creating innovative quizzes and fostering new classroom debates. Educators can also use these tools to create textbooks [76].

The third category, *Suggestion Provision*, appeared twenty-two times in the literature. GenAI can be used to generate the final content, such as an essay or a lab report, and as a means to seek inspiration and guidance [43]. Based on the literature, most articles briefly mentioned this category [14], [85], [114], [115], [116], indicating that genAI provides suggestions. Specific examples include Siegle [14], who noted that the genAI tool Bing AI can be used to generate suggestions on any specific concern, and Cox and Tzoc [10], who observed that ChatGPT can serve as a research assistant by providing useful ideas for the investigation process.

Kovačević [54] concluded that students could benefit from the recommendations generated by ChatGPT. Additionally, Lyu et al. [108] concluded that the suggestions provided by these tools can enhance the educational process. Moreover, this technology can assist in the writing process by generating suggestions at any point [110], [114] and supporting educators in preparing teaching materials [104].

The fourth category is *Code Generation*, mentioned in eighteen articles. Some tools, like ChatGPT, are particularly effective at generating code for beginners [39]. A study by Rahman et al. [115] demonstrated the accuracy of code generated by this tool, reaching nearly eighty-six percent accuracy. GenAI tools can generate code and explain its functionality [14], enriching the learner's path by detailing how and why it generated the code that way and fostering critical thinking.

Aside from ChatGPT, other AI tools that are specifically code-focused, such as Codex [67], have demonstrated superior performance in certain exams compared to student-written code. Additionally, Copilot's answers can surpass those provided by students when they work independently [66], highlighting genAI's effectiveness in introductory programming education [115].

The fifth category is *Brainstorming*, mentioned in eleven articles. During brainstorming sessions, blockages may occur due to a lack of structured plans or overwhelming information. In these scenarios, these tools can assist in the brainstorming process [8], by supporting the creative process [57], [112], [128] and providing suggestions [3], enabling students to focus on their thought processes and organise their ideas more effectively.

6.1.3. Learning Support and Enhancement

The third theme corresponds to categories that fall into the support and enhancement of student learning. As seen in Table 6.3, five categories of applications and benefits associated with using genAI tools in learning and student development were identified.

Table 6.3. Frequency of Mentions of Categories for Learning Support and Enhancement.

Category	**Count**	**Article References**
Enhance Learning	47	[2] [5] [13] [15] [17] [20] [29] [38] [42] [51] [52] [55] [56] [59] [61] [62] [64] [65] [66] [68] [70] [73] [75] [76] [78] [82] [85] [86] [90] [94] [95] [96] [98] [100] [107] [108] [109] [111] [115] [116] [117] [118] [120] [122] [124] [132] [133]
Interactive Learning	17	[1] [13] [14] [21] [23] [29] [36] [50] [54] [56] [65] [84] [96] [99] [107] [109] [135]
Skill Development	11	[14] [21] [59] [62] [70] [96] [107] [115] [117] [120] [124]
Inspiration and Motivation to Learn	10	[11] [10] [14] [20] [43] [53] [52] [56] [59] [70]
Multimodal Understanding (Text & Images)	3	[29] [57] [132]

The most mentioned category in the literature was *Enhance Learning*. In a total of one hundred and twenty-one articles, this category appeared in forty-seven. As mentioned in section 6.1.2, genAI can support students by serving as a writing assistant, helping them express their thoughts and ideas more effectively, thereby improving their writing skills [51]. By providing continuous support and enhancement, students can focus on refining their work without the anxiety of making mistakes [85], allowing them to engage more deeply with their assignments.

These tools can also provide valuable information on various topics [86], supporting students' understanding and exploration of their subjects. For example, chatbots like ChatGPT can significantly enhance student learning [42], [52] in multiple areas, such as STEM education [55], medical education [69], and surgical education [111]. This suggests that genAI can play a crucial role in assisting students throughout their educational path by providing multiple opportunities to enhance learning [5].

The second category is *Interactive Learning*, which is mentioned seventeen times in the literature. The way students interact with AI through conversation can be very engaging depending on the style they instruct it to respond in [61]. Some more sophisticated tools can remember previous conversations and continue answering the user's questions with the previous context [65], making the flow of the interaction more human-like [107], thus enhancing the interactive experience. Additionally, some tools can adjust to the user’s age by styling the response accordingly [107].

When interacting with these technologies, students can ask the model to generate learning materials like quizzes and games [54]. It can also be used in group discussions [135] and to help young students develop their language and interaction skills [91]. This adaptability to generate content according to the learner's age, subject, and type of material makes the user experience interactive and personalised [13] when using genAI tools.

The third category, *Skill Development,* was mentioned eleven times. According to the review, it was noted that if the interactions with the genAI tools are properly made [96], they can aid in developing a range of skills useful in education. These tools can provide students with the possibility to develop the ability to learn independently by exploring different answers and problem-solving methods and thinking critically about the generated answers [62].

An example brought by Sallam et al. [21] in dental education, where ChatGPT helped develop skills due to the tools' ability to explain each step in detail and in an interactive manner. Siegle [14] stated that apart from ChatGPT, Bing AI also supports the development and improvement of multiple skills. For instance, Crawford et al. [70] concluded that this tool can improve students' computational skills, and Ivanov and Soliman [96] mentioned that by interacting with these technologies, students can develop AI-related skills and become more aware of AI. Overall, the literature mentions that these tools help develop reading, writing, language development, critical thinking, and problem-solving skills [59], [61], [70], [107], [115], [120], [124].

The fifth category *Inspiration and Motivation to Learn* was mentioned in ten publications. These tools can inspire students [61] and motivate them to learn new things or improve their understanding of current matters [14], [109]. A study by Yilmaz et al. [70] investigated the effects of using genAI in programming education and revealed that using AI as an assistant caused an improvement in students' motivation. Banić et al. [53] also studied the impact of ChatGPT in programming and concluded that it significantly helps improve the learner's motivation.

Aside from programming education, Pataranutaporn et al. [20] analysed having an AI virtual tutor that is similar to a person the student admires and found that it significantly impacts their motivation. Cox and Tzoc [10] provide an example of how it can improve motivation, which can be by generating a first draft on a certain topic so students can use it as inspiration. This improvement in motivation allows them to feel more interested in the subjects they are studying and be more motivated to continue learning.

Finally, the fifth category, *Multimodal Understanding (Text & Images)*, appeared three times in the review. Although some genAI tools are only text-based, others, like ChatGPT-4 [132] have a multimodal understanding, meaning they can generate both text and images. This capability allows them to process images and diagrams, thereby assisting students with their assignments beyond text limitations.

6.1.4. Personalized Assistance and Feedback

This section explores the four categories extracted from the literature, as seen in Table 6.4, which demonstrates how genAI's constant availability can benefit both educators and students and the purpose they can serve by assisting them in their daily tasks.

Table 6.4. Frequency of Mentions of Categories for Personalized Assistance and Feedback.

Category	Count	Article References
Personalised Feedback	38	[5] [8] [10] [11] [13] [14] [16] [21] [23] [50] [51] [52] [54] [56] [61] [68] [69] [70] [71] [73] [75] [76] [78] [84] [85] [90] [96] [97] [99] [104] [105] [109] [110] [117] [120] [126] [131] [135]
Virtual Tutor	37	[1] [3] [11] [13] [14] [16] [20] [23] [29] [39] [50] [54] [56] [61] [65] [70] [79] [83] [85] [89] [95] [96] [97] [99] [104] [105] [107] [109] [110] [112] [115] [116] [120] [123] [131] [134] [135]
Research Assistant	36	[8] [10] [11] [14] [19] [29] [43] [50] [56] [57] [59] [60] [69] [73] [75] [76] [81] [83] [86] [89] [93] [96] [98] [99] [102] [103] [106] [110] [113] [115] [116] [118] [119] [122] [128] [132]
Information Summarization	22	[6] [10] [62] [67] [69] [72] [75] [76] [84] [88] [89] [95] [96] [99] [102] [104] [106] [108] [113] [114] [128] [132]

The first category mentioned in this section is *Personalized Feedback*, with thirty-eight mentions in total. Innovations like ChatGPT have revolutionised how students access information [110]. These tools are always available to users, providing feedback on any topic whenever needed [85]. This is particularly beneficial for students who may require immediate feedback on their work when educator availability is limited. The literature often highlights the ability of genAI to provide personalised and immediate feedback [1], [8], [15], [61], [68], [70], [73], [84], [99], [104], [109], [117], [120], which Naidu and Sevnarayan [110] consider essential for learning.

The feedback from these tools can be used to improve the quality of work produced [85] across various educational areas. It can enhance the overall writing quality of an assignment [75], [105], whether by providing personalised feedback on a document [71] or by improving the grammar, as explained by Hwang and Nurtantyana [51]. Additionally, Sallam et al. [21] mentioned that students can benefit from genAI's ability to provide instant feedback on the techniques they are utilising to produce their work. This availability for feedback and the resultant improvements can significantly impact the students' work quality and overall learning experience [105].

The second category is *Virtual Tutor*, which has been mentioned in thirty-seven publications. Some authors in the literature list the benefits of using genAI tools as virtual tutors, such as Su and Yang [13], who cite immediate feedback and tailored experience for students, and how it can assist them in managing their time. Qadir [11] notes that the ability to ask questions and receive answers is a significant application of genAI tutoring, and Yilmaz and Karaoglan [70] concluded that these tutors can positively impact the learner's grades and motivations towards their subjects. Therefore, research findings demonstrate that genAI can serve effectively as a virtual tutor [23].

Educators can also leverage this technology by assigning various roles to these tools [97]. GenAI can act as a content creator, grading assistant, or administrative assistant. This constant support and availability make these tools effective virtual tutors [14]. They can help personalise students' learning paths [15] and provide tailored learning experiences [110] that meet individual needs. This personalised approach can assist students by answering questions, summarising information [67], preparing them for assessments [104], and providing personalised feedback with human-like responses [135], enhancing the interactive learning experience discussed in 6.1.3.

The third category is *Research Assistant*, which is mentioned in thirty-six articles. According to the literature, AI can function as a research assistant [75], [99], [118], aiding in the overall process of research and investigation. It can help retrieve and analyse information by processing substantial amounts of data and transforming it into valuable insights that form the foundation of research [83]. This ability to process vast amounts of text [43] coupled with summarising extensive information [99], [102], significantly reduces the time required for research tasks [75], [132], allowing researchers to concentrate on other critical aspects of their work.

Additionally, in an investigation of the role of genAI tools in academic research, Lund et al. [106] provided examples of how AI can assist researchers, from language assistance to formatting references. Therefore, genAI can support multiple stages of the research process [69], from data collection and analysis to writing and synthesising [119].

The fourth category mentioned in this section is *Information Summarization*, with twenty-two mentions in the literature. Research findings indicate that most mentions cite this category briefly, concluding that genAI can assist in summarising information [6], [62], [67], [75], [76], [84], [96], [99], [102], [104], [106], [108], [113], [114], [128], [132]. Crawford et al. [69] and Rudolph et al. [72], noted that it can summarise documents, a skill necessary for the learning process, as described by Emenike et al. [95].

When used transparently and ethically, genAI tools can provide numerous benefits for students, educators, and researchers [73]. These tools provide personalised assistance and feedback, significantly enhancing teaching and learning experiences. With their continuous support, availability, and tailored interactions, they deliver immediate feedback and research assistance.

6.1.5. Language and Communication

This section encompasses the categories that help and transform the overall communication process. In this research, three categories were extracted from the literature, as demonstrated in Table 6.5.

Table 6.5. Frequency of Mentions of Categories for Language and Communication.

Category	Count	Article References
Multilingual Communication	25	[4] [29] [43] [50] [55] [56] [62] [70] [71] [75] [77] [81] [84] [88] [99] [104] [106] [108] [114] [115] [116] [117] [120] [122] [128]
Accurate Definitions	15	[6] [7] [19] [21] [29] [57] [62] [69] [75] [86] [89] [90] [92] [96] [100]
Reduction of Language Anxiety	2	[52] [56]

The first category is *Multilingual Communication,* as mentioned in twenty-five publications. The literature reveals that genAI is proficient in supporting multilingual communication [62], [88]. Some publications mentioned the ability of these tools to translate texts into multiple languages [50], [55], [62], [71], [115], [116], [128], providing advanced support in those languages [56], [70]. Thus, making them valuable in educational settings, particularly for non-native speakers [114]. Dwivedi et al. [77] provide examples of document types genAI can write in different languages, including poetry and essays. These tools are particularly beneficial in assisting non-native speakers by reducing language barriers [29], [43], [62], [71], [75], [84], [88], [99], [104], [106], [108], [114], [115], [116], [117], [120], [122], [128].

In addition to translations, these tools can make text more understandable for their audience. For instance, Peres et al. [4] highlighted the benefit of transforming academic style into a more conversational tone, and Lyu et al. [108] discussed translating medical terms into a similar style. By utilising these capabilities, genAI can produce grammatically correct content in several languages and styles [55].

The second category presented in this theme is *Accurate Definitions*, appearing fifteen times in publications. According to the literature, GenAI tools are capable of providing precise and accurate definitions [6], [75], [89], [100], which is particularly useful in educational contexts. In medical education, Xie et al. [19], stated that ChatGPT provided correct definitions in a specific area of study, while Sallam et al. [21] discussed how these tools can understand and explain concepts. Das et al. [86] found that ChatGPT can answer microbiology questions, providing accurate definitions to students. Similarly, genAI has been shown to offer accurate clinical insights in its responses [92]. This capability demonstrates how this technology supports students in their learning processes.

This section's third and final category is *Reduction of Language Anxiety*, mentioned in two publications. GenAI tools can significantly reduce language anxiety among students [52]. The access to extensive educational resources that AI provides allows these tools to increase equality of opportunity in education, as they can offer more inclusive materials, personalised support, and interactive communication, as stated by Bahrini et al. [56].

This technology provides a safe and non-judgmental platform for practising language skills, helping students gain confidence in their abilities. It allows them to practice speaking, writing, and comprehension without fear of making mistakes in front of peers or teachers, ultimately improving their overall language proficiency.

6.1.6. Student Engagement and Assessment

This theme highlights the ways genAI tools enhance student engagement and support various assessment processes. Four categories were extracted from the literature, as seen in Table 6.6.

Table 6.6. Frequency of Mentions of Categories for Student Engagement and Assessment.

Category	Count	Article References
Assessment Support	20	[39] [42] [56] [61] [76] [77] [83] [86] [88] [91] [92] [99] [100] [104] [105] [110] [122] [125] [132] [135]
Exam Passing Support	12	[16] [29] [47] [48] [58] [66] [72] [73] [75] [100] [101] [117]
Outperforming Student Responses	2	[18] [66]
Comfortable Asking AI Instead of Teachers	1	[55]

The first category mentioned in this theme is *Assessment Support*, appearing in twenty publications. The research indicates that genAI tools can provide substantial support for several types of assessments [56], helping students improve their studying and learning abilities. According to Lo [104], Lodge et al. [105] and Sevgi et al. [122], these tools can generate content to help students practice for their assessments, grade students' work, and provide feedback to enhance their learning. Additionally, Tsang [132] noted that genAI can explain definitions and concepts.

The type of assignments generated to assist students in their learning process can include multiple-choice, figure-based and essay questions [77]. Furthermore, these tools can create scenarios for students to practice with real examples. In medical education, genAI can enhance students' diagnostic and treatment-planning abilities by providing realistic case scenarios for practice [99]. Consequently, as concluded by Yan [135], reliance on these technologies to support studying and exam preparation has demonstrated improvements in the quality of student work.

The second category presented in this theme is *Exam Passing Support*, which is mentioned in twelve articles. According to the literature, genAI tools have demonstrated their ability to pass exams and answer assignments correctly [48]. For instance, ChatGPT has been shown to pass MBA exams

[75], perform comparably to students on physics exams [100], achieve remarkable results in multiple-choice exams [101], and pass exams in the law and business fields [29]. These tools can also assist with introductory exam answers [16], [47], [48], [117], offering significant support in exam preparation.

Additionally, According to Limna et al. [73], the premium version of ChatGPT has performed well in graduate-level exams, demonstrating a passing grade. This showcases the technology's ability to not only provide correct answers to assessments but also support students in preparing for them [75], as discussed in the previous category.

The third category *Outperforming Student Responses* was mentioned twice in the literature. According to Li et al. [18], these tools can generate responses that often outperform those of students—an example provided by Finnie-Ansley et al. [66] concluded that genAI tools like Codex can generate better responses than students, and tools like Copilot have demonstrated the ability to create superior solutions compared to those produced by students working independently.

Finally, the fourth category, *Comfortable Asking AI Instead of Teachers*, appeared in one publication. As previously discussed in section 6.1.5, this technology provides a non-judgemental tool for students, allowing them to feel confident and comfortable making questions without expecting judgment in return [55]. This environment encourages frequent interactions and fosters a deeper understanding of subjects where students have doubts.

6.1.7. Equity and Accessibility

This section refers to the final theme identified for the applications and benefits of using genAI in education. It introduces the role of genAI in promoting equity and accessibility in education and consists of two categories, as mentioned in Table 6.7.

Table 6.7. Frequency of Mentions of Categories for Equity and Accessibility.

Category	Count	Article References
Equity and Access to Education	21	[2] [4] [15] [29] [52] [56] [66] [68] [81] [83] [90] [96] [108] [110] [111] [115] [116] [122] [123] [127] [132]
Learning Disabilities Support	6	[29] [56] [68] [81] [83] [115]

The most frequently mentioned category in this theme is *Equity and Access to Education*, with twenty-one mentions in the literature. The constant availability of genAI tools ensures that learning resources are also easily accessible [52]. GenAI offers easy and fast access to vast amounts of information [90], [110], providing personalised feedback and learning experiences adapted to individual needs. This adaptability allows genAI to support different literacy levels within a classroom, helping students with lower levels improve their skills [15].

Beyond language literacy, genAI can also support and better integrate students with communication disabilities by providing tools to improve their communication skills [81]. Additionally, the widespread availability of genAI helps students, educators, and faculties [96], as well as researchers [4] across the academic spectrum. For instance, a study conducted by Finnie-Ansley et al. [66] exploring the impact of genAI coding tools in informatics education reported that tools like Copilot, freely accessible to students, generate better solutions to beginner-level coding problems than unassisted efforts. The literature indicates that the accessibility these tools provide results in improved access to education [108], [123], [127], leading to more significant equity throughout educational institutions [2].

The category *Learning Disabilities Support* was mentioned in six publications. This technology offers tailored support for students with learning disabilities, enabling them to engage with educational content more effectively [68]. These tools can support various areas, such as language and social skills, and generate multiple types of learning materials tailored to the learner's level and ability, thereby improving inclusivity among students [56].

For instance, ChatGPT offers functionalities such as transforming speech to text or vice versa [115], making learning more accessible and adapted to individual needs. As stated by Wagholikar et al. [29], this tool benefits users with disabilities by allowing them to use the voice option to record their thoughts and translate them into text, enhancing accessibility. Additionally, Corsello and Santangelo [83], noted that LLMs are capable of detecting developmental delays in young students through speech and language analysis, enabling early intervention and enhancing their development outcomes. Therefore, genAI can ensure that all students have the opportunity to succeed regardless of their challenges [81].

6.2. Limitations, Risks, and Concerns of GenAI in Education

While genAI offers numerous advantages, as outlined in the previous subchapter, it also presents several limitations and risks that must be addressed. According to the SLR, twenty categories have been identified that discuss potential challenges such as ethical considerations, privacy issues, bias, plagiarism, over-reliance on genAI tools, and concerns regarding the reliability and accuracy of generated content. These categories will be analysed and reported in the following sections.

The categories have been grouped into six main themes to enhance clarity and structure; each aggregating related challenges identified in the literature. These themes are:

- Academic Integrity and Ethical Challenges
- Accuracy, Bias, and Plagiarism Issues
- Authorship and Content Attribution
- Security and Privacy Risks
- Technical, Usability, and Interaction Challenges
- Cognitive and Educational Impact

The following subsections will explore each one, with accompanying tables detailing the corresponding categories, the number of articles that mention each category, and the relevant references. Their significance or frequency of mentions does not rank these themes. It is intended to provide a comprehensive and organised view of the various limitations and concerns that arise with the emergence of genAI in education.

The following subsections will present these themes along with the findings from the literature review, offering a deeper exploration of the challenges and risks posed by genAI integration in educational contexts.

6.2.1. Academic Integrity and Ethical Challenges

This section reports on the academic integrity and ethical challenges faced when integrating genAI into education. Three categories were extracted from the literature that fit within this theme, as seen in Table 6.8.

Table 6.8. Frequency of Mentions of Categories for Academic Integrity and Ethical Challenges.

Category	Count	Article References
Academic Integrity	49	[2] [6] [8] [11] [13] [15] [17] [19] [21] [14] [16] [36] [48] [50] [55] [56] [57] [58] [59] [60] [65] [69] [70] [72] [75] [76] [77] [80] [81] [84] [85] [90] [98] [99] [104] [105] [106] [109] [112] [115] [118] [119] [121] [122] [128] [130] [131] [132] [135]
Ethical Concerns	38	[5] [8] [10] [13] [16] [29] [42] [46] [52] [55] [56] [57] [68] [71] [76] [82] [83] [94] [95] [96] [106] [107] [108] [109] [110] [112] [113] [114] [116] [118] [122] [125] [127] [128] [130] [131] [132] [134]
Misuse	11	[16] [17] [18] [42] [56] [65] [97] [101] [114] [122] [132]

The first category, *Academic Integrity*, is the third most frequently mentioned concern in the literature regarding the limitations, risks, and challenges of integrating genAI in education, with fortynine references. Some authors mention academic integrity as a concern [8], [59], [60], [69], [75], [76], [77], [81], [90], [105], [112], [132], without providing detailed examples of how it might be compromised. However, approximately sixty-one percent of the articles in this category specifically highlight plagiarism as a significant threat to academic integrity [2], [6], [11], [13], [14], [15], [21], [48], [50], [55], [56], [57], [65], [72], [75], [76], [80], [84], [90], [98], [99], [104], [106], [112], [118], [119], [121], [122], [130], [135].

Beyond plagiarism, other concerns regarding the use of genAI in education were identified. Susnjak [36] points out that students can use AI-generated information to cheat on evaluations and assignments. Rahman et al. [115] further emphasise that cheating is exacerbated in online environments, as students have more access to genAI tools, and Nikolic et al. [39] conclude that online assessments pose a significant risk to academic integrity. Overall, approximately thirty-nine percent of the articles within this category discuss cheating and academic dishonesty as critical issues associated with the use of genAI in education [16], [21], [36], [48], [57], [58], [65], [70], [75], [81], [85], [99], [109], [110], [115], [128], [131], [135]. These concerns highlight the need for robust strategies and policies to maintain academic integrity in the face of increasing genAI integration in educational settings.

The second category, *Ethical Concerns*, was reported thirty-eight times in the literature. Approximately seventy-six percent of the articles within this category mentioned that using genAI tools in education comes with ethical concerns [5], [10], [13], [29], [46], [52], [71], [76], [82], [83], [94], [95], [96], [106], [107], [108], [109], [112], [114], [116], [118], [123], [125], [127], [128], [130], [131], [132], [134], while the remaining articles provide specific examples of these concerns. From an ethical standpoint, Rasul et al. [8] emphasise that educational stakeholders must use these tools appropriately for educational purposes. For instance, Giannos and Delardas [16] argue that using genAI tools to complete admission assignments can create inequality in the admission process, aggravating educational disparities, as noted by Adiguzel et al. [68].

Another prominent ethical concern is the potential for plagiarism. Elder et al. [55] and Bahrini et al. [56] point out that these tools can facilitate plagiarism. This represents a significant issue given that traditional cheating detection tools may struggle to identify AI-generated content, as highlighted by Wu et al. [57]. Beyond plagiarism, Sevgi et al. [122] report that using genAI tools to generate and disseminate false information is a critical ethical concern. Strzelecki [42] further underscores the importance of the accuracy of information generated by these tools, as inaccuracies can have widespread implications.

The third category in this section, *Misuse*, was mentioned eleven times in the literature. This category refers to how individuals can dishonestly interact with genAI, such as creating and disseminating false information using these tools, as stated by Bahrini et al. [56]. According to Li et al.

[18], it is crucial to be aware of the misuse of genAI in education to differentiate between students' and AI-generated content.

Sullivan et al. [17], in their analysis of the disruption provoked by ChatGPT in universities, reported the categories frequently mentioned in their study, aligning with the ones in this section. This match demonstrates the importance of addressing misuse to maintain the integrity and reliability of educational content and processes, as well as addressing academic integrity and ethical concerns.

6.2.2. Accuracy and Bias

This section reports on the relevancy of the information generated by genAI tools and its accuracy. To do so, three categories were identified in the literature and will be reported, as seen in Table 6.9.

Table 6.9. Frequency of Mentions of Categories for Accuracy and Bias.

Category	Count	Article References
Bias and Falsified Information	57	[2] [8] [9] [10] [11] [13] [14] [20] [21] [22] [35] [39] [43] [49] [50] [52] [56] [57] [58] [59] [60] [61] [64] [65] [68] [69] [71] [72] [73] [75] [76] [77] [78] [80] [83] [88] [89] [90] [94] [95] [96] [103] [104] [105] [106] [107] [108] [110] [116] [118] [119] [123] [128] [129] [130] [133] [134]
Accuracy and Reliability	51	[3] [6] [8] [11] [13] [16] [23] [35] [39] [42] [47] [49] [52] [56] [57] [60] [61] [62] [63] [64] [66] [68] [69] [73] [77] [87] [88] [89] [90] [92] [93] [94] [98] [102] [104] [105] [108] [109] [111] [117] [118] [120] [122] [124] [125] [126] [127] [128] [130] [131] [133]
Fake References	20	[1] [11] [19] [23] [42] [43] [49] [65] [69] [77] [87] [93] [96] [104] [109] [112] [118] [124] [125] [128]

The first category in this section, *Bias and Falsified Information*, represents the most discussed limitation in the literature, with fifty-seven publications mentioning this concern. According to Siegle [14], the content that genAI models are trained on predominantly comes from the internet, implying that the quality of data generated by these models is contingent upon the quality of internet data. As Han and Cai [9] point out, the Internet contains biased and false information, leading scholars to conclude that these models can generate outputs containing prejudiced and biased content.

Therefore, there is a significant risk of these models producing false information [76]. For example, Malinka et al. [65] provide instances where ChatGPT generates fake information, such as incorrect links or citations. This aligns with Bekeš et al. [52], who highlight the potential of these tools to disseminate false information. The quality of these models is a crucial consideration due to the inherent risks associated with their use [89], [90].

The potential bias in the AI model training process can influence and impact equality in education. According to Alasadi et al. [22], this bias can worsen societal disparities, with one segment of society advancing more rapidly than others. In the educational context, the quality of data used to train these models can significantly affect the educational materials they generate, often leading to falsified and biased information, as noted by Lodge et al. [105]. Consequently, poor-quality data can cause these models to produce incorrect educational materials, a concern stated by Naidu and Sevnarayan [110].

The second category, *Accuracy and Reliability*, with fifty-one mentions in publications, represents a significant concern regarding the risks and limitations of integrating genAI into education. This category was extracted from the SLR and highlights the potential inaccuracies and unreliability of content generated by genAI models. Some authors explicitly discuss accuracy and reliability [42], [61], [68], [69], [87], [89], [92], [94], [102], [104], [109], [120], [122], [127], [128], [131], [133], while others refer to these inaccuracies as hallucinations [6], [11], [23], [35], [56], [57], [60], [64], [69], [73], [77], [118], [124].

According to Qadir [11], this lack of accuracy can be attributed to the quality of the data with which the models are trained. As mentioned in the previous category, the data sourced from the internet often contains biased and false information. Therefore, the trust users place in the content generated by these models must be tempered with critical analysis skills. Students must be trained to evaluate the information critically and discard inaccuracies, as emphasised by Joyner [63] and Nikolic et al. [39].

Bekeš et al. [52] analysed the accuracy of genAI responses on historical topics, finding several inaccuracies. This underscores the need for examination by students when using genAI tools. Additionally, Oh et al. [111] compared the performance between GPT-3.5 and GPT-4, noting that the newer model exhibits a higher level of accuracy, particularly in comprehending medical-related questions.

Johinke et al. [98] introduce the concept of explainable AI, highlighting the challenge that users often do not understand how AI models arrive at their generated content or the sources of their information, complicating the assessment of the quality of AI-generated content [90]. Moreover, Lyu et al. [108] note that the variability of responses from these models, even when using the same prompt, adds to the unpredictability and potential unreliability of the answers provided.

The third category, *Fake References*, was mentioned in twenty publications. According to multiple articles, genAI tools occasionally provide fake references [23], [39], [65], [69], [77], [87], [104], [109], [112], [118], [124], [125], [128]. For instance, Xie et al. [19] concluded in their evaluation study of

ChatGPT responses that the references it generated were fake and did not match any actual publications. Furthermore, Ivanov and Soliman [96] highlighted a significant limitation of these models as they often fail to cite the texts they use to generate answers.

6.2.3. Authorship and Content Attribution

This section introduces concerns from the literature related to authorship problems and the efficiency of plagiarism detector tools in detecting AI-generated content in students' assignments. Three categories were extracted, as demonstrated in Table 6.10.

Table 6.10. Frequency of Mentions of Categories for Authorship and Content Attribution.

Category	Count	Article References
Authorship	21	[4] [46] [55] [59] [67] [75] [77] [81] [82] [87] [89] [96] [103] [106] [113] [114] [121] [122] [127] [129] [134]
Distinguish AI from Human Writing	13	[18] [36] [43] [48] [57] [59] [79] [84] [88] [112] [124] [130] [135]
Plagiarism Tools Not Reliable	6	[18] [48] [57] [75] [77] [88]

The first category, *Authorship*, was mentioned twenty-one times in the literature. This category addresses issues surrounding the attribution of content generated by genAI in the educational field, as these tools are not recognised as authors. According to an investigation by Lee [103] on whether ChatGPT can be cited as an author, the conclusion is that it cannot, which raises multiple concerns within the academic community. Konecki et al. [67] further highlight the problem for institutions, noting that it is challenging to verify whether a work was exclusively done by students or if genAI was also used. If this last case is valid, students cannot acknowledge genAI as a co-author as these tools are not considered authors.

Overall, the SLR reports that some publications mention the category briefly [59], [75], [81], [89], [96], [103], [106], [113], [121], [122], [127], [129], [134]. The inability to attribute authorship to genAI tools presents significant challenges in academic contexts, including accountability, originality, and ethical considerations.

The second category, *Distinguishing AI from Human Writing*, was mentioned in thirteen publications. This limitation relates to the challenge of differentiating text written by students from that generated by genAI due to the technology's capability to produce human-like text, as noted by Shoufan [59]. Other authors in the literature echo this concern. For instance, in an assessment of ChatGPT conducted by Breeding et al. [79], it was found that distinguishing AI-generated content from human-produced content is remarkably difficult. Similarly, Cotton et al. [84] concluded that the text generated by genAI is challenging to differentiate from human writing.

Further exploration by Perkins [112] and Shaw et al. [124] revealed that it remains unclear whether humans or AI generated specific texts. This ambiguity presents a significant limitation, as it complicates the verification process in academic contexts. Thurzo et al. [130] also concluded that it is difficult to determine whether AI or students authored the text.

Finally, the third category, *Plagiarism Tools Not Reliable*, was mentioned in six publications. This category highlights a critical concern faced by educational institutions, which lack effective tools to detect AI-generated content with certainty [75]. Li et al. [18] provided an example where the AI plagiarism detector used in their study failed to distinguish between AI-generated content and content produced by students. This aligns with the conclusions drawn by Geerling et al. [48], who noted the difficulty in identifying when assignments and assessments have been completed with the assistance of AI.

The challenge of detecting AI-generated content is further amplified by the advanced capabilities of genAI tools, making plagiarism detection more difficult [57]. Since these tools can produce content that is undetectable by conventional plagiarism detectors [77], institutions must reevaluate their approaches and tools to address this issue effectively [88].

6.2.4. Security and Privacy Risks

This section highlights concerns regarding the safety of the student's private data and the data given to genAI models, mainly if it contains personal information. To address these issues, two categories were extracted from the literature, as detailed in Table 6.11.

Table 6.11. Frequency of Mentions of Categories for Security and Privacy Risks.

Category	**Count**	**Article References**
Privacy and Safety	21	[9] [13] [14] [21] [56] [59] [68] [75] [78] [83] [88] [94] [95] [107] [109] [110] [118] [123] [130] [131] [132]
Cybersecurity	9	[14] [29] [56] [59] [73] [75] [83] [110] [118]

The first category reported in this section is *Privacy and Safety*, with twenty-one publications mentioning this concern. According to the literature, most authors briefly address this category without providing detailed examples of how privacy and safety issues may affect students or how students might inadvertently contribute to these issues. They generally refer to this concern as an invasion of student privacy [9], [13], [21], [14], [56], [59], [75], [83], [88], [94], [95], [107], [109], [110], [118], [123], [130], [131], [132].

However, Adiguzel et al. [68] highlight a specific concern where genAI models collect the data inputted by users. This emphasises the importance of students being aware of the data they share with these models, as it can compromise their privacy and the privacy of other educational stakeholders.

The second category, *Cybersecurity*, was mentioned in nine articles. According to the literature, most articles briefly mention that integrating genAI tools in education can pose security risks [56], [29], [59], [73], [75], [83], [110], [118]. However, in his exploration of AI in gifted education, Siegle [14] concluded that there are significant cybersecurity risks in addition to safety and privacy concerns associated with students' usage of genAI. These risks include the potential for these tools to be misused for cyberbullying.

6.2.5. Technical, Usability, and Interaction Challenges

This section addresses the technical, usability, and interaction challenges of integrating genAI into educational settings. Three categories were extracted from the literature analysis, as seen in Table 6.12, highlighting the complexities and limitations that educators and students may encounter when incorporating genAI into the learning environment.

Table 6.12. Frequency of Mentions of Categories for Technical and Interaction Challenges.

Category	**Count**	**Article References**
Inefficient Communication	8	[3] [6] [13] [29] [52] [56] [59] [99]
Lack of Human Interaction	7	[56] [68] [73] [115] [120] [123] [134]
Technical Expertise Barriers	2	[54] [59]

The first category, *Inefficient Communication*, was mentioned in eight articles. According to the SLR, some authors reported that genAI tools can be inefficient at communicating. Su and Yang [13], Cascella et al. [6], and Dakhel et al. [3] noted that these tools often struggle with complex instructions, leading to ineffective communication. Bahrini et al. [56] and Khan et al. [99] further pointed out that these models can have limitations in understanding context, resulting in communication issues where the tools do not behave as the user intends.

Additionally, the analysis revealed that genAI tools could struggle with comprehending conversational language across different cultures, as noted by Shoufan [59]. They can also be less accurate for non-European languages, according to Wagholikar et al. [29]. This cultural discrepancy can lead to misunderstandings and irrelevant content generation, as highlighted by Bekeš et al. [52]. These communication inefficiencies emphasise the limitations of genAI in effectively responding to user inputs.

The second category, *Lack of Human Interaction*, was mentioned in seven publications. This category addresses the reduced amount of human interaction when students use genAI tools, as these do not require a human presence, such as a teacher, to assist. Limna et al. [73] highlight that this reduction in teacher-student interaction time is a significant concern when using these tools.

Furthermore, the lack of human interaction is particularly problematic for subjects that require hands-on experience and in-person engagement. Vartiainen and Tedre [134] discuss this issue in their exploration of genAI in craft education, noting that the experience traditionally gained in a classroom setting is transformed into an interaction with a digital system, resulting in a loss of tactile and personal engagement.

The third category, *Technical Expertise Barriers*, was mentioned in two publications. This category addresses the knowledge and expertise required to understand and integrate genAI tools into education. According to Kovačević [54], integrating these tools necessitates a certain level of computer science knowledge to comprehend the tool's functionalities and customise the experience effectively. This finding aligns with Shoufan [59], who concluded that a significant limitation is teachers' lack of technical knowledge, which interferes with their ability to interact with these tools in their professional roles.

6.2.6. Cognitive and Educational Impact

This section addresses the final theme identified concerning the limitations, risks and concerns associated with using genAI in education. It explores the potential deterioration of essential educational skills due to over-reliance on genAI, as well as significant impacts, as outlined in Table 6.13.

Table 6.13. Frequency of Mentions of Categories for Cognitive and Educational Impact.

Category	Count	Article References
Dependence on GenAI (over-reliance)	23	[14] [39] [42] [43] [56] [59] [61] [68] [81] [83] [90] [96] [97] [106] [109] [110] [115] [116] [117] [118] [120] [132] [135]
Deterioration of Cognitive and Communication Skills	14	[18] [21] [55] [56] [58] [65] [67] [75] [81] [90] [97] [110] [115] [117]
Educational Equity	8	[11] [13] [68] [84] [89] [95] [117] [135]
Educational Assessments	7	[1] [5] [12] [23] [29] [38] [110]
Teachers Replacement	3	[20] [95] [117]
Performance in School Subjects	2	[100] [104]

The first category, *Dependence on GenAI*, was mentioned in twenty-three publications. This category addresses the potential risks associated with educators and students becoming overly dependent on these tools and the implications of such reliance. Many publications raise concerns about this issue [56], [61], [68], [81], [83], [97], [106], [109], [110], [115], [116], [117], [118], [120], [132], [135].

For instance, Shoufan [59] specifically highlights concerns about the overuse of ChatGPT, while Dergaa et al. [43] express similar worries for both students and researchers. Farrokhnia et al. [90], in their exploration of ChatGPT's impact on education, concluded that this over-reliance could have negative consequences. Strzelecki [42], in a study on the usage of ChatGPT in education, provides an example of how students' over-reliance on GenAI can lead to them neglecting to verify the sources of AI-generated content, ultimately limiting their understanding.

The second category, *Deterioration of Cognitive and Communication Skills*, was mentioned in fourteen articles. This category represents an indirect consequence of students' dependence on genAI models. Authors such as Sallam et al. [21] and Li et al. [18], concluded that over-reliance on these models can negatively impact students' ability to think critically and analyse information. Furthermore, the interaction with these models, primarily through computers or smartphones, can reduce the time students spend communicating with their peers and educators, further aggravating the issue.

The literature highlights concern regarding the deterioration of critical thinking and communication skills [18], [21], [58], [65], [75], [81], [90], [97], [115], [117]. In terms of cognitive skills, Elder et al. [55] and Dwivedi et al. [75] mention the potential decline in writing skills essential for academic success. Dwivedi et al. [75] also emphasise that over-reliance on genAI can lead to a lack of problem-solving skills and creativity, which are crucial for students' overall cognitive development. This concern is mentioned by other authors who note the potential diminishment of problem-solving abilities and creativity as a consequence of excessive reliance on these tools [75], [110], [115].

The third category presented in this section is *Educational Equity*, mentioned in eight publications. The literature analysis reveals concerns about how the integration of genAI tools in education could impact educational equity [89], [135]. According to Su and Yang [13], institutions in economically disadvantaged areas may face significant challenges in adopting genAI, as these tools require investments in technological infrastructure, such as computers and reliable internet access.

Adiguzel et al. [68] further emphasise that the emergence of genAI tools could bring more educational inequities, as some regions or institutions may have immediate access to these technologies while others may struggle to keep pace. In a study on engineering education, Qadir [11] concluded that the field needs to be updated with new technology, underscoring the importance of ensuring that all institutions have equal access to these tools.

The fourth category, *Educational Assessments*, was mentioned in seven articles. This category pertains to the threats that genAI poses to the integrity and effectiveness of educational assessments. For instance, Eager and Brunton [38], in their study on the challenges of AI in higher education, concluded that traditional assessments commonly used in schools could be negatively impacted by genAI, as these tools can provide correct answers. Dwivedi et al. [75] offer a similar concern, providing an example of how genAI can be utilised to complete assessments and even draft an academic thesis.

In addition to the challenges facing traditional assessments, Naidu and Sevnarayan [110] expands the discussion to include the potential disruption of online classes and assignments caused by genAI. This concern is reinforced by Daun and Brings [1] and Wagholikar et al. [29], who highlight the risks that genAI poses to the integrity and reliability of online educational activities.

The fifth category, *Teachers Replacement*, was mentioned in three publications. This category reflects concerns in the literature regarding the potential for AI to replace teachers in educational settings, as well as the consequent reduction in interactions between students and their educators. Pataranutaporn et al. [20], in their exploration of using virtual characters as teachers, concluded that such practices could disrupt students' educational experiences. Similarly, Sánchez-Ruiz et al. [117] found that students' over-reliance on genAI could lead to a significant decrease in interactions with their educators. In terms of job security, Iskender [95] expressed concern about teachers losing their jobs as AI becomes more capable of fulfilling educational roles.

Finally, the last category in this section, *Performance in School Subjects*, was mentioned in two publications. This category highlights the inadequate performance of genAI models in specific academic subjects, as analysed in the literature. Kortemeyer [100] evaluated whether ChatGPT could pass a physics course and found that the model produced poor answers, mainly when dealing with mathematical problems. Similarly, Lo [104] concluded that the performance of the same model was underwhelming in fields such as law and medicine.

6.3. Future Suggestions for GenAI in Education

This subchapter compiles future suggestions from the literature on enhancing the effectiveness and acceptance of genAI tools in educational settings. Key recommendations include the development of innovative and updated assessments and assignments, the establishment of clear guidelines for the ethical and effective use of genAI, increased research into the impact of genAI on education, the development of more genAI-powered educational tools, and strategies to encourage the development of critical thinking among students.

These suggestions were categorised into seven categories and grouped into two main themes, which will be discussed in the following sections. These themes are:

- Curriculum and Educational Practices
- Interaction, Training, and Ethical Considerations

These two themes will be explored with accompanying tables that detail the associated categories, the number of articles that mentioned each category, and the relevant references. Their significance or frequency of mentions do not rank them. The tables are intended to provide a comprehensive and organised view of future suggestions regarding genAI in education. The following subsections will present these themes along with the findings from the literature review, offering a deeper exploration of the recommendations proposed by the literature.

6.3.1. Curriculum and Educational Practices

This section explores future suggestions for integrating genAI into education, particularly concerning the evolution of curricula and educational practices. Based on the literature, three main categories were identified, as summarised in Table 6.14.

Table 6.14. Frequency of Mentions of Categories for Curriculum and Educational Practices

Category	Count	Article References
AI Integration in Teaching and Curriculum	34	[38] [39] [42] [53] [58] [59] [61] [65] [67] [72] [76] [77] [78] [82] [96] [97] [107] [110] [111] [115] [116] [117] [118] [120] [122] [123] [124] [128] [130] [131] [132] [133] [134] [135]
Innovative Assessment Design and Curricula Adaptation	21	[8] [12] [11] [17] [29] [36] [39] [48] [66] [72] [77] [81] [84] [88] [90] [96] [100] [104] [105] [107] [110]
Development and Research of AI in Education	10	[4] [5] [7] [13] [53] [57] [78] [105] [109] [110]

The first category discussed in this section is *AI Integration in Teaching and Curriculum*, mentioned in thirty-four publications. This category encompasses various suggestions authors make across a wide range of articles. Approximately seventy-one percent of these articles emphasise the necessity of integrating genAI into the classroom setting [39], [42], [58], [59], [61], [72], [76], [96], [97], [110], [111], [115], [116], [117], [118], [120], [122], [123], [124], [131], [132], [133], [134]. For instance, Oh et al. [111], in their evaluation of the performance of OpenAI's GPT-4 in medical education, highlight the importance of integrating AI to support users and enhance the current educational curriculum. Similarly, Naidu and Sevnarayan [110] argue that institutions must embrace these modern technologies as part of their curricula.

In addition to calls for integration into the curriculum, the literature also suggests that genAI tools could serve as valuable assistants to educators, functioning as educational tools [38], [53], [72], [77], [107]. However, for this integration to be effective, Wagholikar et al. [59] suggest that teachers must have opportunities to interact with and analyse these tools. Furthermore, as Konecki et al. [67] noted, the teaching curriculum itself must evolve to accommodate and effectively utilise these technologies.

Overall, the literature advocates for a balanced adoption of genAI tools in education [65], [82], [135]. Sharma et al. [123] and Tsang [132] emphasise that these tools should complement existing educational practices rather than replace them. Tafferner et al. [128] further suggest that this balance should involve weighing human capabilities against those of AI, ensuring that both are effectively integrated into the educational process.

The second category, *Innovative Assessment Design and Curricula Adaptation*, was mentioned in twenty-one publications. This category highlights the need to reassess and modernise educational practices, particularly the methods used for student assessments. As explored in earlier sections, integrating genAI into education offers advantages and challenges, with academic dishonesty being one significant concern. Lodge et al. [105] argued that to mitigate this issue, there must be changes in how assessments are designed.

Susnjak [36] suggests that incorporating multimodal exam formats and favouring oral assessments could be effective strategies. These approaches would limit students' opportunities to engage with genAI tools during exams, as they would be under direct supervision. Complementing Susnjak's perspective, Kortemeyer [100] proposes dividing assessments into two parts: a discussion-based exam and an exam conducted without internet access, thereby minimising the risk of students interacting with genAI models.

Another perspective from Lo [104] emphasises the importance of innovating exams to include multi-modal assessments and questions that require students to demonstrate analytical and critical thinking skills. This approach aligns with the findings of Arghavan et al. [66], who, in their study on computer education and coding AI tools, suggest that as AI tools become capable of solving exam-level coding problems, educators need to adapt by crafting questions that encourage deeper critical thinking and problem-solving. The literature suggests that the current educational curriculum must innovate to follow the technological advancements [88], [90].

This subsection's third and final category is *Development and Research of AI in Education*, mentioned in ten articles. This category emphasises the importance of expanding research and development efforts in the field of genAI in education. Some authors argue that there is a pressing need for more studies to better understand the implications and potential of genAI within educational contexts [4], [7], [13], [57], [109]. Lodge et al. [105] underscore the necessity of being proactive in researching genAI, given its rapid growth and the significant impacts it could have on education.

Beyond increasing research efforts, Firat [5] advocates for developing more AI tools that can be integrated into educational settings. The literature generally supports the idea of not only conducting more research on genAI but also focusing on the creation of innovative educational tools [13]. Bauer et al. [78] suggest that universities should take the lead in developing these tools, which could also be applied in administrative software to enhance overall educational infrastructure.

The research identifies four AI tools developed to assist different educational stakeholders. The first is Smart Uenglish, created by Hwang and Nurtantyana [51], designed to help students improve their English skills. The study concluded that the suggestions provided by this application significantly enhanced the quality of student essays. The second tool, EconBot, was mentioned by Konecki et al. [67] and supports students in their economics classes. The third tool, Smart Sparrow, highlighted by Adiguzel et al. [68], aims to provide engaging learning materials to increase student involvement. Finally, Aisha, an educational tool developed by Lappalainen and Narayanan [102], is targeted at librarians and is designed to assist students with research when the library is closed.

6.3.2. Interaction, Training, and Ethical Considerations

The second theme within the future suggestions for genAI in education focuses on how students and educators can benefit from interacting with these tools. Additionally, it addresses the ethical considerations that must be presented to ensure that these interactions occur safely and responsibly within educational environments. Through this research, four main categories were identified, as summarised in Table 6.15.

Table 6.15. Frequency of Mentions of Categories for Interaction, Training, and Considerations.

Category	Count	Article References
Guidelines and Ethical Education for AI Use	44	[1] [2] [4] [5] [6] [7] [8] [9] [10] [11] [14] [16] [35] [39] [50] [55] [68] [72] [77] [79] [82] [83] [84] [43] [89] [92] [104] [105] [106] [107] [109] [110] [112] [113] [114] [115] [116] [118] [120] [122] [127] [132] [134] [135]
Awareness and Skill Development for Emerging Technologies	28	[2] [5] [9] [14] [42] [43] [46] [56] [58] [64] [65] [70] [72] [76] [77] [81] [82] [84] [85] [89] [94] [98] [104] [117] [120] [128] [132] [134]
Teacher Training with AI	11	[52] [54] [70] [73] [97] [105] [109] [110] [120] [127] [135]
AI Detection Tools	8	[16] [18] [36] [65] [84] [100] [115] [132]

The first category, *Guidelines and Ethical Education for AI Use*, is the most frequently mentioned within the future suggestions theme, with forty-four publications addressing it. Yan [135] highlights the importance of providing educators and students with resources and training to understand the implications of using genAI. This includes the ethical considerations that must be considered when interacting with these tools, as well as the considerations of using the content generated by them. The literature advises the creation of courses that educate students, educators, and other members on how to effectively use, interact with, and benefit from genAI, while also addressing the ethical concerns it raises, such as issues related to academic integrity [1],[4], [7], [8], [9], [10], [14], [16], [35], [39], [43], [50], [72], [83], [104], [105], [112], [114], [132], [135].

Approximately seventy-seven percent of the articles emphasise the need to establish academic guidelines to ensure that institutions, teachers, and students are well-informed and prepared for the implications of using genAI tools [2], [4], [5], [6], [11], [14], [43], [55], [68], [72], [77], [79], [82], [84], [89], [104], [105], [106], [107], [109], [110], [112], [113], [114], [115], [116], [118], [120], [122], [127], [132], [134], [135]. Firat [5] stresses the necessity of establishing these guidelines, while Adiguzel et al. [68] argue that for such guidelines to be practical, they must involve educators, students, administrative departments, and researchers.

These guidelines should be comprehensive, covering all levels of education, not just higher education, as suggested by Luo et al. [107]. They should also clearly define how genAI tools can be used to benefit students [113]. Once these policies are in place, teachers will be better equipped to instruct their students on applying these guidelines in their coursework, ensuring that genAI is ethical and productive, as noted by Cotton et al. [84]. This approach aims to create an educational environment that maximises the benefits of genAI while minimising potential ethical risks.

The second category, *Awareness and Skill Development for Emerging Technologies*, is mentioned in twenty-eight publications. This category includes the suggested approaches for leveraging genAI to enhance students' skill development and the need for institutions and educators to stay informed about emerging technologies, including new educational frameworks. The literature emphasises the importance of maintaining constant awareness of emerging tools [2], [46], as institutions must be proactive in integrating these advancements into educational practices, as stated by Johinke et al. [98]. This awareness extends beyond administrative departments and educators to include students, as highlighted by Ajevski et al. [77]. Effective communication about these tools and technologies should flow across all departments to ensure comprehensive understanding and adoption.

The literature also emphasises the need for ongoing awareness as these tools evolve rapidly. With this awareness comes the opportunity to recognise and develop the skills that genAI can enhance in students. As AI becomes more accessible, cultivating AI literacy among students is increasingly important. One practical approach suggested in the literature is to teach students how to design effective prompts [70], [94], enabling them to ask precise questions that yield the most accurate and relevant responses from these tools.

However, there is a concern that certain fundamental skills, such as critical thinking, could be undermined as genAI tools increasingly generate text quickly and easily. The literature emphasises the importance of encouraging critical thinking in the classroom and through corresponding assignments or assessments [9], [14], [42], [43], [56], [58], [64], [65], [70], [76], [81], [82], [84], [85], [89], [94], [104], [117], [120], [128], [132], [134]. Malinka et al. [65] suggest the need for educators and students to analyse and verify the content generated by these tools critically. Cotton et al. [84] further recommend that teachers create assessments that demand students to develop critical thinking skills, thus ensuring that this essential skill is not lost but strengthened through genAI.

The third category, *Teacher Training with AI*, is highlighted in eleven articles. Given that genAI tools represent a rapidly emerging technology, many educators have little to no prior experience or familiarity with these tools. To promote educational equity, it is crucial to encourage and support teachers in engaging with genAI. This engagement will enable educators to understand the capabilities of these tools, how they can assist in their teaching, and how students can benefit from their use. The literature encourages providing teachers with the necessary training and support to learn and integrate genAI into their educational practices effectively [52], [54], [70], [73], [97], [105], [109], [110], [120], [127], [135]. This training is essential not only for enhancing teachers' technological proficiency but also for ensuring that they can guide their students in using genAI in ways that enrich learning and maintain educational standards.

The final category, *AI Detection Tools*, is mentioned in eight articles and represents a point of argument within the literature. This category revolves around the debate regarding the effectiveness and necessity of tools designed to detect AI-generated content in student submissions. The primary concern is that current plagiarism detection tools cannot identify with certainty whether a student's work includes content generated by genAI. As a result, there is a strong call in the literature for the development and integration of AI detection tools within educational settings [16], [18], [36], [65], [84], [100], [115], [132]. However, some authors, like Rudolph et al. [72], express significant concerns about the current state of AI detection tools. These tools are still in their early development stage and may not always function reliably, leading to the potential for false positives. Such inaccuracies could result in students being unfairly penalised for using AI tools even when they have not, which raises serious ethical and practical concerns.

6.4. Educational Level

As genAI tools are accessible to users across all age groups, their application in education covers multiple educational levels. This section analyses the educational levels discussed in the articles reviewed in the SLR, presenting a distribution of the extracted data. The primary objective of this analysis is to identify the educational levels most frequently discussed throughout the literature and the ones being overlooked in research.

The levels defined in the analysis include kindergarten, elementary school, middle school, high school, and university, as well as a general category for articles that do not specify a particular educational level. Figure 6.2 illustrates the distribution of these categories, visually representing the number of articles that mention each level. Most publications focus on higher education, particularly university-level studies, experiments, or explorations related to genAI in education.

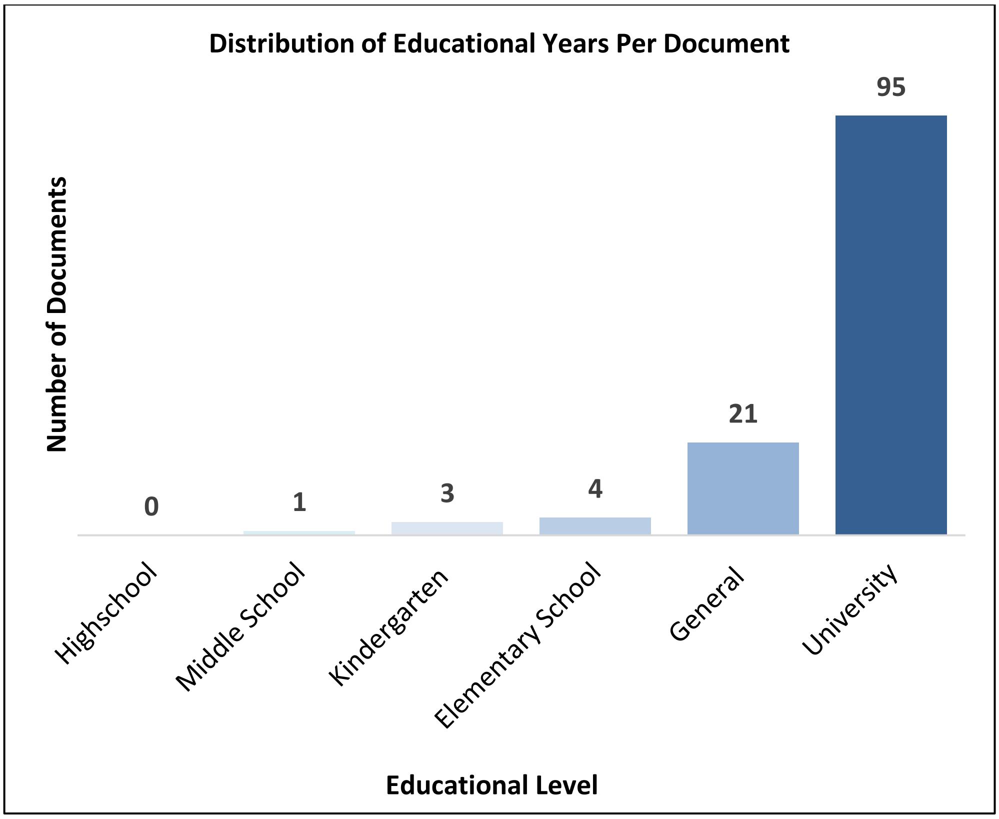

Figure 6.2. Distribution of Educational Years Per Document.

According to the literature, approximately seventy-seven percent of the articles reference university-level education. This is followed by seventeen percent of the articles that fall under the general category, where no specific educational level is mentioned. In contrast, three percent of the articles address elementary school, while two percent mention kindergarten. Middle school is scarcely referenced, with only one article discussing it, and there are no mentions of high school education. This data is elaborated in Table A.1, which provides a detailed breakdown of the articles by educational level they reference.

As reflected in the graph, the substantial emphasis on higher education reveals a significant gap in the literature concerning the exploration of genAI at other educational levels. Only Su and Yang [13] and Han and Cai [9] mention earlier stages of education, specifically kindergarten and elementary school. Of the 121 publications reviewed, only Bekeš and Galzina [52] mention middle school. This underrepresentation at the earlier stages of the educational path is notable as learners develop their foundational skills in the early years of schooling. Addressing this gap is crucial for understanding the broader impact of genAI across the entire educational spectrum.

Given that most articles focus on higher education, Figure 6.3 provides further insight by representing the distribution of university-level articles, distinguishing between those targeting introductory or undergraduate classes and those addressing advanced-level courses. The distribution shows that eighty percent of the university-level articles are focused on introductory or undergraduate education, while twenty percent focus on advanced-level education. This set consists of twenty articles within the university-level dataset, specifically mentioning whether they are introductory classes or advanced.

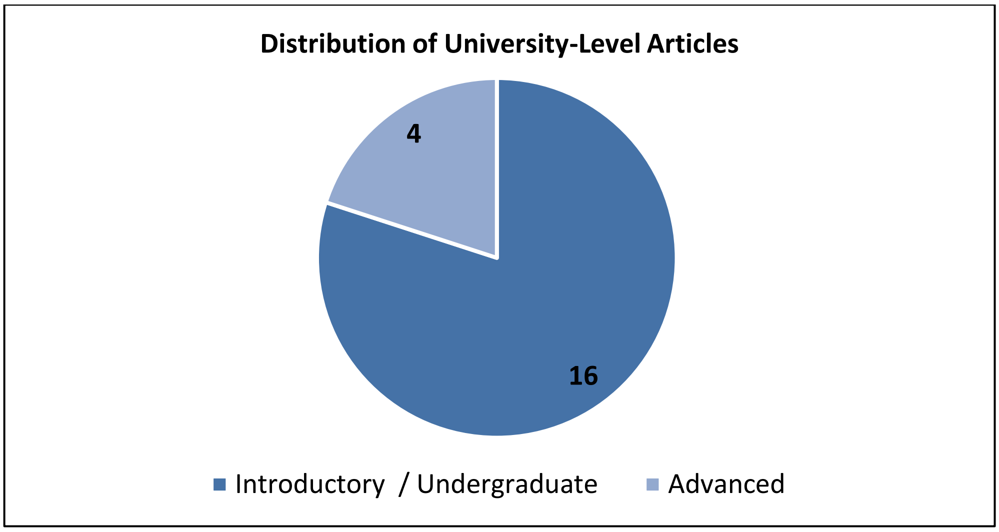

Figure 6.3. Distribution of University-Level Articles: Introductory vs. Advanced.

This additional data is detailed in Table A.2 and reflects a strong focus on undergraduate education in the literature. While this focus is valuable for understanding how genAI tools can support early university students, there remains a need for more in-depth exploration of genAI's applications at more advanced levels of education. This section further expands the focus on higher education by analysing the specific areas of study that genAI research targets. As shown in Figure 6.4, the articles analysed cover a broad range of academic disciplines, offering insights into the fields where genAI has been most actively researched and implemented in educational settings.

The areas of study explored in the literature range from medicine and health, tourism, and legal studies to more specialised fields like economics and sports management. Based on the research, the most frequently discussed area is medicine and health, which is targeted by twenty-seven articles. This category includes multiple subfields. For instance, the study by Sallam et al. [21] covers pharmacy, public health, dental, and medical education. Several authors specifically focus on the implications of genAI in pharmacy education [18], [29], [91], while Subramani et al. [47] investigate the use of ChatGPT in medical physiology.

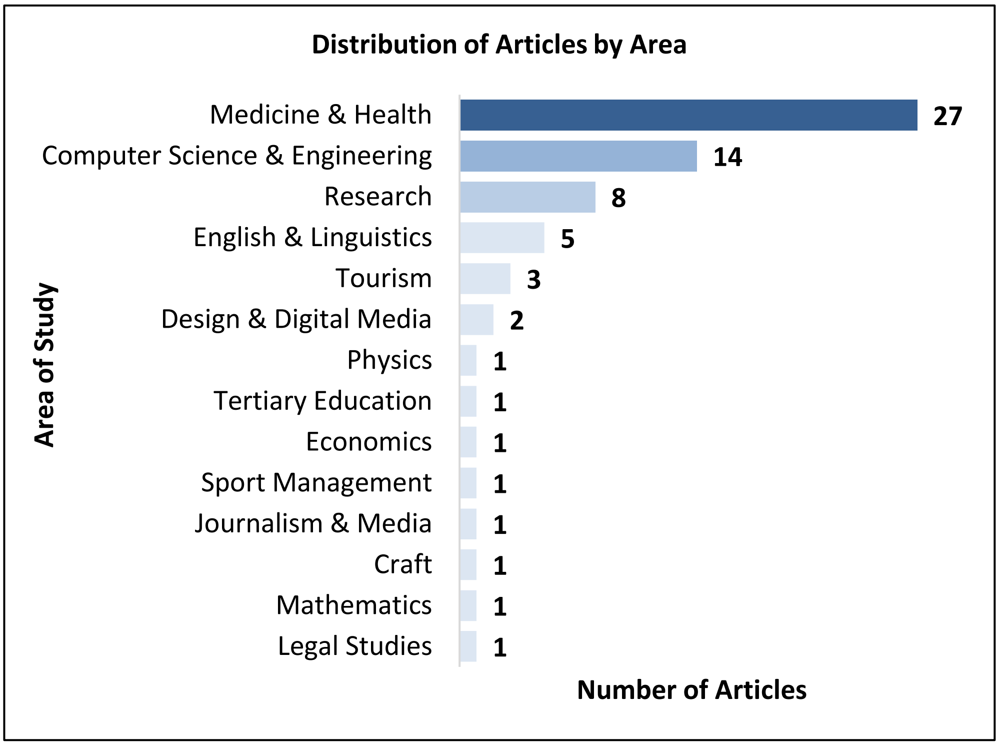

Figure 6.4. Distribution of Articles per Area of Study.

Following medicine and health, the next most prominent field is computer science and engineering, with fourteen publications focused on this area. Daun and Brings [1] explore the potential of genAI in software engineering, concluding that while it can benefit teachers, its use should be carefully managed for introductory-level students. Other studies, such as those by Qadir [11] and Shoufan [59], also focus on software engineering. Some explore computer science more broadly [58], [65], [66], [70], while others, like Tafferner et al. [128], extend their research into electrical engineering.

The third most discussed area is academic research, with eight articles examining how genAI can support or influence this field. Following academic research, English and linguistics is represented by five mentions. This area includes studies on the impact of genAI on students learning English as a foreign language [109], [112] and on students in general English classes [58], [135]. Other areas of study mentioned in the literature, though less frequently, include tourism with three mentions, design and digital media with two mentions, and fields such as physics, tertiary education, economics, sports management, journalism and media, mathematics, craft, and legal studies, each with one article targeting their exploration of genAI in education.

These findings emphasise the wide-ranging interest in genAI across various disciplines but also underscore certain disparities in focus. Fields like medicine, health, and computer science receive the most attention, while others remain underexplored. This data is elaborated in Table A.3, which provides a detailed breakdown of the articles referencing the corresponding area of study.

6.5. GenAI Tools

The analysis of the various genAI tools mentioned throughout the studies included in this SLR aims to provide a comprehensive overview of the most frequently referenced tools within education. The literature identified 135 distinct genAI tools with varying levels of discussion and emphasis. Table B.1 contains the complete list of these tools, highlighting that while tools like ChatGPT are extensively referenced, others such as BioGPT or Stability AI receive comparatively fewer mentions. For this section, the top 18 most mentioned tools were selected and represented in a distribution chart, as illustrated in Figure 6.5.

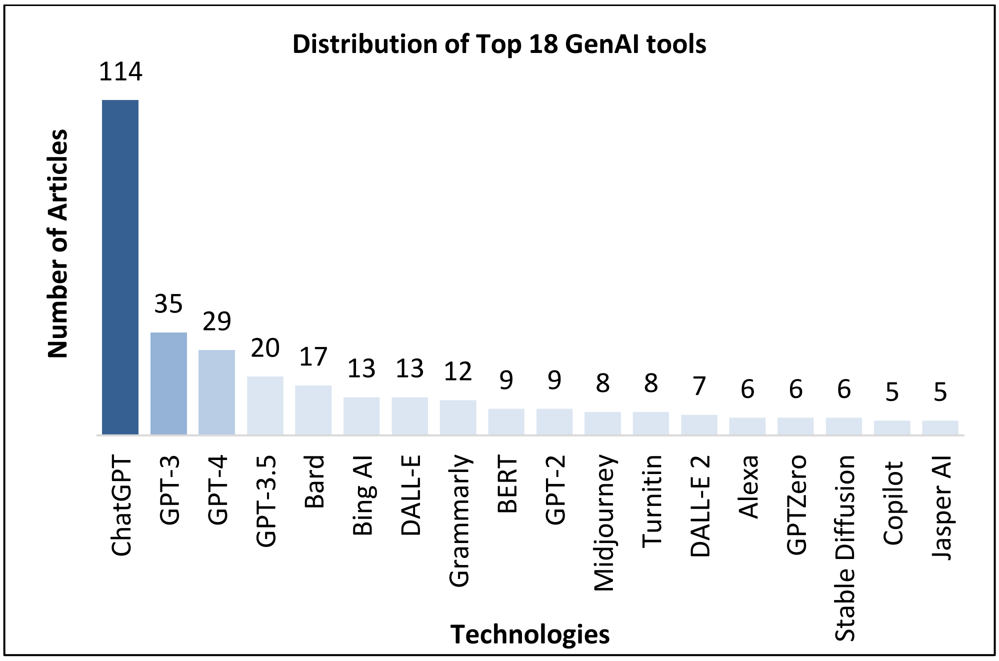

Figure 6.5. Distribution of Top 18 GenAI mentioned tools.

ChatGPT stands out significantly among the identified tools, being mentioned in approximately ninety-four percent of the articles. This overwhelming dominance emphasises the prevalent popularity and relevance of ChatGPT in discussions related to genAI, mainly as it represents an accessible and recently developed technology—much of the current debate and research surrounding genAI in education centres on this tool. For instance, Khan et al. [99] discuss the benefits and limitations of ChatGPT in medical education, while Nikolic et al. [39] examine its outputs when completing engineering assessments at the undergraduate level. Sedaghat [119] also explores its advantages in medical education and research, indicating its broad application in educational settings.

Following ChatGPT, GPT-3 and GPT-4 hold the second and third positions, with thirty-five and twenty-nine mentions, respectively. GPT-4, being a paid version, has been the subject of comparative studies. For instance, Oh et al. [111] analysed the performance of GPT-4 in medical education, particularly in understanding complex information, and concluded that GPT-4 offers greater accuracy and reliability when compared with its previous models.

Next in the distribution of mentioned tools are the chatbots Bard and Bing AI, with seventeen and thirteen mentions, respectively. These tools function similarly to ChatGPT but have not reached the same prominence in academic literature. ChatGPT paid version has multimodal capabilities, but there are tools specifically designed to generate outputs different from text, like images. According to the literature, the most mentioned tools to create images are DALL-E, DALL-E 2, and Midjourney, with thirteen, seven, and eight mentions, respectively. Han and Cai [9] explored the use of these tools in visual storytelling for children in preschool and elementary settings, demonstrating their potential to enhance educational content through visual media.

Furthermore, the literature highlights code generation tools like Copilot, which has five mentions. Dakhel et al. [3] evaluated Copilot's code generation capabilities and found that the code it generated was superior to that produced by some junior developers. Conversely, tools such as GPTZero, which is referenced six times, are in development to identify AI-generated content. This highlights the continuous efforts to tackle the challenges that come with utilising genAI in educational settings.

This analysis demonstrates the diverse applications and growing interest in various genAI tools, each offering unique contributions to education. The focus on ChatGPT and its versions, alongside the emerging use of other tools, highlights the dynamic nature of this field and the need for continued research and exploration.

CHAPTER 7

Reflections

The proposed methodology for this research, an SLR, has provided valuable insights into the implications of integrating genAI tools into educational contexts. This analysis revealed both opportunities and challenges, offering a deeper understanding of the current state of research.

The findings underscore numerous applications and benefits of genAI when effectively integrated into education. These tools have demonstrated their potential to enhance productivity in various ways. For example, genAI can assist educators by automating repetitive and time-consuming tasks, such as grading assignments. This automation not only saves time and resources but also ensures greater objectivity in assessments. Additionally, genAI can support students in managing their time effectively by providing rapid access to information and personalised assistance, enabling them to focus on other academic tasks.

GenAI tools are recognized for generating high-quality, “human-like” text, a feature that benefits both students and educators. For students, these tools serve as writing assistants, helping them refine their ideas and improve the quality of their work. For educators and institutions, genAI can enhance administrative tasks, such as creating polished emails, guidelines, and curriculum materials.

The literature also highlights the creative potential of genAI, with its ability to generate engaging materials such as educational games or brainstorming suggestions. Its versatility enables educators to tailor content for diverse audiences, accounting for factors such as age, language, and learning needs, thereby advancing accessibility and inclusivity in education.

Moreover, genAI tools foster skill development by exposing students to new technological paradigms and promoting interaction with AI. These tools, especially multi-modal models, can generate and interpret both text and images, offering students a richer, more dynamic learning experience. Personalised feedback from genAI, available outside of traditional school hours, further enhances learning by reducing response times and providing judgment-free interactions.

GenAI has also been likened to virtual tutors or research assistants, capable of summarising, analysing, or translating large volumes of text efficiently. These features benefit students, particularly those who face language barriers, by improving communication skills and reducing anxiety associated with language proficiency. In terms of assessment, the literature notes genAI's ability to excel in some university-level exams, even outperforming students in certain cases, demonstrating its utility in preparing learners for evaluations.

Importantly, genAI tools are accessible at low or no cost, ensuring that their benefits can extend to students even without institutional adoption, thereby democratizing educational resources.

Despite its advantages, the integration of genAI in education raises significant concerns and limitations. Academic integrity and ethical challenges are among the most frequently cited issues. Plagiarism poses a primary concern, as students can misuse these tools to complete assignments or cheat on exams. The ease of access and interaction with genAI exacerbates this issue, and its ability to generate convincing "human-like" text complicates the detection of misconduct.

Concerns also extend to the accuracy, reliability, and quality of genAI-generated content. Since these models are trained on information that comes from multiple sources, such as the Internet or books, they can contain biases or inaccuracies in the training data. Instances of genAI tools generating falsified information, such as fabricated references, have been reported, highlighting the need for caution. Furthermore, the use of user-provided data to train these tools raises significant privacy and safety concerns, especially when sensitive information is involved.

Over-reliance on genAI is another issue, with potential negative impacts on essential skills such as critical thinking, problem-solving, and writing. Students who depend excessively on these tools risk diminishing their ability to perform independently. Additionally, the use of genAI may reduce teacher-student interactions, particularly in practical or hands-on subjects like music or craft education, where human interaction is crucial.

Authorship and accountability present further challenges. GenAI tools cannot be cited as authors, which raises questions about the ownership of content generated with their assistance. Moreover, existing plagiarism detection tools are not adequately equipped to identify AI-generated content, increasing the challenges for educators.

To address these limitations and concerns, the literature strongly advocates for the development of robust strategies and policies. Institutions should establish clear guidelines for genAI usage to ensure that all educational stakeholders know the appropriate practices. Proactive measures, such as integrating genAI into curricula and teaching methods, can help institutions keep up with technological advancements while minimizing risks.

Educators should be provided with training opportunities to familiarize themselves with genAI tools, enabling them to interact with these technologies effectively in their classes. Simultaneously, there is a need for innovative approaches to assessment. Suggestions include reverting to pen-and-paper exams, conducting oral assessments, or designing tasks that require critical thinking and problem-solving skills, which are less amenable to genAI assistance.

In addition to updating assessments, the development of educational tools tailored to specific learning objectives is essential. Research efforts should focus on creating tools that complement traditional teaching methods while enhancing students' learning experiences.

The integration of genAI in education presents both advantages and challenges. While its applications offer transformative potential, the associated risks necessitate careful consideration and strategic planning. By addressing concerns through the development of guidelines, innovative assessment methods, and targeted training, educational institutions can harness the benefits of genAI while safeguarding the integrity of learning environments.

CHAPTER 8

Conclusion

Since the appearance of genAI tools accessible to the public, there has been a growing interest in studying and exploring the impact of this technology across various fields, including education. This research aimed to investigate the ongoing research of genAI in education by examining its benefits and challenges for students, educators, and institutions. Additionally, the study sought to determine whether the research conducted so far has addressed all educational levels or focused primarily on specific ones. The study also explored which genAI tools are most frequently referenced when discussing this technology in education.

This investigation was carried out through a Systematic Literature Review of 121 articles published between 2018 and 2023. The review provided a comprehensive overview of the state of research on genAI in education. The analysis revealed various applications, benefits, limitations, and risks associated with integrating genAI into educational settings.

One of the main advantages discussed in the literature is the ability of genAI to enhance the learning experience for students by acting as writing assistants, thereby improving the quality of their work throughout the writing process—from idea generation to final composition. Many studies also highlighted the high quality and versatility of the text generated by these tools, which can support not only students and educators but also the administrative tasks performed within institutions. These tools, capable of generating content in multiple languages, contribute to diverse communication needs within educational settings. Another prominent theme identified was the personalised experience that these tools could provide, with genAI offering continuous support outside of traditional school hours, enabling students to interact with a virtual tutor in a judgment-free environment. Overall, twenty-eight applications and benefits were mapped out in this research and reported in section 6.1.

However, the study also identified twenty concerns and limitations associated with genAI use, as reported in section 6.2. A primary concern is the potential for these tools to produce biased or falsified information. Given that models like GPT-4 are trained on vast datasets sourced from the Internet, which may contain inherent biases. Likewise, the accuracy and reliability of the information generated by these tools remain a concern, as they can disseminate both inaccurate and unreliable content. Beyond content reliability, integrating genAI in education raises issues related to academic integrity and ethical risks. Plagiarism is one of the most frequently cited concerns, alongside the potential for students to misuse these tools for cheating on assessments or completing assignments dishonestly. Such misuse could lead to an over-reliance on AI, ultimately diminishing essential skills like critical thinking and problem-solving, which are crucial for educational and personal growth.

These concerns demonstrate the need for strategies and policies to maintain academic integrity as genAI becomes more integrated into educational settings. The study's findings indicate that it is crucial to develop clear guidelines for interacting with these tools, ensuring that all educational stakeholders know what constitutes appropriate use. Moreover, research suggests that incorporating genAI into educational programs and pedagogical methods can be advantageous. It has the potential to lighten the workload of educators, enabling them to allocate more time for direct interaction and assistance with students.

Regarding the educational levels addressed in the SLR, most studies focused on higher education, accounting for approximately seventy-seven percent of the reviewed articles. This suggests a significant gap in research exploring the impact of genAI across different educational levels, from kindergarten to high school education.

Within higher education, most articles focused on fields such as medicine and health, computer science and engineering, and academic research. This highlights another gap in the literature, as there is a need for more studies that explore the application of genAI in other academic disciplines. ChatGPT emerged as the most prominently mentioned genAI tool, emphasising its popularity and widespread use in educational research. However, this dominance may shift as new genAI models and tools are being developed by various entities such as OpenAI, Google, and emerging startups.

In conclusion, this research highlights the potential of genAI to transform education, provided that its integration is approached with caution and guided by clear policies and ethical considerations. Future research should focus on exploring the impact of genAI across all educational levels and disciplines.

7.1. Limitations

While this SLR provided valuable insights, it is not without its limitations. First, the dataset for this research was restricted to articles published between 2018 and 2023, as it reflects the time of extraction from the databases. This limitation may have resulted in the exclusion of more recent studies and developments. Additionally, excluding non-English articles could have omitted significant research contributions from other linguistic contexts, potentially narrowing the scope of the review.

Moreover, given that genAI is an emerging technology, relying only on white literature may have excluded essential findings from grey literature sources, such as reports, industry publications, or educational blogs. These sources could provide a more comprehensive and practical understanding of genAI's impact on education.

References

[1] Daun, M., & Brings, J. (2023). How ChatGPT Will Change Software Engineering Education. *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, 110–116. https://doi.org/10.1145/3587102.3588815

[2] Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *International Journal of Management Education*, *21*(2). https://doi.org/10.1016/j.ijme.2023.100790

[3] Dakhel, A. M., Majdinasab, V., Nikanjam, A., Khomh, F., Desmarais, M. C., Ming, Z., & Jiang. (2022). *GitHub Copilot AI pair programmer: Asset or Liability?* http://arxiv.org/abs/2206.15331

[4] Peres, R., Schreier, M., Schweidel, D., & Sorescu, A. (2023). Editorial: On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice. *International Journal of Research in Marketing*, *40*(2), 269–275. https://doi.org/10.1016/j.ijresmar.2023.03.001

[5] Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching*, *6*(1), 57–63. https://doi.org/10.37074/jalt.2023.6.1.22

[6] Cascella, M., Montomoli, J., Bellini, V., & Bignami, E. (2023). Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *Journal of Medical Systems*, *47*(1), 33. https://doi.org/10.1007/s10916-023-01925-4

[7] Pavlik, J. v. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism and Mass Communication Educator*, *78*(1), 84–93. https://doi.org/10.1177/10776958221149577

[8] Rasul, T., Nair, S., Kalendra, D., Robin, M., Santini, F. O., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching*, *6*(1), 41–56. https://doi.org/10.37074/jalt.2023.6.1.29

[9] Han, A., & Cai, Z. (2023). Design implications of generative AI systems for visual storytelling for young learners. *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 470–474. https://doi.org/10.1145/3585088.3593867

[10] Cox, C., & Tzoc, E. (2023). ChatGPT Implications for academic libraries. *College and Research Libraries News*, *84*(3), 99–102. https://doi.org/10.5860/crln.84.3.99

[11] Qadir, J. (2023). Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education. *2023 IEEE Global Engineering Education Conference (EDUCON)*, 1–9. https://doi.org/10.1109/EDUCON54358.2023.10125121

[12] Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays — should professors worry? *Nature*. https://doi.org/10.1038/d41586-022-04397-7

[13] Su (苏嘉红), J., & Yang (杨伟鹏), W. (2023). Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education. *ECNU Review of Education*, *6*(3), 355–366. https://doi.org/10.1177/20965311231168423

[14] Siegle, D. (2023). A Role for ChatGPT and AI in Gifted Education. *Gifted Child Today*, *46*(3), 211–219. https://doi.org/10.1177/10762175231168443

[15] Murgia, E., Pera, M. S., Landoni, M., & Huibers, T. (2023). Children on ChatGPT Readability in an Educational Context: Myth or Opportunity? *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, 311–316. https://doi.org/10.1145/3563359.3596996

[16] Giannos, P., & Delardas, O. (2023). Performance of ChatGPT on UK Standardized Admission Tests: Insights from the BMAT, TMUA, LNAT, and TSA Examinations. *JMIR Medical Education*, *9*. https://doi.org/10.2196/47737

[17] Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching*, *6*(1), 31–40. https://doi.org/10.37074/jalt.2023.6.1.17

[18] Li, Y., Sha, L., Yan, L., Lin, J., Raković, M., Galbraith, K., Lyons, K., Gašević, D., & Chen, G. (2023). Can large language models write reflectively. *Computers and Education: Artificial Intelligence*, *4*. https://doi.org/10.1016/j.caeai.2023.100140

[19] Xie, Y., Seth, I., Rozen, W. M., & Hunter-Smith, D. J. (2023). Evaluation of the Artificial Intelligence Chatbot on Breast Reconstruction and Its Efficacy in Surgical Research: A Case Study. *Aesthetic Plastic Surgery*. https://doi.org/10.1007/s00266-023-03443-7

[20] Pataranutaporn, P., Leong, J., Danry, V., Lawson, A. P., Maes, P., & Sra, M. (2022). AI-Generated Virtual Instructors Based on Liked or Admired People Can Improve Motivation and Foster Positive Emotions for Learning. *2022 IEEE Frontiers in Education Conference (FIE)*, 1–9. https://doi.org/10.1109/FIE56618.2022.9962478

[21] Sallam, M., Salim, N. A., Barakat, M., & Al-Tammemi, A. B. (2023). ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study

highlighting the advantages and limitations. *Narra J*, *3*(1). https://doi.org/10.52225/narra.v3i1.103

[22] Alasadi, J., AlHilli, A., Atrey, P. K., & Singh, V. K. (2022). A Generative Approach to Mitigate Bias in Face Matching using Learned Latent Structure. *2022 IEEE Eighth International Conference on Multimedia Big Data (BigMM)*, 150–157. https://doi.org/10.1109/BigMM55396.2022.00033

[23] Eysenbach, G. (2023). The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. *JMIR Medical Education*, *9*, e46885. https://doi.org/10.2196/46885

[24] Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

[25] Mccarthy, J. (2007). *WHAT IS ARTIFICIAL INTELLIGENCE?* http://www-formal.stanford.edu/jmc/

[26] Stryker, C., & Kavlakoglu, E. (n.d.). What is artificial intelligence (AI)? *IBM*. Retrieved from https://www.ibm.com/topics/artificial-intelligence

[27] Jajal, T. (2018, May 21). Distinguishing between narrow AI, general AI, and super AI. *AIDEN Global*. Retrieved from https://medium.com/aiden-global/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22

[28] Musthafa, M. (n.d.). Narrow AI vs. General AI. *How to Learn Machine Learning*. Retrieved from https://howtolearnmachinelearning.com/articles/narrow-ai-vs-general-ai/

[29] Wagholikar, S., Chandani, A., Atiq, R., Pathak, M., & Wagholikar, O. (2023). ChatGPT - Boon or Bane: A Study from Students Perspective. *2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, 207–212. https://doi.org/10.1109/InCACCT57535.2023.10141820

[30] EITC. (n.d.). The stages and categorizations of AI. *EITC*. Retrieved from http://www.eitc.org/research-opportunities/new-media-and-new-digital-economy/ai-machine-learning-deep-learning-and-neural-networks/ai-research-and-applications/the-stages-and-categorizations-of-ai

[31] IBM. (n.d.). What is machine learning (ML)? *IBM*. Retrieved from https://www.ibm.com/topics/machine-learning

[32] Crabtree, M. (2023, July 19). What is machine learning? Definition, types, tools & more. *DataCamp*. Retrieved from https://www.datacamp.com/blog/what-is-machine-learning

[33] IBM. (n.d.). Deep Learning. *IBM*. Retrieved from https://www.ibm.com/topics/deep-learning

[34] AWS. (n.d.). What is deep learning? *AWS*. Retrieved from https://aws.amazon.com/what-is/deep-learning/

[35] Fayyad, U. M. (2023). From Stochastic Parrots to Intelligent Assistants—The Secrets of Data and Human Interventions. *IEEE Intelligent Systems*, *38*(3), 63–67. https://doi.org/10.1109/MIS.2023.3268723

[36] Susnjak, T. (2022). *ChatGPT: The End of Online Exam Integrity?* http://arxiv.org/abs/2212.09292

[37] Toner, H. (2023, May 12). What are generative AI, large language models, and foundation models? *Center for Security and Emerging Technology*. Retrieved from https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/

[38] Eager, B., & Brunton, R. (2023). Prompting Higher Education Towards AI-Augmented Teaching and Learning Practice. *Journal of University Teaching and Learning Practice*, *20*(5). https://doi.org/10.53761/1.20.5.02

[39] Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., Neal, P., & Sandison, C. (2023). ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European Journal of Engineering Education*, 48(4), 559–614. https://doi.org/10.1080/03043797.2023.2213169

[40] Kitchenham, Barbara. (2004). Procedures for Performing Systematic Reviews. Keele, UK, Keele Univ.. 33.

[41] Garousi, V., Felderer, M., & Mäntylä, M. V. (2019). Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology, 106*, 101-121. https://doi.org/10.1016/j.infsof.2018.09.006

[42] Strzelecki, A. (2023). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments*, 1–14. https://doi.org/10.1080/10494820.2023.2209881

[43] Dergaa, I., Chamari, K., Zmijewski, P., & ben Saad, H. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, *40*(2), 615–622. https://doi.org/10.5114/biolsport.2023.125623

[44] Webster, Jane & Watson, Richard. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly. 26. 10.2307/4132319.

[45] Garcia, M. B. (2023). Using AI Tools in Writing Peer Review Reports: Should Academic Journals Embrace the Use of ChatGPT? *Annals of Biomedical Engineering*. https://doi.org/10.1007/s10439-023-03299-7

[46] Milton, C. L. (2023). ChatGPT and Forms of Deception. *Nursing Science Quarterly*, *36*(3), 232–233. https://doi.org/10.1177/08943184231169753

[47] Subramani, M., Jaleel, I., & Mohan, S. K. (2023). Evaluating the performance of ChatGPT in medical physiology university examination of phase I MBBS. *Advances in Physiology Education*, *47*(2), 270–271. https://doi.org/10.1152/ADVAN.00036.2023

[48] Geerling, W., Mateer, G. D., Wooten, J., & Damodaran, N. (2023). ChatGPT has Aced the Test of Understanding in College Economics: Now What? *The American Economist*, *68*(2), 233–245. https://doi.org/10.1177/05694345231169654

[49] Hurlburt, G. (2023). What If Ethics Got in the Way of Generative AI? *IT Professional*, *25*(2), 4–6. https://doi.org/10.1109/MITP.2023.3267140

[50] Ahmad, N., Murugesan, S., & Kshetri, N. (2023). Generative Artificial Intelligence and the Education Sector. *Computer*, *56*(6), 72–76. https://doi.org/10.1109/MC.2023.3263576

[51] Hwang, W.-Y., & Nurtantyana, R. (2022). The integration of multiple recognition technologies and artificial intelligence to facilitate EFL writing in authentic contexts. *2022 6th International Conference on Information Technology (InCIT)*, 379–383. https://doi.org/10.1109/InCIT56086.2022.10067490

[52] Bekeš, E. R., & Galzina, V. (2023). Exploring the Pedagogical Use of AI-Powered Chatbots Educational Perceptions and Practices. *2023 46th MIPRO ICT and Electronics Convention (MIPRO)*, 636–641. https://doi.org/10.23919/MIPRO57284.2023.10159734

[53] Banić, B., Konecki, M., & Konecki, M. (2023). Pair Programming Education Aided by ChatGPT. *2023 46th MIPRO ICT and Electronics Convention (MIPRO)*, 911–915. https://doi.org/10.23919/MIPRO57284.2023.10159727

[54] Kovačević, D. (2023). Use of ChatGPT in ESP Teaching Process. *2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH)*, 1–5. https://doi.org/10.1109/INFOTEH57020.2023.10094133

[55] Elder, C., Pozek, G., Horine, S., Tripaldelli, A., & Butka, B. (2023). Can Artificial Intelligence Pass a Sophomore Level Digital Design Laboratory? *SoutheastCon 2023*, 861–868. https://doi.org/10.1109/SoutheastCon51012.2023.10115116

[56] Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R. J., Esmaeili, M., Majdabadkohne, R. M., & Pasehvar, M. (2023). ChatGPT: Applications, Opportunities, and

Threats. *2023 Systems and Information Engineering Design Symposium (SIEDS)*, 274–279. https://doi.org/10.1109/SIEDS58326.2023.10137850

[57] Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., & Tang, Y. (2023). A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development. *IEEE/CAA Journal of Automatica Sinica*, *10*(5), 1122–1136. https://doi.org/10.1109/JAS.2023.123618

[58] Ibrahim, H., Asim, R., Zaffar, F., Rahwan, T., & Zaki, Y. (2023). Rethinking Homework in the Age of Artificial Intelligence. *IEEE Intelligent Systems*, *38*(2), 24–27. https://doi.org/10.1109/MIS.2023.3255599

[59] Shoufan, A. (2023). Exploring Students' Perceptions of ChatGPT: Thematic Analysis and Follow-Up Survey. *IEEE Access*, *11*, 38805–38818. https://doi.org/10.1109/ACCESS.2023.3268224

[60] de Silva, D., Mills, N., El-Ayoubi, M., Manic, M., & Alahakoon, D. (2023). ChatGPT and Generative AI Guidelines for Addressing Academic Integrity and Augmenting Pre-Existing Chatbots. *2023 IEEE International Conference on Industrial Technology (ICIT)*, 1–6. https://doi.org/10.1109/ICIT58465.2023.10143123

[61] Zhu, C., Sun, M., Luo, J., Li, T., & Wang, M. (2023). How to harness the potential of ChatGPT in education? *KNOWLEDGE MANAGEMENT & E-LEARNING-AN INTERNATIONAL JOURNAL*, *15*(2), 133–152. https://doi.org/10.34105/j.kmel.2023.15.008

[62] Ouh, E. L., Gan, B. K. S., Jin Shim, K., & Wlodkowski, S. (2023). ChatGPT, Can You Generate Solutions for My Coding Exercises? An Evaluation on Its Effectiveness in an Undergraduate Java Programming Course. *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, 54–60. https://doi.org/10.1145/3587102.3588794

[63] Joyner, D. A. (2023). ChatGPT in Education: Partner or Pariah? *XRDS*, *29*(3), 48–51. https://doi.org/10.1145/3589651

[64] Murgia, E., Abbasiantaeb, Z., Aliannejadi, M., Huibers, T., Landoni, M., & Pera, M. S. (2023). ChatGPT in the Classroom: A Preliminary Exploration on the Feasibility of Adapting ChatGPT to Support Children's Information Discovery. *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, 22–27. https://doi.org/10.1145/3563359.3597399

[65] Malinka, K., Peres\'{\i}ni, M., Firc, A., Hujnák, O., & Janus, F. (2023). On the Educational Impact of ChatGPT: Is Artificial Intelligence Ready to Obtain a University Degree?

Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1, 47–53. https://doi.org/10.1145/3587102.3588827

[66] Finnie-Ansley, J., Denny, P., Luxton-Reilly, A., Santos, E. A., Prather, J., & Becker, B. A. (2023). My AI Wants to Know If This Will Be on the Exam: Testing OpenAI's Codex on CS2 Programming Exercises. *Proceedings of the 25th Australasian Computing Education Conference*, 97–104. https://doi.org/10.1145/3576123.3576134

[67] Konecki, M., Konecki, M., & Biškupić, I. O. (2023). Using Artificial Intelligence in Higher Education. In U. J. M. B. Jovanovic J. Chounta I.-A. (Ed.), *International Conference on Computer Supported Education, CSEDU - Proceedings* (Vol. 2, pp. 136–141). Science and Technology Publications, Lda. https://doi.org/10.5220/0012039700003470

[68] Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, *15*(3), ep429. https://doi.org/10.30935/cedtech/13152

[69] Crawford, J., Cowling, M., Ashton-Hay, S., Kelder, J.-A., Middleton, R., & Wilson, G. S. (2023). Artificial Intelligence and Authorship Editor Policy: ChatGPT, Bard Bing AI, and beyond. *Journal of University Teaching and Learning Practice*, *20*(5). https://doi.org/10.53761/1.20.5.01

[70] Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, *4*. https://doi.org/10.1016/j.caeai.2023.100147

[71] Ellaway, R. H., & Tolsgaard, M. (2023). Artificial scholarship: LLMs in health professions education research. *Advances in Health Sciences Education*. https://doi.org/10.1007/s10459-023-10257-4

[72] Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching*, *6*(1), 364–389. https://doi.org/10.37074/jalt.2023.6.1.23

[73] Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., & Chocksathaporn, P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching*, *6*(1), 64–74. https://doi.org/10.37074/jalt.2023.6.1.32

[74] Khademi, A. (2023). Can ChatGPT and Bard generate aligned assessment items? A reliability analysis against human performance. *Journal of Applied Learning and Teaching*, *6*(1), 75–80. https://doi.org/10.37074/jalt.2023.6.1.28

[75] Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*. https://doi.org/10.1016/j.ijinfomgt.2023.102642

[76] Keiper, M. C., Fried, G., Lupinek, J., & Nordstrom, H. (2023). Artificial intelligence in sport management education: Playing the AI game with ChatGPT. *Journal of Hospitality, Leisure, Sport and Tourism Education, 33*. https://doi.org/10.1016/j.jhlste.2023.100456

[77] Ajevski, M., Barker, K., Gilbert, A., Hardie, L., & Ryan, F. (2023). ChatGPT and the future of legal education and practice. *The Law Teacher*, *57*(3), 352–364. https://doi.org/10.1080/03069400.2023.2207426

[78] Bauer, E., Greisel, M., Kuznetsov, I., Berndt, M., Kollar, I., Dresel, M., Fischer, M. R., & Fischer, F. (2023). Using natural language processing to support peer-feedback in the age of artificial intelligence: A cross-disciplinary framework and a research agenda. *British Journal of Educational Technology*, *54*(5), 1222–1245. https://doi.org/10.1111/bjet.13336

[79] Breeding, T., Martinez, B., Patel, H., Nasef, H., Arif, H., Nakayama, D., & Elkbuli, A. (2024). The Utilization of ChatGPT in Reshaping Future Medical Education and Learning Perspectives: A Curse or a Blessing? *The American Surgeon™*, *90*(4), 560–566. https://doi.org/10.1177/00031348231180950

[80] Chatterjee, J., & Dethlefs, N. (2023). This new conversational AI model can be your friend, philosopher, and guide ... and even your worst enemy. *Patterns*, *4*(1), 100676. https://doi.org/10.1016/j.patter.2022.100676

[81] Chaudhry, I. S., Sarwary, S. A. M., el Refae, G. A., & Chabchoub, H. (2023). Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT — A Case Study. *Cogent Education*, *10*(1). https://doi.org/10.1080/2331186X.2023.2210461

[82] Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*, *32*(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

[83] Corsello, A., & Santangelo, A. (2023). May Artificial Intelligence Influence Future Pediatric Research?—The Case of ChatGPT. *Children*, *10*(4), 757. https://doi.org/10.3390/children10040757

[84] Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

[85] Crawford, J., Cowling, M., & Allen, K.-A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching and Learning Practice*, *20*(3). https://doi.org/10.53761/1.20.3.02

[86] Das, D., Kumar, N., Longjam, L. A., Sinha, R., Deb Roy, A., Mondal, H., & Gupta, P. (2023). Assessing the Capability of ChatGPT in Answering First- and Second-Order Knowledge Questions on Microbiology as per Competency-Based Medical Education Curriculum. *Cureus*, *15*(3). https://doi.org/10.7759/cureus.36034

[87] Day, T. (2023). A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT. *The Professional Geographer*, 1–4. https://doi.org/10.1080/00330124.2023.2190373

[88] Eggmann, F., Weiger, R., Zitzmann, N. U., & Blatz, M. B. (2023). Implications of large language models such as <scp>ChatGPT</scp> for dental medicine. *Journal of Esthetic and Restorative Dentistry*. https://doi.org/10.1111/jerd.13046

[89] Emenike, M. E., & Emenike, B. U. (2023). Was This Title Generated by ChatGPT? Considerations for Artificial Intelligence Text-Generation Software Programs for Chemists and Chemistry Educators. *Journal of Chemical Education*, 100(4), 1413–1418. https://doi.org/10.1021/acs.jchemed.3c00063

[90] Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 1–15. https://doi.org/10.1080/14703297.2023.2195846

[91] Fergus, S., Botha, M., & Ostovar, M. (2023). Evaluating Academic Answers Generated Using ChatGPT. *Journal of Chemical Education*, 100(4), 1672–1675. https://doi.org/10.1021/acs.jchemed.3c00087

[92] Ghosh, A., & Bir, A. (2023). Evaluating ChatGPT's Ability to Solve Higher-Order Questions on the Competency-Based Medical Education Curriculum in Medical Biochemistry. *CUREUS JOURNAL OF MEDICAL SCIENCE*, 15(4). https://doi.org/10.7759/cureus.37023

[93] Hegde, A., Srinivasan, S., & Menon, G. (2023). Extraventricular Neurocytoma of the Posterior Fossa: A Case Report Written by ChatGPT. *Cureus*, 15(3). https://doi.org/10.7759/cureus.35850

[94] Hsu, Y.-C., & Ching, Y.-H. (2023). Generative Artificial Intelligence in Education, Part One: the Dynamic Frontier. *TechTrends*, 67(4), 603–607. https://doi.org/10.1007/s11528-023-00863-9

[95] Iskender, A. (2023). Holy or Unholy? Interview with Open AI's ChatGPT. *European Journal of Tourism Research*, 34, 3414. https://doi.org/10.54055/ejtr.v34i.3169

[96] Ivanov, S., & Soliman, M. (2023). Game of algorithms: ChatGPT implications for the future of tourism education and research. *Journal of Tourism Futures*, 9(2), 214–221. https://doi.org/10.1108/JTF-02-2023-0038

[97] Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11834-1

[98] Johinke, R., Cummings, R., & di Lauro, F. (2023). Reclaiming the technology of higher education for teaching digital writing in a post—pandemic world. *Journal of University Teaching and Learning Practice*, 20(2). https://doi.org/10.53761/1.20.02.01

[99] Khan, R. A., Jawaid, M., Khan, A. R., & Sajjad, M. (2023). ChatGPT - Reshaping medical education and clinical management. *Pakistan Journal of Medical Sciences*, 39(2), 605–607. https://doi.org/10.12669/pjms.39.2.7653

[100] Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory physics course? *Physical Review Physics Education Research*, 19(1), 010132. https://doi.org/10.1103/PhysRevPhysEducRes.19.010132

[101] Kumah-Crystal, Y., Mankowitz, S., Embi, P., & Lehmann, C. U. (2023). ChatGPT and the clinical informatics board examination: the end of unproctored maintenance of certification? *Journal of the American Medical Informatics Association*, 30(9), 1558–1560. https://doi.org/10.1093/jamia/ocad104

[102] Lappalainen, Y., & Narayanan, N. (2023). Aisha: A Custom AI Library Chatbot Using the ChatGPT API. *Journal of Web Librarianship*, 17(3), 37–58. https://doi.org/10.1080/19322909.2023.2221477

[103] Lee, J. Y. (2023). Can an artificial intelligence chatbot be the author of a scholarly article? *Journal of Educational Evaluation for Health Professions*, 20, 6. https://doi.org/10.3352/jeehp.2023.20.6

[104] Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), 410. https://doi.org/10.3390/educsci13040410

[105] Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in tertiary education. *Australasian Journal of Educational Technology,* 39(1), 1–8. https://doi.org/10.14742/ajet.8695

[106] Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). <scp>ChatGPT</scp> and a new academic reality: <scp>Artificial Intelligence-written</scp> research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, 74(5), 570–581. https://doi.org/10.1002/asi.24750

[107] Luo, W., He, H., Liu, J., Berson, I. R., Berson, M. J., Zhou, Y., & Li, H. (2023). Aladdin's Genie or Pandora's Box for Early Childhood Education? Experts Chat on the Roles, Challenges, and Developments of ChatGPT. *Early Education and Development,* 1–18. https://doi.org/10.1080/10409289.2023.2214181

[108] Lyu, Q., Tan, J., Zapadka, M. E., Ponnatapura, J., Niu, C., Myers, K. J., Wang, G., & Whitlow, C. T. (2023). Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. *Visual Computing for Industry, Biomedicine, and Art*, 6(1), 9. https://doi.org/10.1186/s42492-023-00136-5

[109] Mohamed, A. M. (2023). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: perceptions of EFL Faculty Members. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11917-z

[110] Naidu, K., & Sevnarayan, K. (2023). ChatGPT: An ever-increasing encroachment of artificial intelligence in online assessment in distance education. *Online Journal of Communication and Media Technologies*, 13(3), e202336. https://doi.org/10.30935/ojcmt/13291

[111] Oh, N., Choi, G.-S., & Lee, W. Y. (2023). ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. *Annals of Surgical Treatment and Research*, 104(5), 269. https://doi.org/10.4174/astr.2023.104.5.269

[112] Perkins, M. (2023). Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2). https://doi.org/10.53761/1.20.02.07

[113] Polonsky, M. J., & Rotman, J. D. (2023). Should Artificial Intelligent Agents be Your Co-author? Arguments in Favour, Informed by ChatGPT. *Australasian Marketing Journal*, 31(2), 91–96. https://doi.org/10.1177/14413582231167882

[114] Rahimi, F., & Talebi Bezmin Abadi, A. (2023). Passive Contribution of ChatGPT to Scientific Papers. *Annals of Biomedical Engineering*. https://doi.org/10.1007/s10439-023-03260-8

[115] Rahman, Md. M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences*, 13(9), 5783. https://doi.org/10.3390/app13095783

[116] Ramos, L., Marquez, R., & Francklin, F. (2023). AI's next frontier: The rise of ChatGPT and its implications on society, industry, and scientific research. *CIENCIA E INGENIERIA*, 44(1), 131–148.

[117] Sánchez-Ruiz, L. M., Moll-López, S., Nuñez-Pérez, A., Moraño-Fernández, J. A., & Vega-Fleitas, E. (2023). ChatGPT Challenges Blended Learning Methodologies in Engineering Education: A Case Study in Mathematics. *Applied Sciences*, 13(10), 6039. https://doi.org/10.3390/app13106039

[118] Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare*, 11(6), 887. https://doi.org/10.3390/healthcare11060887

[119] Sedaghat, S. (2023). Early applications of ChatGPT in medical practice, education and research. *Clinical Medicine*, 23(3), 278–279. https://doi.org/10.7861/clinmed.2023-0078

[120] Seetharaman, R. (2023). Revolutionizing Medical Education: Can ChatGPT Boost Subjective Learning and Expression? *Journal of Medical Systems*, 47(1), 61. https://doi.org/10.1007/s10916-023-01957-w

[121] Teixeira da Silva, J. A. (2023). Is ChatGPT a valid author? *Nurse Education in Practice*, 68, 103600. https://doi.org/10.1016/j.nepr.2023.103600

[122] Sevgi, U. T., Erol, G., Doğruel, Y., Sönmez, O. F., Tubbs, R. S., & Güngor, A. (2023). The role of an open artificial intelligence platform in modern neurosurgical education: a preliminary study. *Neurosurgical Review*, 46(1), 86. https://doi.org/10.1007/s10143-023-01998-2

[123] Sharma, S., Pajai, S., Prasad, R., Wanjari, M. B., Munjewar, P. K., Sharma, R., & Pathade, A. (2023). A Critical Review of ChatGPT as a Potential Substitute for Diabetes Educators. *Cureus*, 15(5). https://doi.org/10.7759/cureus.38380

[124] Shaw, D., Morfeld, P., & Erren, T. (2023). The (mis)use of ChatGPT in science and education. *EMBO Reports*, 24(7). https://doi.org/10.15252/embr.202357501

[125] Skavronskaya, L., Hadinejad, A. (Hana), & Cotterell, D. (2023). Reversing the threat of artificial intelligence to opportunity: a discussion of ChatGPT in tourism education. *Journal*

of Teaching in Travel & Tourism, 23(2), 253–258. https://doi.org/10.1080/15313220.2023.2196658

[126] Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: an autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20(1), 35. https://doi.org/10.1186/s41239-023-00404-7

[127] Zumsteg, J. M., & Junn, C. (2023). Will ChatGPT Match to Your Program? *American Journal of Physical Medicine & Rehabilitation*, 102(6), 545–547. https://doi.org/10.1097/PHM.0000000000002238

[128] Tafferner, Z., Balázs, B., Krammer, O., & Géczy, A. (2023). Can ChatGPT Help in Electronics Research and Development? A Case Study with Applied Sensors. *Sensors*, 23(10), 4879. https://doi.org/10.3390/s23104879

[129] Tang, G. (2023). Academic journals cannot simply require authors to declare that they used ChatGPT. *Irish Journal of Medical Science* (1971 -). https://doi.org/10.1007/s11845-023-03374-x

[130] Thurzo, A., Strunga, M., Urban, R., Surovková, J., & Afrashtehfar, K. I. (2023). Impact of Artificial Intelligence on Dental Education: A Review and Guide for Curriculum Update. *Education Sciences*, 13(2), 150. https://doi.org/10.3390/educsci13020150

[131] Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15. https://doi.org/10.1186/s40561-023-00237-x

[132] Tsang, R. (2023). Practical Applications of ChatGPT in Undergraduate Medical Education. *Journal of Medical Education and Curricular Development*, 10, 238212052311784. https://doi.org/10.1177/23821205231178449

[133] Uddin, S. M. J., Albert, A., Ovid, A., & Alsharef, A. (2023). Leveraging ChatGPT to Aid Construction Hazard Recognition and Support Safety Education and Training. *Sustainability*, 15(9), 7121. https://doi.org/10.3390/su15097121

[134] Vartiainen, H., & Tedre, M. (2023). Using artificial intelligence in craft education: crafting with text-to-image generative models. *Digital Creativity*, 34(1), 1–21. https://doi.org/10.1080/14626268.2023.2174557

[135] Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11742-4

Appendix A

Table A.1. Educational Level Per Article.

Educational level	Total	Articles
University	95	[1] [3] [5] [7] [8] [10] [11] [12] [16] [17] [18] [19] [21] [23] [29] [36] [39] [42] [45] [46] [47] [48] [50] [53] [54] [55] [56] [58] [59] [62] [65] [66] [67] [38] [69] [70] [71] [72] [73] [75] [76] [77] [78] [79] [81] [83] [84] [85] [86] [87] [43] [88] [89] [90] [91] [92] [93] [95] [96] [98] [99] [100] [101] [102] [103] [104] [105] [106] [108] [109] [110] [111] [112] [113] [114] [115] [116] [117] [118] [119] [120] [122] [123] [124] [125] [126] [127] [128] [129] [130] [132] [133] [134] [135]
General	21	[4] [2] [6] [14] [20] [22] [35] [49] [51] [57] [60] [61] [63] [68] [74] [80] [82] [94] [97] [121] [131]
Elementary School	4	[9] [13] [15] [64]
Kindergarten	3	[9] [13] [107]
Middle School	1	[52]
Highschool	0	

Table A.2. Higher Education Sub-level Per Article.

Higher Education Sub-level	Total	Articles
Introductory / Undergraduate	16	[16] [36] [39] [47] [53] [56] [58] [62] [70] [76] [91] [92] [100] [130] [132] [135]
Advanced	4	[58] [59] [79] [130]

Table A.3. Higher Educational Area of Study Per Article.

Area of Study	Total	Articles
Medicine & Health	27	[18] [19] [21] [23] [29] [46] [47] [71] [79] [83] [86] [88] [91] [92] [93] [99] [101] [108] [111] [118] [119] [120] [122] [123] [127] [130] [132]
Computer Science & Engineering	14	[1] [11] [39] [53] [56] [58] [59] [62] [65] [66] [70] [115] [117] [128]
Research	8	[45] [46] [103] [106] [113] [114] [116] [118]
English & Linguistics	5	[54] [58] [109] [112] [135]
Design & Digital Media	2	[55] [98]
Tourism	3	[95] [96] [125]
Journalism & Media	1	[7]
Economics	1	[48]
Mathematics	1	[58]
Sport Management	1	[76]
Legal Studies	1	[77]
Physics	1	[100]
Tertiary Education	1	[105]
Craft	1	[134]

Appendix B

Table B.1. GenAI Tools Mentioned Per Article.

Technologies	Total	Article
ChatGPT	114	[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [13] [12] [14] [16] [17] [18] [19] [21] [23] [29] [35] [36] [38] [39] [42] [43] [45] [46] [47] [48] [49] [50] [52] [53] [54] [55] [56] [57] [58] [59] [60] [61] [62] [63] [64] [65] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80] [81] [82] [83] [84] [85] [86] [87] [88] [89] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100] [101] [102] [103] [104] [105] [106] [107] [108] [109] [110] [111] [112] [113] [114] [115] [116] [117] [118] [119] [120] [121] [122] [123] [124] [125] [126] [127] [128] [129] [130] [131] [132] [133]
GPT-3	35	[3] [5] [6] [7] [9] [18] [39] [43] [49] [52] [55] [56] [57] [60] [66] [67] [70] [75] [77] [80] [84] [89] [95] [98] [100] [106] [108] [112] [113] [115] [121] [125] [128] [130] [131]
GPT-4	29	[5] [8] [9] [13] [14] [39] [49] [56] [57] [60] [61] [62] [68] [70] [72] [77] [78] [94] [101] [102] [104] [105] [107] [108] [111] [117] [126] [128] [132]
GPT-3.5	20	[5] [6] [8] [14] [16] [56] [62] [72] [73] [77] [101] [102] [104] [111] [112] [117] [126] [130] [132] [135]
Bard	17	[2] [8] [14] [38] [49] [64] [67] [71] [72] [74] [77] [90] [91] [94] [102] [39] [121]
Bing AI	13	[8] [14] [38] [39] [47] [49] [64] [67] [69] [71] [72] [91] [94]
DALL-E	13	[2] [7] [10] [11] [14] [23] [65] [75] [94] [106] [122] [130] [134]
Grammarly	12	[55] [68] [69] [71] [75] [77] [85] [87] [95] [112] [125] [135]
BERT	9	[39] [43] [56] [57] [60] [75] [106] [108] [112]
GPT-2	9	[39] [43] [51] [56] [57] [70] [100] [112] [131]
Midjourney	8	[4] [9] [72] [82] [94] [102] [130] [134]
Turnitin	8	[48] [67] [69] [72] [77] [81] [100] [135]
DALL-E 2	7	[4] [9] [57] [67] [82] [98] [102]
Alexa	6	[48] [52] [72] [99] [102] [116]
GPTZero	6	[48] [50] [65] [67] [81] [90]
Stable Diffusion	6	[4] [7] [9] [23] [102] [130]
Copilot	5	[3] [35] [62] [65] [66]
Jasper AI	5	[69] [72] [99] [117] [125]
Siri	5	[48] [52] [72] [99] [116]
Codex	4	[3] [6] [62] [67]
GPT-1	4	[56] [57] [70] [39]

Table B.1. Continued from the previous column.

Technologies	Total	Article
XLNet	4	[39] [56] [106] [108]
BioGPT	3	[35] [72] [111]
Copyleaks	3	[50] [67] [81]
DeepL	3	[65] [75] [100]
Ernie	3	[8] [57] [72]
Translate	3	[65] [100] [135]
Watson	3	[72] [75] [102]
AlphaCode	2	[62] [66]
Alpa	2	[98] [100]
Alpha	2	[39] [117]
BLOOM	2	[56] [98]
Chatsonic	2	[53] [72]
Cortana	2	[48] [102]
CopyAI	2	[72] [117]
DialoGPT	2	[69] [99]
ELMo	2	[57] [60]
Eleven Labs	2	[4] [72]
LLaMA	2	[102] [111]
GPT-5	2	[49] [56]
GPT-output detector	2	[58] [130]
InstructGPT	2	[6] [80]
ProWritingAid	2	[43] [125]
Quillbot	2	[68] [135]
QnABot	2	[72] [75]
Rytr	2	[72] [117]
Writer AI	2	[67] [130]
ZeroGPT	2	[75] [100]
Aigiarism	1	[50]
AlphaGo	1	[95]
Alpaca	1	[102]
Amper Music	1	[94]
ALEKS	1	[68]
Autolab	1	[68]
AutoGradr	1	[68]
Blenderbot 3	1	[98]
Blender	1	[39]
BloombergGPT	1	[102]
Blockly	1	[1]
Chanel	1	[4]
Chinchilla	1	[102]
CodeWhisperer	1	[66]
CopyCatch	1	[77]
CoVe	1	[60]
Crossplag AI	1	[67]
Copysmith	1	[117]

Table B.1. Continued from the previous column.

Technologies	Total	Article
ChibiAI	1	[72]
Cerebras-GPT	1	[102]
DataRobot	1	[113]
DetectGPT	1	[48]
Dialogflow	1	[104]
D-ID	1	[72]
Educational copilot	1	[52]
EconBot	1	[67]
Elicit	1	[98]
Leonardo.ai	1	[94]
FaceApp	1	[65]
First Order Motion Model	1	[20]
Fermat	1	[98]
Google Assistant	1	[48]
GPT-NeoX	1	[102]
Gradescope	1	[68]
Google AI	1	[29]
IBM watson	1	[52]
InstaText	1	[68]
InVideo	1	[94]
iThenticate	1	[72]
Lensa	1	[7]
LaMDA	1	[102]
LanguageTool	1	[125]
Kafkai	1	[117]
Knewton	1	[68]
Koala	1	[102]
Make-a-video	1	[57]
Magic Write	1	[4]
Meena	1	[39]
MuseNet	1	[94]
MusicLM	1	[98]
Nvidia AI	1	[98]
OpenAI Text Classifier	1	[48]
Originality-ai	1	[67]
Otter.ai	1	[99]
PaLM 2	1	[102]
PanGu-Sigma	1	[102]
Perusall	1	[112]
Repl.it	1	[68]
rTutor.ai	1	[75]
Research Rabbit	1	[75]
Replika	1	[99]
RoBERTa	1	[106]
Samsung Bixby	1	[52]

Table B.1. Continued from the previous column.

Technologies	Total	Article
Scratch	1	[1]
Sapling AI	1	[67]
SenseChat robot	1	[57]
Smart Sparrow	1	[68]
Socratic	1	[69]
Stability AI	1	[98]
StableLM	1	[102]
Tabnine	1	[66]
Tay	1	[102]
TruthGPT	1	[72]
Unicheck	1	[67]
VIRTA	1	[67]
Vicuna	1	[102]
Wav2Lip	1	[20]
WaveNet	1	[20]
Writesonic	1	[117]
WordTune	1	[112]
Wolfram	1	[117]
Xiaolce	1	[39]
YouChat	1	[64]
You.com	1	[72]