

Article

# Large Language Models (LLMs) in Engineering Education: A Systematic Review and Suggestions for Practical Adoption

Stefano Filippi \*  and Barbara Motyl 

Polytechnic Department of Engineering and Architecture (DPIA), University of Udine, 33100 Udine, Italy; motyl@uniud.it

\* Correspondence: filippi@uniud.it; Tel.: +39-0432-558289

**Abstract:** The use of large language models (LLMs) is now spreading in several areas of research and development. This work is concerned with systematically reviewing LLMs' involvement in engineering education. Starting from a general research question, two queries were used to select 370 papers from the literature. Filtering them through several inclusion/exclusion criteria led to the selection of 20 papers. These were investigated based on eight dimensions to identify areas of engineering disciplines that involve LLMs, where they are most present, how this involvement takes place, and which LLM-based tools are used, if any. Addressing these key issues allowed three more specific research questions to be answered, offering a clear overview of the current involvement of LLMs in engineering education. The research outcomes provide insights into the potential and challenges of LLMs in transforming engineering education, contributing to its responsible and effective future implementation. This review's outcomes could help address the best ways to involve LLMs in engineering education activities and measure their effectiveness as time progresses. For this reason, this study addresses suggestions on how to improve activities in engineering education. The systematic review on which this research is based conforms to the rules of the current literature regarding inclusion/exclusion criteria and quality assessments in order to make the results as objective as possible and easily replicable.

**Keywords:** engineering education; large language models—LLMs; LLM-based tools; systematic review; PRISMA



**Citation:** Filippi, S.; Motyl, B. Large Language Models (LLMs) in Engineering Education: A Systematic Review and Suggestions for Practical Adoption. *Information* **2024**, *15*, 345. <https://doi.org/10.3390/info15060345>

Academic Editors: Jui-Long Hung, Xu Du and Juan Yang

Received: 22 April 2024

Revised: 14 May 2024

Accepted: 22 May 2024

Published: 12 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Large language models (LLMs) are text-generating AI systems. Their use is spreading in several areas of research and development. The literature about real implementations of LLMs in everyday activities started to appear in 2019 following the availability of the first version of ChatGPT (GPT-2) [1,2]. However, most of the publications and systematic reviews about LLMs and LLM-based tools started to be available in 2023, as witnessed in databases such as SCOPUS and WOS [3–8] (PRISMA3).

The advent of LLMs can be considered as a sort of revolution at both the educational and professional levels. This is why the correct approach to them in university courses is mandatory to ensure that students, as future engineering professionals, are able to address these new technologies properly and conscientiously in order to give research, development, and innovation an effective boost. As engineering educators at different levels (undergraduate, postgraduate, etc.), we have been introducing LLMs in our courses since 2021 from both theoretical and practical points of view [9,10]. Although the effective help provided by LLMs has been evident since the beginning, the approach was initially empirical since adoption guidelines or best practices were not available. In the last three years, meaningful information has been generated and made available. Although the landscape of the literature dealing with LLMs offers the most-cited systematic reviews covering domains such as medicine [5,11], industry or robotics [7,12], and education [8,13],

it seems that reviews focusing on LLMs' involvement in undergraduate/postgraduate engineering education have not yet appeared in the literature. Furthermore, suggestions and guidelines for best practices of involving LLMs in everyday education activities are still missing. All of this suggests that we should analyze the literature while focusing only on this domain. The investigation spread across several dimensions, from areas that involve LLMs to the engineering disciplines in which they are most present, as well as how this involvement takes place and which LLM-based tools are used, if any.

Under these premises, the initial research question—RQ0—was defined as follows.

**RQ0.** *What is the current status of LLMs' involvement in engineering education?*

The term “involvement” was carefully chosen, as was the generality of RQ0. This aimed to capture the presence of LLMs in the engineering education activities reported in the literature as much as possible. Using terms such as “implementation”, “adoption”, or something else or defining a more focused question would have unnecessarily narrowed the scope of the review a priori. We decided to attempt to answer RQ0 through a focused, systematic review [14]. Moreover, given the availability of guidelines and checklists for making systematic reviews as rigorous and replicable as possible, we mapped this systematic review to the PRISMA checklist [15]. For this reason, labels with “PRISMA<sub>n</sub>” appear throughout the article. They are references to the items of the PRISMA checklist, as reported in Appendix A.

Once we defined the general research question (RQ0), we built queries for the selection of articles from the literature, collected them, developed and applied inclusion/exclusion criteria, read the articles, and analyzed the data. The results provide a clear overview of the current involvement of LLMs in engineering education. These results, in turn, will help address structured ways to involve LLMs and measure their effectiveness as time progresses (PRISMA4). The Discussion section also deals with practical suggestions for involving LLMs in engineering education activities.

This article opens with the Materials and Methods section, which describes the research background and approach. The activities conducted as part of the systematic review are described in the next section. Then, the Results section reports the review's outcomes, and the following Discussion section analyzes them critically and offers suggestions about the use of the research results in undergraduate and postgraduate engineering courses. The conclusion, which also contains some research perspectives, closes the study.

## 2. Materials and Methods

Regarding LLMs and LLM-based tools, the widespread use and ubiquitous integration of artificial intelligence (AI) is now commonplace in professional, educational, and everyday life. Specifically, given the focus of this research, it is appropriate to delve into aspects related to generative AI (GenAI). GenAI is an artificial intelligence technology that generates content in response to prompts within natural-language conversational interfaces. In contrast to systems that merely curate existing webpages, GenAI produces entirely new content, encompassing various forms of representation of human thought, including natural-language texts, images, videos, music, and software code. It undergoes training using data from webpages, social media conversations, and online media, wherein it analyzes statistical distributions of words, pixels, or other elements to identify and replicate common patterns, such as word associations [2–4].

The technologies underlying GenAI belong to the machine learning (ML) family of artificial intelligence. ML uses algorithms to continuously improve performance through data. A major contributor to recent advancements in AI is a type of ML known as artificial neural networks (ANNs). ANNs are inspired by the human brain and the synaptic connections between its neurons. There are several types of ANNs. Text-generating AI uses a special type of ANN called a general-purpose transformer. Text-generating AI systems are commonly referred to as large language models (LLMs). Within this category, a specific type of LLM known as a generative pre-trained transformer (GPT) plays a central

role. This is the origin of the GPT acronym in the name ChatGPT. ChatGPT, specifically, is built upon GPT-3, a product of OpenAI and the third generation in their GPT series. The first GPT model was launched in 2018, and the latest iteration, GPT-4, was released in March 2023 [1,2].

Several LLMs based on transformer architectures similar to ChatGPT are currently available. Notable examples include Gemini (formerly known as Bard) by Google [16], Alpaca by Stanford University Center for Research on Foundation Models (CRFM) [17], and Elicit by Ought [18]. These models are often pre-trained on large datasets and tuned for specific tasks. Each has its own strengths and weaknesses that make it better suited for specific applications or use cases.

### 3. Systematic Review

As mentioned before, the planning of this systematic review occurred by following the PRISMA checklist. This helped define the scope of the systematic review, identify key research questions, establish inclusion and exclusion criteria, process data, and formulate the outcomes. This review is not registered (PRISMA 24a). Regarding the assessment of the risk of bias, the considerations leading to both the first search and the subsequent adoption of the exclusion criteria were objective and strong enough to keep the risk of bias as low as possible (PRISMA11). Regarding the protocol used, the precise references to the PRISMA checklist occurring in the different sections of the study highlight that the research occurred rigorously and made it replicable by other researchers and practitioners (PRISMA 24b, PRISMA24c).

Two researchers took part in the review activities. They screened the records independently using Microsoft Excel spreadsheets for data analysis. At the end of their work, they compared the results and generated the research outcomes (PRISMA9, PRISMA13a).

The selection/evaluation of articles occurred as follows. Two databases, SCOPUS and IEEE X, were searched on 6 March 2024. The SCOPUS database was searched using the following query:

```
“(TITLE-ABS-KEY ((chatgpt OR bard OR gemini OR “large language models” OR llms) AND engineering AND education) AND LANGUAGE (english))”
```

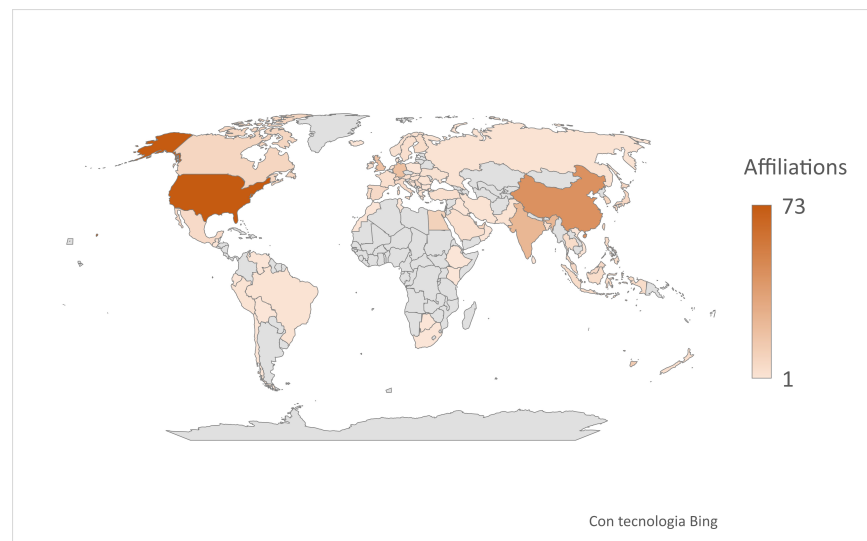
This query returned 202 papers. The IEEE X database was searched using the following query:

```
“(“All Metadata”:ChatGPT OR “All Metadata”:Bard OR “All Metadata”:GEMINI OR “All Metadata”:Llms OR “All Metadata”:“large language models”) AND (“All Metadata”:engineering) AND (“All Metadata”:education)”
```

In this case, the results consisted of 168 papers. Thus, the total number of papers selected from the two databases was 370 (PRISMA6; PRISMA7). By eliminating 39 duplicates, the number of papers dropped to 331, which was the starting point for the following activities. These papers were numbered in order to code them, and this coding is used hereafter.

Before progressing to the next stage—the content analysis—the first exclusion criteria were implemented (PRISMA5). Initially, from the pool of 331 papers, those authored prior to 2018 were excluded, as this was the year of the first appearance of LLM-based tools such as ChatGPT. This step reduced the number of papers to 319. Additionally, papers categorized as “conference reviews”, “books”, and “editorials” were further removed, resulting in a total of 306 papers for subsequent analysis.

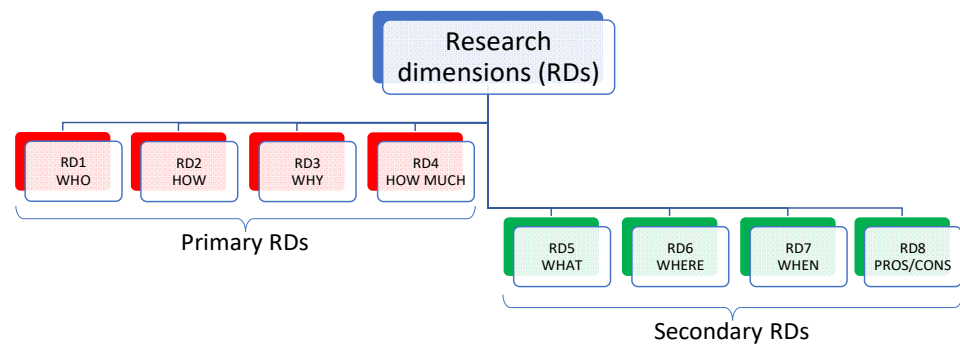
A first analysis was performed on these 306 papers. It regarded the countries to which the authors belonged. The aim was to gain an insight into the geographical distribution of the involvement of LLMs in engineering education activities worldwide at the time of the database query. The results showed the prevalence of the USA (73 affiliations), followed by China (44), India (25), Germany (20), and the United Kingdom (19). Many other countries followed, showing that the coverage was quite equally distributed. Figure 1 shows the worldwide coverage.



**Figure 1.** Worldwide coverage of the affiliations of the authors of papers in some way related to the involvement of LLMs in engineering education at the time of the database query.

Next, starting from these 306 papers, there was an initial screening through the reading of titles, abstracts, and keywords (authors’ keywords, indexed keywords, or IEEE terms). This reading led to the definition of the second exclusion criterion. Papers that were not deemed to be focused on the theme posed in RQ0, namely, the use of LLMs in engineering education, were discarded, thus reducing the number of papers to 151 (PRISMA16b).

The reading of the titles, abstracts, and keywords of the 151 papers helped refine the initial research question by distributing the interest over several topics, which are called research dimensions (RDs) here. Figure 2 shows the eight RDs considered in the research.



**Figure 2.** Research dimensions (RDs) highlighted during the analysis of the titles, abstracts, and keywords of the 151 selected papers.

These eight RDs, which were as orthogonal as possible to each other, had the following peculiarities.

- RD1—WHO. This refers to the actors involving LLMs in engineering education activities. Examples thereof are students, educators, or any other stakeholders.
- RD2—HOW. This dimension represents the ways in which LLMs are involved. Examples thereof—grouped as reference activities—are tests of use, case studies, use method proposals, etc.
- RD3—WHY. This describes the reasons/goals for the involvement of LLMs. Examples thereof span from the enhancement of understanding to the enrichment of problem solving, teaching improvement, etc.

- RD4—HOW MUCH. Papers could report qualitative/quantitative evaluations of the involvement of LLMs in tasks or activities in engineering education. Examples thereof are a qualitatively measured low impact, a quantitatively measured high impact, etc.
- RD5—WHAT. Since more LLM-based tools are made available day by day, this dimension allows the description of those that are involved paper by paper, if any. Examples thereof are ChatGPT, Bard/Gemini, etc.
- RD6—WHERE. This dimension represents the domains of engineering education in which the involvement of LLMs takes place. Examples thereof are software engineering, mechanical engineering, chemical engineering, etc.
- RD7—WHEN. It is important to highlight the moment of the educative path at which the involvement of LLMs takes place. This dimension allows this to be expressed. Examples thereof are undergraduate courses, postgraduate courses, etc.
- RD8—PROS/CONS. Some papers are quite clear about the advantages and drawbacks of the involvement of LLMs. Examples of PROS are enhanced understanding, adoption of real-world examples and practical applications, etc. Examples of CONS are confusing and contradictory answers, inaccuracies in responses, ethical concerns, etc.

As a first important consequence of the definition of the RDs, they allow the general research question proposed in the introduction (RQ0) to be refined. The RDs could be logically combined to obtain research questions whose answers would better represent the state-of-the-art involvement of LLMs in engineering education. Three research questions that were more precise and focused were the result of these considerations. They were developed by paying attention to the mixing of “primary” dimensions (RD1 to RD4) with “secondary” dimensions (RD5 to RD8) (see Figure 2). The reason for this classification will be made clear in the following.

The first new research question (RQ1) investigated the interactions between people and LLMs. This RQ was based on RD1—WHO, RD2—HOW, RD7—WHEN, and RD8—PROS/CONS. RQ1 was the following:

**RQ1.** *Are the roles and duties of people clear regarding the involvement of LLMs in engineering education?*

The second new research question (RQ2) referred to the engineering domains of the involvement of LLMs and their possible influences on the modalities of this involvement. RQ2 was based on RD2—HOW, RD6—WHERE, and RD7—WHEN.

**RQ2.** *Is there evidence of relationships between engineering disciplines and the ways that LLMs are involved in related educational activities?*

Finally, the third new research question (RQ3) focused on LLM-based tools. It dealt with possible suggestions for their adoption in educational activities. RQ3 was based on RD4—HOW MUCH, RD5—WHAT, RD7—WHEN, and RD8—PROS/CONS.

**RQ3.** *Can clear indications of which LLM-based tools should be involved in order to improve the effectiveness of education activities and impact measurements be obtained?*

These new RQs will help formulate suggestions for the improvement of current educational activities.

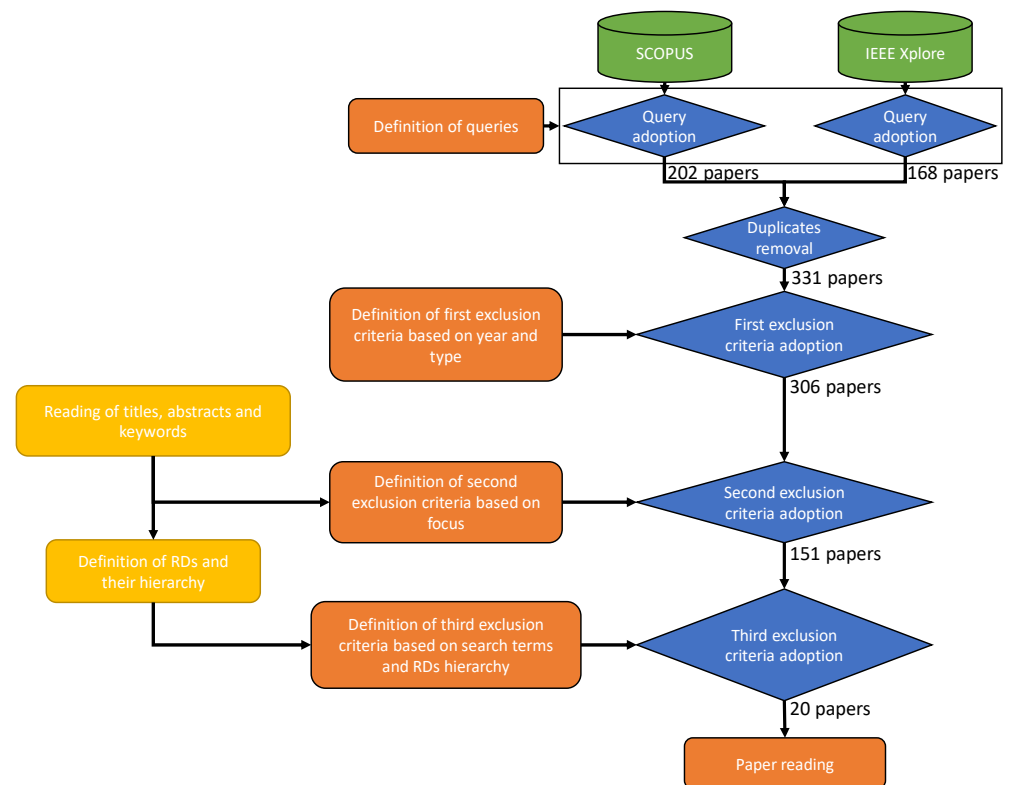
Before starting to read the full text of the papers, a third set of exclusion criteria was implemented in order to focus on specific topics from time to time (PRISMA5; PRISMA8). The first criterion referred to the relationships among the search terms used for the formulation of the queries. It aimed to exclude papers where the terms appeared in the title, abstract, and/or keywords, but their meaning did not belong to the research scope. For example, both the terms “ChatGPT” and “engineering education” appeared in one of the papers, but the focus of the work was on the ways of detecting and managing cases of plagiarism, and this topic was not covered here; therefore, that paper was excluded. Moreover, in order to filter the selected papers to focus on the research objective even more closely, a hierarchy was defined for the RDs. The WHO (RD1), HOW (RD2), WHY (RD3), and

HOW MUCH (RD4) dimensions were considered primary. This decision was based on the authors' experience as researchers and educators, as well as on precise considerations about the eight RDs. For example, knowing the "who" (RD1) of involving LLMs in education or how (RD2) this happens was considered fundamental in order to understand the state of the art and list practical suggestions for improving educational activities. On the contrary, the other four RDs, WHAT (RD5), WHERE (RD6), WHEN (RD7), and PROS/CONS (RD8), were considered secondary. For example, knowing where (RD6) the involvement occurs and when (RD7) it happens is considered important information but not at the same level as the first four RDs. Consequently, papers that did not present clear references to the four main dimensions were excluded. Once these exclusion criteria were applied, 20 papers remained from the 151 papers. Table 1 contains the titles of these papers, along with the numerical codes used to represent them throughout this research.

**Table 1.** The 20 selected papers with their reference codes.

Title	Code
Roadmap for software engineering education using ChatGPT [19]	2
ChatGPT as a full-stack web developer—Early Results [20]	4
Students' use of large language models in engineering education: a case study on technology acceptance, perceptions, efficacy, and detection chances [21]	20
Designing a worksheet for using ChatGPT: towards enhancing information retrieval and judgment skills [22]	36
How ChatGPT will change software engineering education [23]	44
Complementary role of large language models in educating undergraduate design of distillation columns: methodology development [24]	94
Practical application of AI and Large Language Models in software engineering education [25]	95
Work-in-Progress: integrating generative AI with evidence-based learning strategies in computer science and engineering education [26]	98
A perspective on the synergistic potential of artificial intelligence and product-based learning strategies in biobased materials education [27]	111
Digital transformation in engineering education: exploring the potential of AI-assisted learning [28]	126
ChatGPT in the classroom. Exploring its potential and limitations in a functional programming course [29]	130
Exploring the application of ChatGPT in mechanical engineering education [30]	131
ChatGPT in computer science curriculum assessment: an analysis of its successes and shortcomings [31]	135
ChatGPT challenges blended learning methodologies in engineering education: a case study in mathematics [32]	145
Exploring students' perceptions of ChatGPT: thematic analysis and follow-up survey [33]	153
Student perceptions of ChatGPT use in a college essay assignment: implications for learning, grading, and trust in artificial intelligence [34]	166
Exploring the use of large language models (LLMs) in chemical engineering education: building core course problem models with ChatGPT [35]	167
Exploring the role of AI assistants in computer science education: methods, implications, and instructor perspectives [36]	180
Investigating the use of AI-generated exercises for beginner and intermediate programming courses: a ChatGPT case study [37]	230
Explicitly introducing ChatGPT into first-year programming practice: challenges and impact [38]	262

Before describing the next research activities, Figure 3 depicts a flow diagram summarizing the search and selection process that led to the dataset used in the research—from the use of the queries to select the 370 records from the databases to the selection of the final 20 papers (PRISMA 16a).



**Figure 3.** Flow diagram of the search and selection process.

These papers were then read carefully to look for correspondences to the RDs (PRISMA10a). They were primarily papers published in 2023 (17 papers) and 2024 (3 papers), mostly in conference proceedings (10 papers) or in indexed scientific journals (10 papers). They mainly described experiences related to engineering education in the IT field, focusing on software engineering courses (9 papers), electrical/electronic engineering (4), or chemical engineering (3). There were very few works related to other engineering fields (4). The papers primarily aimed to understand the influence that the use of GenAI tools, mainly ChatGPT, can have in educational settings. Eleven papers also included a component of investigating the opinions of different users through the use of questionnaires. In particular, many papers discussed the possibility of using LLMs for exercises related to programming and code production, evaluating the situation both before and after the introduction of new LLM-based tools or assessing the reliability of solving exercises assigned during classes. Some papers also concerned the evaluation of the degree of reliability and the correctness of the solutions obtained. In some papers that were primarily related to non-computer-science subjects, the possibility of using LLM-based tools for text production and in-depth exploration of topics of interest (essay production) was evaluated, thus assessing the reliability of the information obtained. Several papers presented, in different ways, the potential advantages and disadvantages of the introduction and use of these tools. In some cases, observations came from both students' and educators' perspectives.

What followed mapped the eight RDs to the peculiarities of the 20 papers (PRISMA17) in detail.

### 3.1. RD1—WHO

The WHO dimension allowed the encoding of whether the actors involving LLMs in education activities were students, educators, or others. Moreover, for these roles, it was assessed whether the participation in the described experiences was direct or indirect. As shown in Table 2, 12 out of the 20 papers depicted the active participation of students in,

e.g., performing coding activities or assignments during lessons or at home. Seven other papers reported indirect student participation, e.g., exercises that were typically assigned to students were performed with LLM-based tools. In five papers, there was indirect participation of educators in the experiences, as these works suggested methodological approaches or provided advice to educators on the use of these tools within their courses. In a single study, direct involvement of the teaching component was observed through interviews.

**Table 2.** Papers referring to RD1—WHO.

RD1—WHO			
Students		Educators	
Direct	Indirect	Direct	Indirect
4, 20, 36, 95, 98, 131, 135, 145, 153, 166, 167, 262	2, 44, 94, 111, 126, 130, 230	180	2, 44, 94, 111, 126

### 3.2. RD2—HOW

The HOW dimension refers to the ways in which LLMs are involved in education. In order to facilitate a comparative analysis, the nuances found in the papers were grouped into types of reference activities, as shown in Table 3. The experiences of use referred mainly to tests of the use of LLM-based tools (11 papers), the proposal of methods of use and guidelines (14 papers), the development of projects (3 papers), the development of specific tools (1 paper), or the description of case studies (4 papers). Tests of use mainly referred to code development or solving programming exercises in different languages (sometimes mathematical problems). The use method proposals also referred to guidelines such as breaking requests into smaller pieces, checking information that was gathered, training educators and students before use, etc. Some of the 20 papers described the type of instructional strategy used. For example, paper 98 referred to the use of “evidence-based learning” practices that emphasized the importance of defining learning goals well to obtain the desired level of student understanding and achievement. Paper 111 reported the use of ChatGPT within a strategy of product-based learning lessons; specifically, this tool was used to support the concept generation phase, produce scientific texts to better understand topics, and propose innovative solutions. Other papers recommended that before using ChatGPT to help carry out some tasks, it is important to educate students on the basic topics. An example is the case of paper 94, which was related to the chemical field, where the authors suggested educating students about “mass transfer” topics before using ChatGPT to develop specific chemical design projects. In addition, paper 153 emphasized the importance of knowing how to ask appropriate questions and have an adequate background in the relevant field of study to interact effectively with ChatGPT. Finally, other papers, such as papers 94 and 167, recommended that educators rethink the structure of their lessons to introduce the use of LLM-based tools, especially when tackling complex problems.

**Table 3.** Papers referring to RD2—HOW.

RD2—HOW				
Tests of Use	Use Method Proposals	Project Work Dev.	Develop Tools	Case Studies
44, 98, 126, 130, 131, 135, 145, 153, 166, 230, 262	2, 4, 20, 44, 94, 98, 111, 126, 135, 153, 166, 167, 180, 230	4, 95, 167	36 (worksheet)	20, 94, 111, 167



### 3.3. RD3—WHY

The WHY dimension showed the reasons/goals for/of the involvement of LLMs. Again, an attempt was made to distribute the 20 papers into a few categories, as reported in Table 4. In all papers, LLMs were used for the creation of different content, e.g., to generate code in different programming languages, to produce essays on specific topics, and to create ad hoc exercises for teaching. Twelve papers also reported the enhancement of understanding of certain learning topics as motivation for use. Nine papers highlighted the enrichment of problem solving for different types of problems that were relevant to computer science, chemistry, and mechanics. Six papers investigated the possibility of improving critical thinking, while another five investigated the enrichment of personalized learning. Thirteen papers verified the possibility of using LLMs to improve teaching in general. Finally, two papers reported reasons related to the possibility of developing projects in collaborative teams.

**Table 4.** Papers referring to RD3—WHY.

RD3—WHY						
Creation of Content	Enhance Understanding	Enrich Problem Solving	Improve Critical Thinking	Enrich Personalized Learning	Develop Teaching Enhancement	Develop Collaborative Projects
2, 4, 20, 36, 44, 94, 95, 98, 111, 126, 130, 131, 135, 145, 153, 166, 167, 180, 230, 262	2, 20, 36, 94, 95, 98, 111, 126, 131, 153, 166, 167	2, 4, 44, 94, 95, 98, 131, 145, 167	36, 94, 98, 111, 131, 145	20, 94, 131, 153, 166	2, 20, 44, 94, 98, 111, 126, 130, 166, 167, 180, 230, 262	4, 135

### 3.4. RD4—HOW MUCH

The HOW MUCH dimension referred to qualitative/quantitative evaluations of the impact of the involvement of LLMs in engineering education activities. Qualitatively speaking, all of the approaches that did not focus on numerical data were considered, rather referring to authors’ observations or collections of users’ opinions, such as in paper 98, where reference was made to the collection of student feedback through simple questionnaires based on open-ended questions that reported students’ personal opinions, or paper 153, where a thematic analysis of students’ perceptions was reported. Only a few of the selected papers reported quantitative evaluations, as shown in Table 5. The only papers that did so mostly used questionnaires that were administered to students. In some cases, e.g., in paper 20, these questionnaires were implemented both before and after the use of LLM-based tools to perform certain tasks. In papers 135 and 262, quantitative evaluations were used to compare groups of people who did and did not use LLM-based tools. Table 5 also highlights the moment of data collection by specifying “pre” (before the LLMs use), “post” (after), or “pre-post” (before and after) for each paper.

**Table 5.** Papers referring to RD4—HOW MUCH.

RD4—HOW MUCH	
Qualitative Eval.	Quantitative Eval.
2 (post), 4 (post), 36 (post), 44 (post), 94 (post), 95 (post), 98 (pre), 111 (post), 126 (post), 130 (pre), 131 (pre), 145 (post), 153 (pre-post), 167 (post), 180 (pre), 230 (post), 262 (post)	20 (pre-post), 135 (post), 166 (post), 262 (post)

### 3.5. RD5—WHAT

The WHAT dimension allows the description of the LLM-based tools made available day by day and adopted paper by paper, if any. All of the experiences described in the selected papers focused on the use of ChatGPT in different versions—mainly version 3 or 3.5 and, in some cases, version 4. Table 6 specifies the version for each paper if this information was available. Only one paper worked with other LLM-based tools: Bard, DALL-E, Bing Images, and Stable Diffusion.

**Table 6.** Papers referring to RD5—WHAT.

RD5—WHAT	
ChatGPT	Others
2, 4 (ver. 3.5 and ver. 4), 20, 36, 44, 94 (ver. 3.5), 95, 98, 111 (ver. 4), 126 (ver. 3.5), 130 (ver. 3.5), 131, 135, 145 (ver. 3.5 and ver. 4), 153, 166, 167 (ver. 3.5), 180, 230 (ver. 3), 262	95 (Bard, DALL-E, Bing Images, Stable Diffusion)

### 3.6. RD6—WHERE

The WHERE dimension describes the engineering domains in which the LLM involvement took place. Reading the papers revealed that descriptions of experiences were predominantly found in software engineering and computer science with nine papers, followed by electrical/electronic engineering with four papers and chemical engineering with three papers. Only two papers referred to studies in mechanical engineering. Two papers referred to other fields, industrial engineering, and aerospace engineering. Table 7 reports all of this.

**Table 7.** Papers referring to RD6—WHERE.

RD6—WHERE				
Software Eng./Computer Science	Electrical/Electronic Eng.	Chemical Eng.	Mechanical Eng.	Other Eng.
2, 4, 44, 95, 98, 130, 135, 180, 262	36, 126, 153, 230	94, 111, 167	20, 166	131 (industrial eng.), 145 (aerospace eng.)

### 3.7. RD7—WHEN

Concerning the dimension WHEN, representing the moment of the educative path at which the involvement of LLMs took place, the majority of the experiences reported in the papers referred to the undergraduate level (16 papers). Only four papers refer to the postgraduate level. It should be noted that in some cases, the experiences of students from both levels were considered, such as in papers 36, 98, and 230 (see Table 8).

**Table 8.** Papers referring to RD7—WHEN.

RD7—WHEN	
Undergraduate	Postgraduate
4, 36, 94, 95, 98, 126, 130, 131, 135, 145, 153, 166, 167, 180, 230, 262	20, 36, 98, 230

### 3.8. RD8—PROS/CONS

The PROS/CONS dimension considered the advantages and disadvantages of the involvement of LLMs reported in the papers analyzed. Some of them were quite clear about the advantages and drawbacks. These are reported in Tables 9 and 10, respectively.

**Table 9.** Pros in the papers referring to RD8—PROS/CONS.

RD8—PROS/CONS	
Papers	Pros
2, 20, 44, 94, 126, 131, 153, 166, 167, 262	Enhanced understanding of different concepts or topics
2, 94, 131	Adoption of real-world examples and practical applications
2, 94, 145, 262	Iterative and guided learning
2, 20, 94, 111, 145, 262	Instant feedback
2, 94, 153	Increase in student engagement and motivation
2, 20, 111	Peer collaboration and knowledge sharing
4, 98, 135	Code development in different programming languages
20, 44	Better task performance in different assignments
2, 20, 44, 94, 166, 167, 180, 230	Supporting educators in teaching organization
111, 145, 167, 262	Improve problem solving and critical thinking

**Table 10.** Cons in the papers referring to RD8—PROS/CONS.

RD8—PROS/CONS	
Papers	Cons
4, 98	Difficulties in handling code errors
4, 20, 36, 94, 95, 98, 126, 131, 145, 153, 166, 230, 262	Confusing and contradictory answers, inaccuracies in responses
44	Unsupervised use by students
20	Inaccuracy of bibliographical sources
20, 94, 111, 145	Ethical concerns and responsible use
111, 130, 135, 145	Plagiarism

#### 4. Results

The results of the review were as follows (PRISMA23a). Based on the comprehensive data collected from the analysis of the 20 selected papers (Table 1) and summarized in Tables 2–10 in relation to the eight research dimensions (RDs), the following reflections provide deeper insights into research questions RQ1 to RQ3, which were posed in order to investigate the current involvement of LLMs in engineering education.

Referring to RQ1, “Are the roles and duties of people clear regarding the involvement of LLMs in engineering education?”, four dimensions were involved.

- RD1—WHO: This indicated that both students and educators were involved in LLM activities, with varying degrees of direct and indirect participation. For example, as shown in Table 2, papers 4, 20, 36, 95, 98, 131, 135, 145, 153, 166, 167, and 262 depicted the active participation of students in coding activities or assignments during lessons or at home. This showed the direct involvement of students in LLM activities.
- RD2—HOW: This described the types of activities involving LLMs, such as tests of use and method proposals, but did not directly specify the roles and duties. For example, looking at Table 3, papers such as 2, 4, 20, 44, 94, 98, 111, 126, 130, 135, 153, 166, 167, 180, and 230 proposed methods of use and guidelines for LLM tools, suggesting roles for educators in implementing these methods within their courses.
- RD7—WHEN: This specified the moment in the educational path at which LLM involvement occurs, as it can influence the clarity of roles and duties at different

stages of education. For example, referring to Table 8, papers such as 4, 36, 94, 95, 98, 126, 130, 131, 135, 145, 153, 166, 167, 180, 230, and 262 focused on undergraduate-level experiences, indicating the moment in the educational path at which LLM involvement occurs.

- RD8—PROS/CONS: This provided insight into the advantages and disadvantages associated with LLM involvement, as they indirectly reflect roles and duties, such as confusion due to contradictory answers. For example, as shown in Table 10, papers such as 4 and 98 reported difficulties in handling code errors as a disadvantage of LLM involvement, which may indicate unclear roles in overseeing LLMs' use.

In summary, the analysis revealed a spectrum of participation levels among students and educators in LLM activities, with some papers depicting direct engagement in coding exercises or assignments, while others portrayed indirect involvement through methodological guidance or advisories. These findings underscored the complexity of roles and responsibilities within the context of engineering education, suggesting a need for clearer delineation and communication of duties to optimize the integration of LLMs into educational practices.

Regarding RQ2, "Is there evidence of relationships between engineering disciplines and the ways that LLMs are involved in the related educational activities?", three dimensions were involved.

- RD2—HOW: This indicated the types of activities involving LLMs across different engineering disciplines, revealing potential patterns in their utilization. For instance, the papers listed in Table 3, such as 44, 98, 126, 130, 131, 135, 145, 153, 166, 230, and 262, primarily focused on tests of the use of LLM-based tools, while papers such as 2, 4, 20, 44, 94, 98, 111, 126, and 230 proposed methodological approaches, indicating the ways in which LLMs were involved across different engineering disciplines.
- RD6—WHERE: This identified the engineering domains where the involvement of LLMs took place, as this could influence the types of activities observed. For example, in Table 7, we observed that software engineering/computer science papers (e.g., papers 2, 4, 44, 95, 98, 130, 135, 180, and 262) predominantly involved LLM activities.
- RD7—WHEN: This specified the moment in the educational path at which LLM involvement occurs, as this could also influence the types of activities observed across engineering disciplines. For example, referring to Table 8, papers such as 4, 36, 94, 95, 98, 126, 130, 131, 135, 145, 153, 166, 167, 180, 230, and 262 focused on undergraduate-level experiences, showing the timing of LLM involvement in engineering education.

To synthesize the findings, discernible patterns emerged regarding the utilization of LLMs across various engineering disciplines, with some disciplines predominantly emphasizing tests of use or method proposals, while others prioritized case studies or project development. These trends suggest that the specific focus of educational activities within each discipline influences the ways that LLMs are incorporated, highlighting the importance of tailoring LLM integration strategies to discipline-specific needs and objectives.

Finally, concerning RQ3, "Can clear indications of which LLM-based tools to involve in order to improve the effectiveness of education activities and of impact measurements be obtained?", four dimensions were involved.

- RD4—HOW MUCH: This examined the evaluations of the impact of the involvement of LLMs, providing insight into the effectiveness of different tools. For example, papers such as 2, 4, 20, 36, 44, 94, 95, 98, 111, 126, 130, 131, 145, 153, 167, 180, 230, and 262, which are listed in Table 5, provide qualitative evaluations of LLM involvement, offering insights into the effectiveness of LLM-based tools.
- RD5—WHAT: This describes the specific LLM-based tools used, as this can inform decisions on tool selection for improving educational activities. For example, in Table 6, we see that ChatGPT—particularly versions 3 and 3.5—was the predominant LLM-based tool used in the analyzed papers (e.g., papers 2, 4, 20, 36, 44, 94, 95, 98, 111, 126, 130, 131, 135, 145, 153, 166, 167, 180, 230, and 262).

- RD7—WHEN: This specified the moment in the educational path at which LLM involvement occurred, as it could influence the effectiveness and impact of LLM-based tools at different stages of education. For example, referring to Table 8, papers such as 20, 36, 98, and 230 focused on postgraduate-level experiences, indicating the timing of LLM involvement for impact measurement at different educational levels.
- RD8—PROS/CONS: This considered the advantages and disadvantages associated with LLM involvement, providing insights into which aspects of LLMs contributed to effectiveness and impact measurement. For instance, papers such as 2, 20, 44, 94, 126, 131, 153, 166, 167, and 262, which are listed in Table 9, outlined advantages such as enhanced understanding and engagement, providing indications on the effectiveness of certain aspects of LLM involvement.

In conclusion, the analysis yielded insights into the effectiveness of specific LLM-based tools for enhancing educational activities and measuring impact, with certain tools demonstrating advantages such as enhanced student engagement, improved problem-solving abilities, and increased task performance. These findings offer valuable guidance for educators and policymakers seeking to optimize educational outcomes through informed selection and implementation of LLM-based tools, emphasizing the importance of considering both the pedagogical context and the desired educational objectives.

## 5. Discussion

As the first point of discussion, in order to follow the PRISMA checklist as much as possible, it is worth stating that the quality of evidence in the studies included in the review ranged from “very low” to “high”, depending on several factors. For example, mainly in papers belonging to conference proceedings, due to the small number of pages allowed, the descriptions of the experiences were rather essential; therefore, in this case, the quality of evidence should be considered “very low” or “low”. On the contrary, papers published in journals are usually complete and more detailed, so their quality can be considered “high” (PRISMA23b).

Some limitations of the review can be highlighted as well. First, since the domain where the research took place is rapidly evolving, the outcome risks being outdated in the near future. Indeed, it is worth saying that this outcome is valid at the time of the queries (6 March 2024). Moreover, the literature allows the situation at the time of the writing of the papers to be depicted; therefore, we can assume a delay of several months with respect to the current situation, which is a long time considering the rapid evolution of the AI field. At the time of the reading of this paper, some issues highlighted by this research might have been solved in new versions of LLMs. Moreover, the novelty of the spread of LLMs necessitates some kind of shortage related information availability, the different LLM-based tools involved (just one, up to now), or the variety of engineering disciplines where LLMs are currently involved (PRISMA23c).

Moreover, this research allowed the identification of some gaps in engineering education and in the involvement of LLM-based tools within courses; these gaps are where further research would be needed. For example, current research lacks insight into the development and evaluation of specific pedagogical approaches to engaging LLMs in engineering education activities. There are few detailed examples of the integration of LLM-based tools in different engineering disciplines and course levels. In addition, there are no examples of evaluations of the impact of the involvement of LLM-based tools on student engagement, participation, and interaction in engineering courses. There are also a few papers that explored the potential of LLM-based tools to personalize and tailor learning experiences for individual students in engineering courses or to help educators make the best use of these tools.

The results of this review have possibilities for practical adoption, and they suggest future research directions. As the proposal of practical suggestions for putting LLMs into practice in engineering education was one of the goals of this research, as claimed in the abstract, the following text focuses on this (PRISMA23d). Table 11 lists suggestions for

improving the effectiveness of involving LLMs in engineering education while ensuring a responsible and ethical approach. Although each suggestion comes from a specific RQ, as is easily recognizable, this information has been reputed to be useless when an educator uses these suggestions to improve their educational activities; thus, this information does not appear in the list.

**Table 11.** Suggestions for improving engineering education activities through LLM involvement.

#	Suggestion
1	Clarify roles and responsibilities: Clearly define the roles and responsibilities of both students and educators in LLM activities to ensure effective integration into educational practices
2	Tailor integration strategies: Tailor LLM integration strategies to discipline-specific needs and objectives considering the distinct educational focus and timing of involvement across different engineering disciplines
3	Utilize effective LLM-based tools: Explore and utilize effective LLM-based tools, such as ChatGPT versions 3 and 3.5, to enhance educational activities and measure impact effectively
4	Promote direct engagement: Encourage direct engagement of students in coding exercises or assignments by leveraging LLMs as tools for active learning, critical thinking, and problem solving
5	Provide methodological guidance: Offer methodological guidance and advice for educators on the effective implementation of LLM tools within their courses, ensuring consistency and clarity in usage
6	Consider the pedagogical context: Consider the pedagogical context and desired educational objectives when selecting and implementing LLM-based tools, ensuring alignment with learning outcomes
7	Address challenges: Address challenges associated with LLM involvement, such as difficulties in handling code errors or confusion due to contradictory answers, through targeted interventions and support mechanisms
8	Stay updated: Stay updated on emerging trends and advancements in LLM technology and education practices and adapt integration strategies accordingly to remain relevant and effective
9	Encourage collaboration: Foster collaboration and knowledge sharing among students through peer collaboration activities facilitated by LLM tools, promoting a collaborative learning environment
10	Evaluate impact: Continuously evaluate the impact of LLM involvement on educational activities and student outcomes by utilizing qualitative and quantitative measures to inform ongoing improvements and optimizations

By implementing these suggestions, educators can enhance their activities in engineering education by leveraging LLMs as valuable tools for facilitating learning, promoting engagement, and achieving educational objectives effectively.

## 6. Conclusions

The research described in this study aimed to systematically review the existing literature on the involvement of LLMs in engineering education, with a focus on how to improve educational activities at different levels using different actors in different engineering domains and with the LLM-based tools that are made available as time progresses. Despite the relatively small number of papers analyzed, which was noted as a limitation, interesting results were obtained. Although LLMs became widely available only a few years ago, the material collected here made it possible to list some practical suggestions that we were the first to put into practice in our undergraduate and postgraduate courses.

Both the limitations of the research and the gaps highlighted by the systematic review, as described in the Discussion section, provide valuable insights into potential areas for future exploration.

Regarding the limitations of this research, to prevent obsolescence and support the updating of outcomes, the suggestion is to evaluate emerging LLM tools across disciplines to understand their efficacy and limitations. Creating dedicated repositories for LLM-based tools could help address information shortages. Identifying new engineering disciplines for LLM applications is crucial, along with assessing their impacts. Longitudinal research studies can be conducted to investigate the long-term impact of integrating LLMs and LLM-based tools into engineering education on student learning outcomes, career readiness,

and post-graduation success by tracking students' academic performance, professional achievements, and attitudes toward AI over time. The aim of following the PRISMA checklist in this research was also to make it somehow robust and replicable so that similar reviews can be performed to keep the outcomes up to date with the evolution of the AI field.

Considering the gaps, there could be further study of deeper teaching strategies, learning activities, and assessment methods that effectively leverage LLMs to improve student learning outcomes. In addition, special attention can be given to describing the design of curricular modules or assignments that incorporate LLM-based tools to support various engineering disciplines. In addition, it could be considered how LLM-based activities influence student motivation, collaboration, and peer learning experiences within the classroom. Finally, it is also important to address the professional development needs of engineering educators to effectively integrate LLM-based tools into their teaching practices and to enhance their pedagogical competence and confidence in using AI technologies by providing training, resources, and technical support.

From a general research perspective, fostering collaboration among engineering education researchers, AI experts, instructional designers, and industry practitioners could facilitate interdisciplinary approaches to exploring the potential of LLM-based tools in engineering education. This could lead to innovative solutions addressing complex challenges and opportunities at the intersection of AI and engineering pedagogy. Moreover, incorporating user feedback is essential for improving the usability of LLM-based tools. Finally, investigating the biases, privacy concerns, and societal impacts of LLM adoption is imperative for ethical and responsible deployment.

**Author Contributions:** The two authors (S.F. and B.M.) worked equally on all parts of the research and sections of the article. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Concerning the availability of material for replicating the research activities, no datasets were generated in this research (PRISMA27).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A. PRISMA Checklist

Table A1 contains the PRISMA checklist as in [15]. The labels in the table are those used throughout the paper to highlight the rigor of the research that has been conducted.

**Table A1.** PRISMA checklist [15].

Section	Topic	Item	Label
Title	Title	Identify the report as a systematic review.	PRISMA1
Abstract	Abstract	See the PRISMA 2020 for Abstracts checklist.	PRISMA2
Introduction	Rationale	Describe the rationale for the review in the context of existing knowledge.	PRISMA3
	Objectives	Provide an explicit statement of the objective(s) or question(s) the review addresses.	PRISMA4
Methods	Eligibility criteria	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	PRISMA5
	Information sources	Specify all databases, registers, websites, organizations, reference lists, and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	PRISMA6

Table A1. Cont.

Section	Topic	Item	Label
	Search strategy	Present the full search strategies for all databases, registers, and websites, including any filters and limits used.	PRISMA7
	Selection process	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and, if applicable, details of automation tools used in the process.	PRISMA8
	Data collection process	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and, if applicable, details of automation tools used in the process.	PRISMA9
	Data items	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g., for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	PRISMA10a
		List and define all other variables for which data were sought (e.g., participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	PRISMA10b
	Study risk of bias assessment	Specify the methods used to assess the risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and, if applicable, details of automation tools used in the process.	PRISMA11
	Effect measures	Specify for each outcome the effect measure(s) (e.g., risk ratio, mean difference) used in the synthesis or presentation of results.	PRISMA12
	Synthesis methods	Describe the processes used to decide which studies were eligible for each synthesis (e.g., tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	PRISMA13a
		Describe any methods required to prepare the data for presentation or synthesis, such as handling missing summary statistics or data conversions.	PRISMA13b
		Describe any methods used to tabulate or visually display the results of individual studies and syntheses.	PRISMA13c
		Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	PRISMA13d
		Describe any methods used to explore possible causes of heterogeneity among study results (e.g., subgroup analysis, meta-regression).	PRISMA13e
		Describe any sensitivity analyses conducted to assess the robustness of the synthesized results.	PRISMA13f
	Reporting bias assessment	Describe any methods used to assess the risk of bias due to missing results in a synthesis (arising from reporting biases).	PRISMA14



Table A1. Cont.

Section	Topic	Item	Label
	Certainty assessment	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	PRISMA15
Results	Study selection	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	PRISMA16a
		Cite studies that might appear to meet the inclusion criteria but were excluded, and explain why they were excluded.	PRISMA16b
	Study characteristics	Cite each included study and present its characteristics.	PRISMA17
	Risk of bias in studies	Present assessments of risk of bias for each included study.	PRISMA18
	Results of individual studies	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g., confidence/credible interval), ideally using structured tables or plots.	PRISMA19
	Results of syntheses	For each synthesis, briefly summarize the characteristics and risk of bias among contributing studies.	PRISMA20a
		Present results of all statistical syntheses conducted. If meta-analysis was performed, present for each the summary estimate and its precision (e.g., confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	PRISMA20b
		Present results of all investigations of possible causes of heterogeneity among study results.	PRISMA20c
		Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	PRISMA20d
	Reporting biases	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	PRISMA21
Certainty of evidence	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	PRISMA22	
Discussion	Discussion	Provide a general interpretation of the results in the context of other evidence.	PRISMA23a
		Discuss any limitations of the evidence included in the review.	PRISMA23b
		Discuss any limitations of the review processes used.	PRISMA23c
		Discuss the implications of the results for practice, policy, and future research.	PRISMA23d
Other information	Registration and protocol	Provide registration information for the review, including the register name and registration number, or state that the review was not registered.	PRISMA24a
		Indicate where the review protocol can be accessed or state that a protocol was not prepared.	PRISMA24b
		Describe and explain any amendments to information provided at registration or in the protocol.	PRISMA24c
		Support	Describe sources of financial or non-financial support for the review and the role of the funders or sponsors in the review.
Competing interests	Declare any competing interests of review authors.	PRISMA26	
	Availability of data, code, and other materials	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	PRISMA27

## References

1. OpenAI. 2024. (Apr 2024 Version). Available online: <https://openai.com/research/overview> (accessed on 18 April 2024).
2. Floridi, L.; Chiriatti, M. GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds Mach.* **2020**, *30*, 681–694. [CrossRef]
3. Yan, L.; Sha, L.; Zhao, L.; Li, Y.; Martinez-Maldonado, R.; Chen, G.; Li, X.; Jin, Y.; Gašević, D. Practical and Ethical Challenges of Large Language Models in Education: A Systematic Scoping Review. *Br. J. Educ. Technol.* **2024**, *55*, 90–112. [CrossRef]
4. Prather, J.; Denny, P.; Leinonen, J.; Becker, B.A.; Abluwi, I.; Craig, M.; Keuning, H.; Kiesler, N.; Kohn, T.; Luxton-Reilly, A.; et al. The Robots Are Here: Navigating the Generative AI Revolution in Computing Education. In Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education, Turku, Finland, 10–12 July 2023; ACM: Turku, Finland, 2023; pp. 108–159.
5. Chatterjee, S.; Bhattacharya, M.; Pal, S.; Lee, S.; Chakraborty, C. ChatGPT and Large Language Models in Orthopedics: From Education and Surgery to Research. *J. Exp. orthop.* **2023**, *10*, 128. [CrossRef] [PubMed]
6. King, D.R.; Nanda, G.; Stoddard, J.; Dempsey, A.; Hergert, S.; Shore, J.H.; Torous, J. An Introduction to Generative Artificial Intelligence in Mental Health Care: Considerations and Guidance. *Curr. Psychiatry Rep.* **2023**, *25*, 839–846. [CrossRef] [PubMed]
7. Kar, A.K.; Varsha, P.S.; Rajan, S. Unravelling the Impact of Generative Artificial Intelligence (GAI) in Industrial Applications: A Review of Scientific and Grey Literature. *Glob. J. Flex. Syst. Manag.* **2023**, *24*, 659–689. [CrossRef]
8. Nikolic, S.; Daniel, S.; Haque, R.; Belkina, M.; Hassan, G.M.; Grundy, S.; Lyden, S.; Neal, P.; Sandison, C. ChatGPT versus Engineering Education Assessment: A Multidisciplinary and Multi-Institutional Benchmarking and Analysis of This Generative Artificial Intelligence Tool to Investigate Assessment Integrity. *Eur. J. Eng. Educ.* **2023**, *48*, 559–614. [CrossRef]
9. Filippi, S. Measuring the Impact of ChatGPT on Fostering Concept Generation in Innovative Product Design. *Electronics* **2023**, *12*, 3535. [CrossRef]
10. Filippi, S. Relationships among Personality Traits, ChatGPT Usage and Concept Generation in Innovation Design. In *Artificial Intelligence, Social Computing and Wearable Technologies. Proceedings of the AHFE (2023) International Conference, San Francisco, CA, USA, 20–24 July 2023*; Karwowski, W., Ahram, T., Eds.; AHFE Open Access: New York, NY, USA, 2023; Volume 113, p. 113. [CrossRef]
11. Tan, T.F.; Thirunavukarasu, A.J.; Campbell, J.P.; Keane, P.A.; Pasquale, L.R.; Abramoff, M.D.; Kalpathy-Cramer, J.; Lum, F.; Kim, J.E.; Baxter, S.L.; et al. Generative Artificial Intelligence through ChatGPT and Other Large Language Models in Ophthalmology. *Ophthalmol. Sci.* **2023**, *3*, 100394. [CrossRef] [PubMed]
12. Zhang, C.; Chen, J.; Li, J.; Peng, Y.; Mao, Z. Large Language Models for Human–Robot Interaction: A Review. *Biomim. Intell. Robot.* **2023**, *3*, 100131. [CrossRef]
13. Bahroun, Z.; Anane, C.; Ahmed, V.; Zacca, A. Transforming Education: A Comprehensive Review of Generative Artificial Intelligence in Educational Settings through Bibliometric and Content Analysis. *Sustainability* **2023**, *15*, 12983. [CrossRef]
14. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
15. PRISMA 2020 Checklist. Available online: [https://www.prisma-statement.org/s/PRISMA\\_2020\\_checklist-fxke.docx](https://www.prisma-statement.org/s/PRISMA_2020_checklist-fxke.docx) (accessed on 18 April 2024).
16. GEMINI. Available online: <https://gemini.google.com/app> (accessed on 18 April 2024).
17. ALPACA. Available online: <https://crfm.stanford.edu/2023/03/13/alpaca> (accessed on 18 April 2024).
18. ELICIT. Available online: <https://ought.org/elicit> (accessed on 18 April 2024).
19. Abdelfattah, A.M.; Ali, N.A.; Elaziz, M.A.; Ammar, H.H. Roadmap for Software Engineering Education Using ChatGPT. In Proceedings of the 2023 International Conference on Artificial Intelligence Science and Applications in Industry and Society (CAISAIS), Galala, Egypt, 3–5 September 2023; IEEE: Galala, Egypt, 2023; pp. 1–6.
20. Abrahamsson, P.; Anttila, T.; Hakala, J.; Ketola, J.; Knappe, A.; Lahtinen, D.; Liukko, V.; Poranen, T.; Ritala, T.-M.; Setälä, M. ChatGPT as a Fullstack Web Developer—Early Results. In *Agile Processes in Software Engineering and Extreme Programming—Workshops*; Kruchten, P., Gregory, P., Eds.; Lecture Notes in Business Information Processing; Springer Nature: Cham, Switzerland, 2024; Volume 489, pp. 201–209. ISBN 978-3-031-48549-7.
21. Bernabei, M.; Colabianchi, S.; Falegnami, A.; Costantino, F. Students’ Use of Large Language Models in Engineering Education: A Case Study on Technology Acceptance, Perceptions, Efficacy, and Detection Chances. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100172. [CrossRef]
22. Chen, L.; Shimada, A. Designing Worksheet for Using ChatGPT: Towards Enhancing Information Retrieval and Judgment Skills. In Proceedings of the 2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), Auckland, New Zealand, 27 November–1 December 2023; IEEE: Auckland, New Zealand, 2023; pp. 1–4.
23. Daun, M.; Brings, J. How ChatGPT Will Change Software Engineering Education. In Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1, Turku, Finland, 10–12 July 2023; ACM: Turku, Finland, 2023; pp. 110–116.
24. Kong, Z.Y.; Adi, V.S.K.; Segovia-Hernández, J.G.; Sunarso, J. Complementary Role of Large Language Models in Educating Undergraduate Design of Distillation Column: Methodology Development. *Digit. Chem. Eng.* **2023**, *9*, 100126. [CrossRef]
25. Kozov, V.; Ivanova, G.; Atanasova, D. Practical Application of AI and Large Language Models in Software Engineering Education. *IJACSA* **2024**, *15*, 690–696. [CrossRef]

26. Lauren, P.; Watta, P. Work-in-Progress: Integrating Generative AI with Evidence-Based Learning Strategies in Computer Science and Engineering Education. In Proceedings of the 2023 IEEE Frontiers in Education Conference (FIE), College Station, TX, USA, 18–21 October 2023; IEEE: College Station, TX, USA, 2023; pp. 1–5.
27. Marquez, R.; Barrios, N.; Vera, R.E.; Mendez, M.E.; Tolosa, L.; Zambrano, F.; Li, Y. A Perspective on the Synergistic Potential of Artificial Intelligence and Product-Based Learning Strategies in Biobased Materials Education. *Educ. Chem. Eng.* **2023**, *44*, 164–180. [[CrossRef](#)]
28. Pham, T.; Nguyen, T.B.; Ha, S.; Nguyen Ngoc, N.T. Digital Transformation in Engineering Education: Exploring the Potential of AI-Assisted Learning. *AJET* **2023**, *39*, 1–19. [[CrossRef](#)]
29. Popovici, M.-D. ChatGPT in the Classroom. Exploring Its Potential and Limitations in a Functional Programming Course. *Int. J. Hum.–Comput. Interact.* **2023**. *ahead of print*. [[CrossRef](#)]
30. Puig-Ortiz, J.; Pàmies-Vilà, R.; Jordi Nebot, L. Exploring the Application Of Chatgpt in Mechanical Engineering Education. In Proceedings of the 51st Annual Conference of the European Society for Engineering Education (SEFI), Dublin, Ireland, 11–14 September 2023. [[CrossRef](#)]
31. Qureshi, B. ChatGPT in Computer Science Curriculum Assessment: An Analysis of Its Successes and Shortcomings. In Proceedings of the 2023 9th International Conference on e-Society, e-Learning and e-Technologies, Portsmouth, UK, 9–11 June 2023; ACM: Portsmouth, UK, 2023; pp. 7–13.
32. Sánchez-Ruiz, L.M.; Moll-López, S.; Nuñez-Pérez, A.; Moraño-Fernández, J.A.; Vega-Fleitas, E. ChatGPT Challenges Blended Learning Methodologies in Engineering Education: A Case Study in Mathematics. *Appl. Sci.* **2023**, *13*, 6039. [[CrossRef](#)]
33. Shoufan, A. Exploring Students’ Perceptions of ChatGPT: Thematic Analysis and Follow-Up Survey. *IEEE Access* **2023**, *11*, 38805–38818. [[CrossRef](#)]
34. Tossell, C.C.; Tenhundfeld, N.L.; Momen, A.; Cooley, K.; De Visser, E.J. Student Perceptions of ChatGPT Use in a College Essay Assignment: Implications for Learning, Grading, and Trust in Artificial Intelligence. *IEEE Trans. Learn. Technol.* **2024**, *17*, 1069–1081. [[CrossRef](#)]
35. Tsai, M.-L.; Ong, C.W.; Chen, C.-L. Exploring the Use of Large Language Models (LLMs) in Chemical Engineering Education: Building Core Course Problem Models with Chat-GPT. *Educ. Chem. Eng.* **2023**, *44*, 71–95. [[CrossRef](#)]
36. Wang, T.; Díaz, D.V.; Brown, C.; Chen, Y. Exploring the Role of AI Assistants in Computer Science Education: Methods, Implications, and Instructor Perspectives. In Proceedings of the 2023 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), Washington, DC, USA, 3–6 October 2023; IEEE: Washington, DC, USA, 2023; pp. 92–102.
37. Speth, S.; Meißner, N.; Becker, S. Investigating the Use of AI-Generated Exercises for Beginner and Intermediate Programming Courses: A ChatGPT Case Study. In Proceedings of the 2023 IEEE 35th International Conference on Software Engineering Education and Training (CSEE&T), Tokyo, Japan, 7–9 August 2023; IEEE: Tokyo, Japan, 2023; pp. 142–146.
38. Hu, M.; Assadi, T.; Mahrooian, H. Explicitly Introducing ChatGPT into First-Year Programming Practice: Challenges and Impact. In Proceedings of the 2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), Auckland, New Zealand, 27 November–1 December 2023; IEEE: Auckland, New Zealand, 2023; pp. 1–6.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.