

# Evaluating the Role of Large Language Models in Lesson Planning: Insights from a Narrative Review

Vassilis A. Failadis<sup>a</sup>, Sotiris K. Tasoulis<sup>a</sup>, Spiros V. Georgakopoulos<sup>b</sup>, Vassilis P. Plagianakos<sup>a</sup>

<sup>a</sup> Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia 35131, Greece

<sup>b</sup> Department of Mathematics, University of Thessaly, Lamia 35100, Greece

© Higher Education Press 2026

**Abstract** While much of the existing literature on AI in education takes a broad approach, this study focuses specifically on lesson planning. By narrowing its scope to this underexplored area, the study offers a structured synthesis of 53 recent publications. This provides a clearer view of how large language models (LLMs) are currently being used, evaluated, and received by educators. The study explores whether LLMs can produce lesson plans that are pedagogically appropriate and practically useful. The analysis follows the search–appraisal–synthesis–analysis framework and organizes the findings around four main themes: quality and pedagogical value, challenges and limitations, teachers’ attitudes and perceptions, and improvements and future directions. Overall, the literature shows that while LLMs can help teachers organize content and save time, their output often lacks the flexibility, contextual awareness, and pedagogical depth needed for classroom use. Most studies emphasize the importance of teacher involvement in reviewing, adapting, and critically evaluating material generated by LLMs. A significant shortfall in the existing literature is the still limited empirical evidence comparing lesson plans designed by teachers and those produced by LLMs. As these tools become more common, there is a growing need for practical research, thoughtful integration into teaching practice, and ongoing professional development to ensure their responsible and effective use in education.

**Keywords** large language models, lesson planning, LLM-generated lesson plans, pedagogical quality, teacher perceptions

## 1 Introduction

Today, education faces new demands because students have grown up surrounded by technology. In this context, the teacher’s role has become especially important, not just for passing on knowledge but also for helping students build skills such as digital literacy, adaptability, and critical thinking. Teachers are expected to update their skills, use emerging technologies, and rethink the ways in which they organize their lessons. Generative AI (GenAI), as part of the broader field of AI, including large language models (LLMs), seems to have the potential to support teaching in a substantive way. In this study, AI is treated as an umbrella term, while LLMs refer specifically to generative systems designed for natural language processing, such as ChatGPT and similar models. As pointed out by Demir and Ev Çimen (2024), when educators use these tools, they can benefit from them in practical ways, especially in designing lessons that are more structured, creative, and up-to-date.

Despite the rapid spread of AI tools in education, their usefulness cannot be judged only by their functionality. Their real value becomes visible only when they are used as part of actual teaching practice, especially in the context of lesson planning. Producing lesson plans is an essential part of a teacher’s daily work and a core element of their professional preparation, as they reflect the educator’s pedagogical choices and are closely linked to decisions about integrating technology into the classroom (Moundridou et al., 2024). At the same time, lesson plans shape the teacher’s instructional decisions and offer a structure for how lessons are organized and implemented in practice (Baytak, 2024).

In the international literature, lesson planning is described as a demanding process that requires time,

Received February 26, 2026; revised June 8, 2026; accepted June 10, 2026

Vassilis A. Failadis (✉)

E-mail: vfayladis@uth.gr

careful preparation of materials, and sound pedagogical judgment. As teachers are required to handle increasingly complex classroom conditions, the need for planning support has become more apparent worldwide (Fan et al., 2024). Despite concerns about possible misinformation or shallow content, LLMs may reduce preparation time, support teachers' critical thinking, and contribute to wider access to teaching materials (van den Berg & du Plessis, 2023). Instead of bans or restrictions, what is needed is a better understanding of the capabilities and limits of these models, along with their thoughtful integration into actual teaching practice. Recent empirical studies have examined the use of LLMs such as ChatGPT and Gemini in lesson planning and classroom-related tasks. These studies show that, while these systems can generate structured instructional material, their outputs often require careful teacher review, particularly in terms of conceptual accuracy and pedagogical adequacy (Bessas et al., 2025a).

In classroom contexts, such as junior high school physics, empirical research has examined the everyday use of ChatGPT in teaching practice, noting both its supportive role and the continued need for teacher oversight (Bessas et al., 2025b). Even though LLMs are increasingly present in everyday teaching practice, four central questions remain unanswered: (1) How complete and reliable are the lesson plans they produce? (2) Do they have real pedagogical value, or do they simply reproduce technical patterns? (3) How do the teachers who use them perceive them? (4) Do they view them as helpful tools, temporary solutions, or threats to their professional role? These issues have already been pointed out in the recent literature. Corp and Revelle (2023) observed that although there have been mentions of using ChatGPT to support critical thinking or general educational practices, there is still a lack of studies that focus specifically on lesson planning with LLMs and what this means for teacher education. This observation underlines the need for a review that organizes the relevant literature critically and offers a clearer picture of how LLMs are currently understood in instructional design. Exploring these issues is particularly important as AI becomes increasingly integrated in everyday teaching practices.

Based on the discussion above, a central question arises: Can LLMs produce lesson plans that possess pedagogical value and can be applied effectively in practice? To explore this question, the review examines two related sub-questions:

(1) Do LLMs generate lesson plans with genuine pedagogical value?

(2) To what extent are these plans suitable for use in classroom settings?

This review approaches these points by examining what has been recorded so far in the relevant literature, both in terms of practical applications and

concerns about their pedagogical use. It does not focus on the technical features of LLMs but on how they are used in actual teaching and instructional planning. The review is structured around four main themes that appeared repeatedly in the studies and relate to the quality and pedagogical value, challenges and limitations, teachers' attitudes and perceptions, and improvements and future directions. Although several papers have examined the use of AI in education more generally, there is still limited research that looks specifically at how LLMs are being used to generate lesson plans, and even less that evaluates these plans in terms of their pedagogical quality and practical value. Most of the available work remains either theoretical or focuses on teacher perceptions, while relatively few studies examine how AI-generated plans function when implemented in authentic classroom settings. By focusing specifically on lesson planning, this review offers a targeted examination of how LLM-generated plans are currently being used in this specific area, how their output is assessed, and what concerns or limitations are repeatedly noted across studies. In this way, the review aims to contribute to a better understanding of what these tools can and cannot offer when it comes to designing effective lessons, while also outlining where further research and support may be needed.

## 2 Methodology

This section outlines the methodological approach of the study, including the type of review adopted and the application of the search–appraisal–synthesis–analysis (SALSA) framework used to guide the analysis.

### 2.1 | Type of Review

This study adopts a narrative review and focuses on the thematic synthesis of sources related to the use of LLMs in lesson planning. The topic is relatively new and rapidly developing and includes studies that differ considerably in terms of type, methodology, and scope. This diversity in the literature makes it difficult to apply the strict research procedures associated with a systematic review. As Snyder (2019) notes, when a topic is at an early stage of development and the available studies are limited or diverse, a narrative approach is more appropriate. This type of review allows researchers to identify general trends, recurring issues, and areas requiring further investigation.

In this study, the aim is not to examine all publications exhaustively but rather to present the main issues related to the use of LLMs in lesson planning in a critical and focused way. The review is organized around four thematic areas that appeared repeatedly

across the studies examined: the quality and pedagogical value of lesson plans produced by LLMs, challenges and limitations, teachers' attitudes and perceptions, and improvements and future directions.

## 2.2 | SALSA Framework

This study adopts a narrative approach. The SALSA framework was used to organize the review process, including search, appraisal, synthesis, and analysis. This approach was chosen because it offers a simple and flexible way to study a new and complex topic, such as the use of LLMs in lesson planning, without requiring the strict procedures of a systematic review. As Mengist et al. (2020) explained, the SALSA framework is well suited to reviews involving varied literature, multiple perspectives, and topics that require an adaptable yet transparent analytical process. These features match the nature of this study, as the use of LLMs for lesson planning is still at an early stage, is covered by a wide range of studies, and is approached from technological, pedagogical, and research angles. Using SALSA allows for a clear presentation of how material was gathered and organized, clarifies the criteria used to select sources, and helps map out the main topics and trends found in the literature. In this way, the study's validity is supported, as each stage of the process is documented and clearly explained. The SALSA framework was used to structure the review process as follows: The search stage was used to guide the identification of relevant studies, the appraisal stage to select them based on relevance, and the synthesis and analysis stages to organize and interpret the findings thematically.

### 2.2.1 Search

The literature search was conducted through Google Scholar in March 2025, which provides access to a wide range of academic and gray literature. This database is frequently used in reviews that aim to include sources of different types. As Haddaway et al. (2015) noted, Google Scholar can assist researchers in locating both peer-reviewed and gray literature and may complement other databases by expanding the scope of the search. This choice reflects the fact that the literature on LLMs in lesson planning has spread across different sources and is rapidly evolving. Since Google Scholar ranks results by relevance and returns a large number of records, screening was limited to the first 250 results, as the relevance of subsequent results had substantially decreased. Of the 250 records screened, 45 studies met the inclusion criteria and were included in the initial review dataset, while 205 were excluded as not directly relevant to LLM use in lesson planning. To further reduce the risk of omitting relevant studies, additional searches were conducted in Scopus, Web of Science, and Educational Resources Information Center (ERIC),

using the same search terms and publication-date cutoff as in the initial search (March 2025) across all sources. The searches were conducted in April 2026 and the results were as follows: Scopus returned 5 records; 2 were already included in the existing set; and the remaining 3 were outside the focus of the review. ERIC returned 9 records, of which 1 was already included, 1 was added to the review, and 7 were excluded after screening. Web of Science returned 11 records; 5 had already been identified, 1 was added, and 5 were not retained. As a result, 2 additional relevant studies were retrieved through this database screening process, increasing the review dataset from 45 to 47 studies.

The search was carried out using a combination of keywords and Boolean operators to retrieve results directly related to the topic. The primary search string was formulated as follows:

“AI lesson plans” OR “AI-generated lesson plans” OR “ChatGPT lesson plans” OR “GPT lesson plans”).

Additional searches were conducted in May 2026 using LLM-related lesson-planning terms, including “LLM lesson plans,” “LLMs' lesson plans,” “LLM-generated lesson plans,” “large language model lesson plans,” and “large language models lesson plans.” These searches were conducted across Google Scholar, Scopus, Web of Science, and ERIC using the same inclusion criteria and publication-date limit as the initial search.

Google Scholar was the main search source for the additional LLM-related lesson-planning terms. Consistent with the initial Google Scholar search, the first 250 results were screened. Among these records, 37 had already been included, 2 new relevant studies were added, and 211 were excluded as not directly relevant to LLM use in lesson planning. Scopus returned 67 records; 6 had already been identified, 3 were added to the review, and 58 were outside the focus of the study. Web of Science returned 34 records; 9 had already been included, 1 was added, and 24 were not retained. ERIC returned 33 records; 6 overlapped with the review dataset, and no additional studies were included, resulting in the inclusion of 6 additional publications and a final dataset of 53 studies.

Although the search mainly aimed to include peer-reviewed publications, non-peer-reviewed sources (such as preprints and conference papers) were also kept because they were relevant to the topic and added useful perspectives. As the number of studies in this area is still limited, these sources were included to make the review more complete. They were read carefully and used to support and complement the peer-reviewed studies. The overall search resulted in a relatively limited number of relevant records, indicating the emerging nature of research on the use of LLMs in lesson planning. Studies not directly relevant to lesson planning were excluded. The study selection process is shown in Figure 1, while an overview of the reviewed

studies is presented in Figure 2. Most studies were from the United States and Türkiye, with fewer studies published in 2024, and a large proportion originated from other regions.

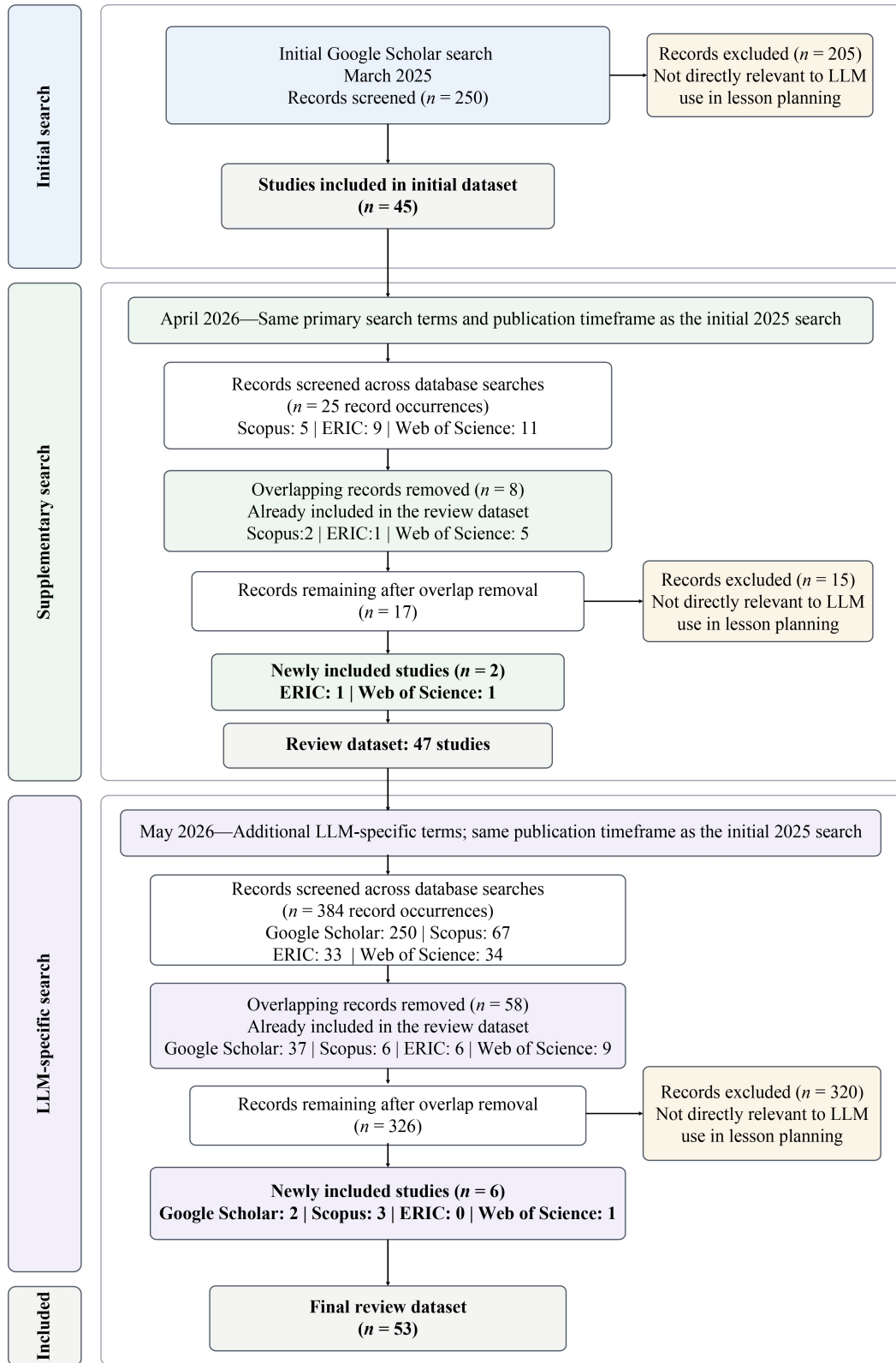


**Figure 1** Overview of the study selection process. *n* indicates the number of studies.

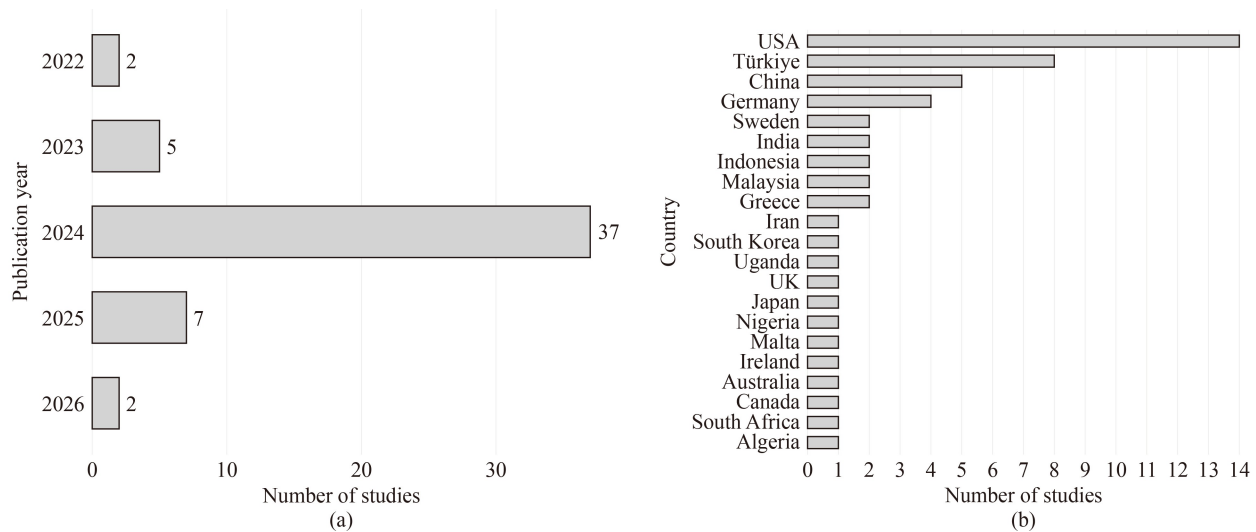


**Figure 2** Overview of the reviewed studies: distribution by (a) publication year and (b) geographical distribution ( $n = 53$ ).

### 2.2.2 Appraisal

The selection of articles was based on their relevance to the research questions, specifically the use of LLMs in lesson planning. Sources that referred to AI in education more generally without examining specific applications or analyses of lesson plans were excluded. Greater attention was given to studies that included empirical data or that followed well-documented research procedures. This involved both empirical and conceptual studies addressing the use of LLMs in lesson planning. The included studies were selected mainly from English-language publications and focused on LLMs or closely related AI tools used in lesson planning. Studies were excluded when they addressed AI in education only in general terms, did not focus on lesson planning, or did not provide sufficient information about the use of LLMs in instructional design. Of the 53 studies included, 47 were empirical and 6 were conceptual. No formal scoring system was used. Instead, they were selected based on their relevance to the research questions and the clarity of their contributions to the topic.

The final set of reviewed studies included publications with considerable variation in methodology, such as qualitative and mixed methods, experimental implementations, case studies, and theoretical papers with detailed discussion, while also varying in terms of subject focus, covering science, technology, engineering, and mathematics (STEM) ( $n = 19$ ), humanities and language education ( $n = 18$ ), and cross-disciplinary or general lesson planning contexts ( $n = 16$ ). In several cases, LLMs were studied as tools that support instructional design, while in others, the focus was on the pedagogical quality of the lesson plans produced. There were also studies that examined the limitations, weaknesses, and challenges associated with using these tools. Most studies concerned teacher education or teacher

professional practice, although some involved school-level classrooms, higher education, vocational education, or subject-specific instructional contexts.

The purpose in selecting these studies was not to classify by type or count quantitatively but to include a range of perspectives relevant to the research questions, such as how LLMs are used, how the generated lesson plans are evaluated, and what teachers report from using them. While the studies differ in context and method, they all deal directly with the use of LLMs in lesson plan creation and provide data or analyses relevant to the central questions of this review. The complete list of the reviewed studies is provided in Electronic Supplementary Material.

### 2.2.3 Synthesis

The literature was organized into four thematic areas, which emerged through the reading and comparative review of the selected studies. Instead of applying a quantitative classification or statistical processing of the articles, a thematic approach was chosen to draw out the main questions that appear repeatedly in the relevant research. This approach was more suitable for the nature of the material, which includes studies that differ in their methodology, content, and teaching context.

Thematic analysis was organized around the four thematic areas of the review. Each article was placed in the section or sections to which it contributes without strict categorization, as some studies touch on more than one aspect of the topic.

### 2.2.4 Analysis

The analysis of the selected studies was guided by the main aim of this review: to assess the capacity of LLMs

to generate lesson plans with pedagogical value. Even in cases where the models were used for broader purposes (such as designing activities, supporting teacher reflection, or professional development), only the parts related to structured lesson planning were analyzed.

The analysis involved more than a simple description of the studies. Attention was given to the clarity of each study's methodology, the relationship between its findings and the research focus and its overall contribution to the review. The aim was to identify both the strengths and limitations of the research, especially when these factors affected the reliability or practical relevance of the conclusions. The analysis does not attempt to produce generalizations but instead seeks to understand different ways in which this emerging phenomenon has been approached. The analysis identified four main themes that appeared frequently across the literature: (1) quality and pedagogical value, (2) challenges and limitations, (3) teachers' attitudes and perceptions, and (4) improvements and future directions, as shown in Figure 3.

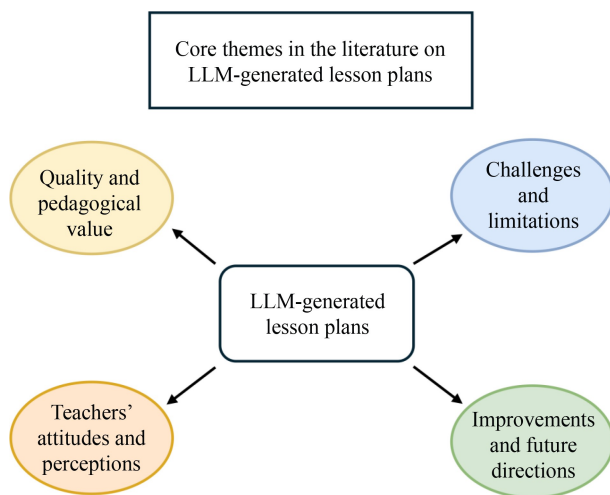


**Figure 3** Core themes identified in the literature on large language model (LLM)-generated lesson plans.

### 3 Findings

This section synthesizes the main findings of the review by identifying recurring patterns across studies and comparing areas of agreement and difference in the literature. The results are grouped into four themes that appeared frequently across the selected studies and represent different ways in which LLMs are used in lesson planning. Together, these themes address the research questions of this study. The theme, quality and pedagogical value, relates to RQ1, which concerns whether LLM-generated lesson plans demonstrate pedagogical value. The remaining themes—challenges and limitations, teachers' attitudes and perceptions, and

improvements and future directions, relate to RQ2, which examines the suitability of these plans for use in classroom settings.

#### 3.1 | Evaluation of the Quality and Pedagogical Value of LLM-Generated Lesson Plans

Across the reviewed studies, lesson-plan quality is primarily discussed in terms of structure, pedagogical depth, differentiation, time planning, and classroom applicability. These dimensions provide criteria for identifying both the strengths and limitations of LLM-generated lesson plans. One recurring issue in the literature concerns whether lesson plans created by LLMs can be considered high quality, complete, and pedagogically useful. Regarding structure, LLM-generated lesson plans are frequently described as well-organized with clear instructional formats. Most studies acknowledge that GenAI platforms, such as OpenAI's ChatGPT, Google's Gemini, Anthropic's Claude, and DeepSeek, can produce structured lesson plans that follow common instructional formats. Baytak (2024), for example, presented lesson plans in which activities are organized into brainstorming, group discussion, guided exploration of real-world problems, presentation, and evaluation stages. However, empirical evidence suggests that, although these plans are often clearly structured, they typically provide only a general framework and require adaptation before they can be effectively implemented in real classrooms (van den Berg & du Plessis, 2023). This lack of classroom readiness has also been examined in empirical studies using established evaluation frameworks. Lammert et al. (2024), for example, analyzed AI-generated lesson plans in relation to universal design for learning and universal design for transition frameworks that emphasize engagement, representation, and expression as basic dimensions of inclusive lesson design and reported limited correspondence with these principles. Demir and Ev Çimen (2024) similarly observed that LLM-generated lesson plans tend to follow basic structures but require adjustments regarding time management, activity design, and pedagogical strategies. Their study also illustrates how such lesson plans may be developed in practice. Using ChatGPT, the researchers initially generated a lesson plan based on the 5E (engage, explore, explain, elaborate, and evaluate) model and subsequently revised it by adjusting elements such as activities, duration, and structure through additional prompts. These findings suggest that producing LLM-generated lesson plans is not a one-step process but involves iterative refinement. Pre-service teachers adjust elements of the initial output through successive prompts, and the quality of the generated material depends partly on the clarity and specificity of the prompts.

Concerning classroom applicability, Cai (2024), van den Berg and du Plessis (2023), and Ivy et al. (2024) reported that although LLM-generated content can support lesson organization, it often does not fully address classroom-specific needs such as student needs and curriculum requirements. Therefore, pedagogical judgment and oversight are required to ensure that the generated material is relevant and pedagogically suitable for practical use.

In terms of pedagogical depth, teaching practices suggested by LLMs tend to promote procedural knowledge rather than conceptual understanding, omitting more complex forms of learning while also providing limited guidance on how content should be taught and offering little support for complex learning processes, such as reasoning, reflection, and justification (Cameron & Mesiti, 2024; Fan et al., 2024).

Regarding differentiation, several studies identify significant limitations, as the plans typically do not include strategies for individualization, nor do they offer alternatives for students with different learning profiles (Buchholtz & Huget, 2024; Clark & van Kessel, 2024; Kim et al., 2025). These limitations appear to be more pronounced in specialized contexts, where Bagus et al. (2024) reported that, although LLM-generated lesson plans may include creative ideas, they often fail to provide adequate differentiation strategies or appropriate assessment criteria, particularly when addressing the needs of students with intellectual disabilities.

Time planning, as one of the major dimensions of lesson-plan quality, also presents challenges. Yilmaz Can and Durmuş (2024), for example, reported that ChatGPT-generated plans assigned very short durations to essential tasks while giving more time to reminder stages, while one participant noted that a plan designed for one hour would require at least three hours in practice. These findings suggest that LLMs may produce unrealistic duration estimates (Sundin, 2024; Yilmaz Can & Durmuş, 2024).

Another important aspect of lesson-plan quality concerns consistency. Dornburg and Davin (2025) reported high variance in scores within the same prompt groups; for example, outputs generated from the same prompt differed by at least five rubric elements. This suggests that LLM-generated lesson plans may not produce stable or reliable results. As a result, at a practical level, lesson plans produced by LLMs often function more as drafts and require teacher editing and adaptation to be usable in the classroom (Deguara et al., 2024; Kalenda et al., 2025; Karaman & Goksu, 2024; Kohnke & Zou, 2025; Povey, 2025).

In addition, some variations across subject areas can also be observed, although clear comparisons are limited because the reviewed studies differ substantially in methodology, educational contexts, evaluation criteria, and the LLMs examined. In STEM contexts,

literature often reports advantages in terms of structure and lesson organization but also reports limitations related to pedagogical depth, conceptual support, and, in some cases, accuracy (Fan et al., 2024; Hu et al., 2024; Zheng et al., 2024). In humanities and language education, studies place greater emphasis on the generation of communicative activities and instructional materials while raising concerns about contextual fit, interpretation, and adaptation to learner needs (Dornburg & Davin, 2025; van den Berg & du Plessis, 2023). At the same time, many studies adopt cross-disciplinary designs, particularly in teacher education, pre-service teacher preparation, instructional design, and general lesson-planning contexts, while also differing considerably in their methodological and evaluation approaches. Furthermore, most studies focus on a single model, usually ChatGPT, making systematic comparisons across subjects or between different LLMs difficult.

Apart from differences across subjects, some variation in the reported quality of LLM-generated lesson plans is also observed in relation to the educational contexts. In particular, studies conducted in teacher education settings focus on how lesson plans are evaluated and refined, whereas classroom-based studies emphasize issues of implementation, adaptation, and correspondence with student needs.

Across the reviewed studies, higher-quality lesson plans are consistently associated with specific conditions, particularly the use of well-defined and iterative prompting, as well as active teacher involvement in adapting and evaluating the generated content. In some cases, the integration of LLMs within structured or guided systems further improves the coherence and pedagogical fit of the outputs (Demir & Ev Çimen, 2024; Fan et al., 2024; Hu et al., 2024; Kalenda et al., 2025). Hu et al. (2024), for example, used mathematical problem chains and a pedagogical content knowledge-based evaluation framework to guide GPT-4 teaching-plan generation, while Fan et al. (2024) developed Lesson Planner, a Gagné-based interactive system that improved lesson-plan quality compared with a baseline ChatGPT interface.

Overall, the evidence suggests that the pedagogical value of LLM-generated lesson plans is inconsistent across contexts; it depends on factors, such as prompt quality, teacher adaptation, and the extent to which the generated content responds to specific pedagogical and classroom needs.

### 3.2 | Challenges and Limitations

Several studies point to important obstacles and limitations in using LLMs for lesson planning. Although AI tools offer structure and help teachers save time by generating plans faster, there are concerns about both the reliability of their content and application in real learning environments.

One of the most frequently reported problems is the lack of adaptability. LLM-generated lesson plans often do not take into account students' level, the socio-cultural context of the classroom, or the specific needs of different learner groups (Karaman & Goksu, 2024; Kussin et al., 2024). Even when prompts include curriculum requirements or instructional objectives, the generated content may still require teacher revision to ensure that these elements are translated into appropriate classroom activities and coherent lesson plans (Corp & Revelle, 2023; Hu et al., 2024; Ivy et al., 2024). Several researchers also mention issues related to accuracy. The teaching suggestions provided by LLMs may include incorrect information, inaccurate content, or even fabricated sources (Kehoe, 2023; Özdemir, 2024; Powell & Courchesne, 2024). This means that teachers need to carefully verify the material before using it. In some cases, the generated plan includes activities or exercises that are overly general or unsuitable for the subject matter (Kloker et al., 2025; Moundridou et al., 2024).

Another point emphasized in the literature is that LLMs often struggle to incorporate pedagogical strategies. While they may produce a typical sequence of teaching steps, they do not take into account differentiation, goal setting, student engagement, and the development of critical thinking competence (Cameron & Mesiti, 2024; Pişkin Tunç, 2024). The suggested strategies often remain generic and do not include differentiated approaches or substantive adaptations to classroom profiling (Bagus et al., 2024; Kasneci et al., 2023).

Technical limitations are also present. There are documented issues related to the need for well-structured and precise prompts to generate useful content. Without this level of precision, LLMs find it difficult to respond to users' needs (Corp & Revelle, 2023; Karpouzis et al., 2024). In addition, evidence suggests that the use of a single prompt often leads to outputs that are not sufficiently accurate or realistic for classroom use, while better results require the use of successive prompts and refinements by users (Goodman et al., 2024). Moreover, there is a risk of overreliance on AI, especially among teachers or pre-service teachers with limited teaching experience or insufficient training in instructional design and the educational use of AI tools (Ogun et al., 2024; Setyaningsih et al., 2024).

Overall, LLMs may serve as useful support tools, but they currently lack the flexibility and pedagogical judgment required to produce fully appropriate and reliable lesson plans. These limitations suggest that LLM-generated lesson plans are not consistently suitable for direct classroom use without careful review and adaptation, as they do not fully meet the contextual and pedagogical demands of practice.

### 3.3 | Teachers' Attitudes and Perceptions

The general picture that emerges from the literature illustrates that most teachers acknowledge the usefulness of AI tools in lesson planning but, at the same time, express reservations about their pedagogical adequacy. Teachers use AI to generate content, but many still prefer to work collaboratively with colleagues and emphasize the need for clear policies to regulate its use in education (Kim et al., 2025). This suggests that teachers prefer to engage with LLMs within collaborative and guided professional contexts rather than treating them as fully autonomous tools for lesson planning (Gurl et al., 2025; Kim et al., 2025). Some educators see it as a helpful aid but do not believe it can replace teachers' pedagogical judgment (Kehoe, 2023; Koenig, 2024; Sun, 2024). Many agree that AI can assist with the initial organization of a lesson, but they stress that the final responsibility for adapting and applying the plan lies with teachers (Kalenda et al., 2025; Kanvaria & Ritika, 2024). Several educators indicated that AI cannot substitute for the experiential knowledge of the teacher and that LLM-generated plans do not always match the real conditions of the classroom (Cameron & Mesiti, 2024; Corp & Revelle, 2023; Karaman & Goksu, 2024).

There appears to be a difference in how teachers respond to AI, depending on their level of experience: Younger educators tend to be more open to using AI, while more experienced ones remain cautious (Mahtoum & Medjdoub, 2024). This difference in responses illustrates how teachers' perspectives are influenced by professional experience and familiarity with classroom practice. At the same time, many acknowledge that AI can be useful for generating ideas but point to the need for training and critical evaluation (Setyaningsih et al., 2024; Sundin, 2024). Relatedly, pre-service teachers in the study by Aydın Yıldız (2026) reported that ChatGPT-assisted lesson planning within a structured teacher-education process that included teacher training, weekly discussions, feedback, observations, and reflection reports, supported pre-service teachers' critical thinking, creativity, communication, and collaboration skills. The general trend is that while AI is seen as a supportive tool for planning, it is not something that can work independently without teacher guidance. From teachers' perspectives, LLM-generated lesson plans are often used as an initial basis for lesson preparation, with their effectiveness depending on the extent and quality of teacher adaptation.

### 3.4 | Improvements and Future Directions

Many studies suggest specific measures to improve the

reliability and pedagogical value of lesson plans generated by LLMs. A recurring finding is the need for structured teacher training. Within this discussion, teacher training emerges as a central aspect, with AI literacy referring to wider ability to critically use AI tools, and prompt design representing a more specific skill. Educators, therefore, need appropriate preparation, both to understand what AI can do and develop the ability to adapt and assess the content it produces (Kanvaria & Ritika, 2024; Kim et al., 2025; Setyaningsih et al., 2024).

Several authors recommended the development of clear policies and guidelines for the use of AI tools to ensure their responsible and educationally appropriate use (Koenig, 2024; Shamsuddin & Mohd Shariff, 2024; Sundin, 2024). The need for ongoing professional development and support was also emphasized so that teachers can make use of AI tools safely and with a critical mindset (Kussin et al., 2024; Mohammadi et al., 2024).

Several studies indicate the need to improve the adaptability of AI tools to classroom teaching and lesson planning by enhancing their features and allowing greater flexibility in instructional design. These features may include guided interfaces, partially structured prompts, lesson-planning functions, and options for teachers to provide contextual information about students, learning objectives, and standards during the planning process (Corp & Revelle, 2023; Moundridou et al., 2024; Sakamoto et al., 2024).

Special attention is given to the importance of improving prompt quality. Karpouzis et al. (2024) and Bagus et al. (2024) emphasized that the pedagogical value of the generated content depends directly on how clear, precise, and well-targeted the prompts are. The need for training in prompt design is mentioned in many studies as an important condition for using LLMs responsibly.

The importance of AI literacy is also emphasized in several studies as essential for the appropriate use of LLMs in education. Teachers need to be able to evaluate the generated content, identify errors or biases, and formulate suitable prompts (Cai, 2024; Fan et al., 2024; Powell & Courchesne, 2024). It has also been

suggested that this kind of training should be included in teacher preparation programs so that educators can use LLMs with awareness and critical thinking (Lee & Zhai, 2024; Povey, 2025). These findings indicate that the effective use of LLM-generated lesson plans in real classroom settings depends on teacher training, AI literacy, and the ability to design and refine prompts.

These points form four main directions that appear frequently in the literature, as shown in Figure 4: (1) teacher training and AI literacy, (2) prompt engineering and reflective use, (3) policy development and ethical guidelines, and (4) model improvement and pedagogical alignment.

## 4 Discussion

The overall picture that emerges from the literature suggests that using LLMs for lesson planning shows promise but also comes with many uncertainties. Four core themes indicate that there is no single consistent conclusion across the studies. Some research focuses on the potential benefits of LLMs, whereas other studies point to their limitations, and many stress the continued importance of teacher involvement in determining outcomes.

### 4.1 | Connections and Contradictions Across Four Core Themes

The analysis of four core themes reveals agreement and differences. Nearly all studies acknowledge that LLMs can produce lesson plans that are organized, structured, and goal-oriented. This corresponds with the generally positive attitude expressed by many teachers who view LLMs as helpful forms of support. However, most studies also indicate that the generated content cannot stand on its own. It needs revision, adaptation, and pedagogical judgment, which is consistent with findings concerning the limitations of LLMs.

A similar contradiction is observed between the third and fourth themes (teachers' attitudes and perceptions, and improvements and future directions). Many

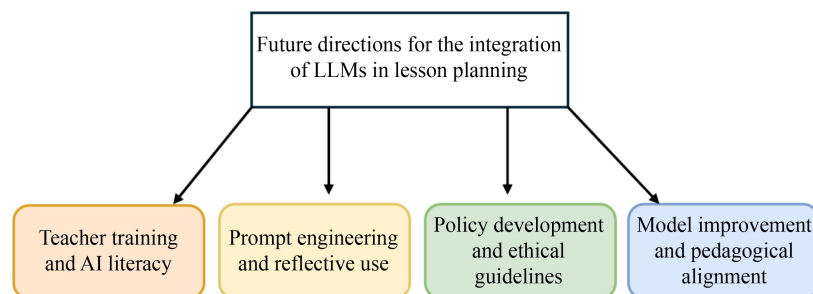


**Figure 4** Future directions for the integration of large language models (LLMs) in lesson planning.

teachers express a generally positive view of AI but, at the same time, raise concerns about how it is actually integrated into classroom practice. This helps explain why many of the future recommendations refer to the need for teacher training, support, and guidance. In other words, accepting these tools does not mean that they are ready to function independently. They still require well-designed structures, clear policies, and ongoing professional development.

Where there is consistent agreement is that the effective use of LLM-generated lesson plans depends on teacher involvement. Whether the focus is on lesson quality or on challenges related to classroom implementation, the reviewed studies frequently indicate that these plans require adaptation and pedagogical judgment. AI can provide material, but its educational value is defined by how it is applied by teachers.

Recent empirical evidence also points to a difference between perceived and actual quality. For example, Cooper (2026) reported that although teacher-created lesson plans were rated higher in quality overall, teachers were unable to reliably distinguish between LLM-generated and teacher-created plans. Zheng et al. (2024), in contrast, found that refined LLM-generated mathematics lesson plans received high evaluation scores and, in several cases, outperformed teacher-created lesson plans, although teacher-generated lesson procedures were often more closely consistent with classroom practice. This suggests that differences in pedagogical quality may not always be clearly identifiable in practice. Similarly, Dornburg and Davin (2025) reported that outputs generated from identical prompts can vary considerably, while increasing prompt specificity does not necessarily result in more consistent or higher-quality lesson plans.

#### 4.2 | Findings in Relation to the Research Questions

The first part of the central question concerns the pedagogical value of lesson plans produced by LLMs (Do LLMs generate lesson plans with genuine pedagogical value?). Studies suggest that these plans offer pedagogical value, mainly in their structure. They provide organization, save time, and include ideas for activities, and in several cases they work as a starting point on which teachers can build. Even so, no study argues that these plans are complete on their own. Most of the literature shows deficiencies in differentiation, links to curriculum goals, and overall detail, and the content often needs to be reviewed and adjusted before it can be considered appropriate for classroom use.

The second part of the question relates to the suitability of these plans for actual classroom use (To what extent are these plans suitable for use in classroom settings?). Here, the findings are more consistent

across studies. The plans can help with preparation, but they are not ready for implementation without substantial changes. Many studies note that teachers have to revise the content, adapt activities to their students, and add practical details, such as time allocation or how specific steps will be carried out. In this sense, the usefulness of the plans depends directly on teachers' contribution. LLMs can support the planning process, but they do not replace it, and pedagogical judgment and knowledge of the class remain essential.

#### 4.3 | Research Gaps in the Use of LLMs for Lesson Planning

Despite the growing number of publications on LLMs and their use in lesson planning, the literature still leaves four important questions about their practical use unanswered. The first issue concerns the fact that most studies remain theoretical or rely on teachers' opinions without examining what happens when these plans are actually applied in the classroom. In particular, there is limited evidence from studies that examine the use of LLM-generated lesson plans in actual classroom teaching settings and their impact on student learning, and relatively few studies consider both teacher and student perspectives.

Although some initial comparisons are beginning to appear (Cooper, 2026; Zheng et al., 2024), empirical studies that directly compare LLM-generated and teacher-created lesson plans remain limited, partly because such research involves their use in practice, which often requires further teacher evaluation and revision before implementation and may raise practical and ethical concerns, such as output accuracy, bias, privacy, and the use of student-related classroom information (Corp & Reville, 2023; Moundridou et al., 2024; Sakamoto et al., 2024). Further investigation of this issue could help clarify how useful these tools are for teaching.

The second issue concerns the type of data being used. Most studies are based on qualitative observations, interviews, or questionnaires. Only a limited number of studies have placed teacher-created and LLM-generated lesson plans side by side to examine differences in quality or outcomes in a systematic way.

The third issue concerns geographical concentration. In this review, country was classified according to the institutional affiliation of the first author of each study. The literature covered a range of countries, although the United States and Türkiye appeared most frequently. This limits the potential to draw general conclusions and reveals the need for greater geographical diversity in future research.

The fourth issue concerns the limited attention given to the processes through which teachers interact with AI systems during lesson planning, particularly in

relation to prompt construction and the evaluation of generated outputs. While most studies focus on the quality of the generated plans or teachers' perceptions, little is known about how these outputs are produced and used in practice. Some recent work has begun to explore this direction, but systematic evidence on how teachers construct prompts and interact with LLM-generated outputs remains limited (Corp & Revelle, 2023; Dornburg & Davin, 2025; Fan et al., 2024).

Overall, the literature gives a fairly clear picture: LLMs can support lesson planning, but they are not yet capable of taking on this role independently. Their use makes sense when they function as support tools and not as substitutes for pedagogical judgment. The need for human involvement remains a consistent theme across all areas. At the same time, research needs to move into more empirical and diverse areas to better understand when, how, and under what conditions LLMs can meaningfully contribute. Until then, their use should remain guided, intentional, and pedagogically aware.

#### 4.4 | Limitations of the Present Review

As this is still a relatively new area of research, the available literature is limited both in scope and in depth. This review focuses exclusively on studies that deal with the use of LLMs for lesson planning. It does not include research on the wider use of AI in education or in other areas of instructional design. Because research on the use of LLMs for lesson planning is still limited, non-peer-reviewed sources (mainly preprints and conference papers) were also included. They were selected due to their thematic relevance and were used only to complement the peer-reviewed studies, with careful consideration of their reliability. Another limitation concerns the geographical and thematic range of the sources. The available literature is unevenly distributed across countries or regions and educational contexts. As a result, the findings may not be fully representative of other education systems, subject areas, or classroom realities. In addition, most of the sources are theoretical or based on teachers' attitudes and perspectives without empirical evidence from real classroom implementation. Comparative studies that directly examine the differences between teacher-created and LLM-generated plans remain limited. These factors should be considered when interpreting findings and drawing conclusions.

A further limitation concerns the search strategy. Although multiple databases were consulted and additional searches were conducted using LLM-related terms, it remains possible that some relevant studies were not captured due to differences in terminology, indexing practices, or publication status.

## 5 Conclusions

This review contributes to the field by mapping current evidence on how LLMs are used for lesson planning and by identifying both practical uses and important deficiencies that need to be addressed in future research.

The overall picture emerging from the literature suggests that LLMs can serve as useful tools for lesson planning, but they are not yet capable of replacing the teacher's pedagogical role. Although lesson plans produced by LLMs are usually well-structured, organized, and goal-oriented, they tend to fall short in adaptation, differentiation, and pedagogy. The need for human intervention, adjustment, and critical evaluation appears across all thematic areas.

At the same time, the positive attitude expressed by many teachers indicates that LLMs can be incorporated into the planning process as long as they are accompanied by appropriate training and support. The discussion on AI in instructional design is still developing and calls for more research, especially in real classroom settings, across different subjects and educational contexts.

This review does not offer a definitive answer to the question of whether LLM-generated lesson plans are pedagogically appropriate, but it does make clear that their use can be useful when placed within a framework of critical engagement, with the teacher remaining at the center of planning process.

To make full use of the potential of LLMs in lesson planning, more systematic research, ongoing teacher training, and policies that frame the use of AI within sound educational practices are needed. The next steps concern both improving the models themselves and preparing educators to use them in a thoughtful, flexible, and responsible way. In particular, future studies should further examine the relative quality and practical effectiveness of lesson plans created by LLMs and teachers in real classroom contexts. Future research may also explore how the quality and applicability of LLM-generated lesson plans vary across subject areas, educational levels, teacher experience, and different LLMs.

Although a small number of comparative studies have recently begun to emerge, the available evidence remains limited and does not yet allow firm conclusions regarding the effectiveness of LLM-generated lesson plans across different educational contexts. For this reason, LLMs should currently be viewed as support for lesson planning rather than as ready-made planning tools and should not yet be considered a substitute for pedagogical expertise.

**Acknowledgment** The authors received no financial support for the research, authorship, and/or publication of this article.

**Declaration on the Use of AI** The authors declare that ChatGPT (OpenAI; model versions varied during manuscript preparation) was used only for language revision and rephrasing. All AI-assisted text was carefully reviewed and edited by the authors, who take full responsibility for the final content of the work.

**Author Contributions** Vassilis A. Failadis: Conceptualization, methodology, screening, data curation, formal analysis, writing—original draft, writing—review & editing, visualization. Sotiris K. Tasoulis: Formal analysis, writing—original draft, writing—review & editing, visualization. Spiros V. Georgakopoulos: Formal analysis, writing—original draft, writing—review & editing, visualization. Vassilis P. Plagianakos: Supervision, formal analysis, writing—original draft, writing—review & editing, visualization.

**Conflict of Interest** The authors declare that they have no competing interests.

**Data Availability Statements** The authors confirm that all data generated or analyzed during this study are included in this published article.

**Electronic Supplementary Material** Supplementary material is available in the online version of this article at <https://doi.org/10.1007/s44366-026-0099-6> and is accessible for authorized users.

## References

- Aydın Yıldız, T. (2026). Exploring the impact of ChatGPT on improving 21st-century skills for future English teachers during lesson planning. *Computers in the Schools*, 43(3), 302–325.
- Bagus, I. G., Phalaguna, G., Kaewsaeng, K., & Worabuttara, T. (2024). Exploring teachers' perceptions of AI-generated English lesson plans for students with intellectual disabilities. *International Journal of Instructions and Language Studies*, 2(2), 19–28.
- Baytak, A. (2024). The content analysis of the lesson plans created by ChatGPT and Google Gemini. *Research in Social Sciences and Technology*, 9(1), 329–350.
- Bessas, N., Tzanaki, E., Vavougiou, D., & Plagianakos, V. P. (2025a). Comparative analysis of ChatGPT and Gemini; Implications for junior high school physics education: Opportunities and ethical challenges. *International Journal of Advanced Multidisciplinary Research and Studies*, 5(1), 7–18.
- Bessas, N., Tzanaki, E., Vavougiou, D., & Plagianakos, V. P. (2025b). The role of ChatGPT in junior high school physics education: Insights from teachers and students and guidelines for optimal use. *Social Sciences & Humanities Open*, 11, 101610.
- Buchholtz, N., & Huget, J. (2024). ChatGPT as a reflection tool to promote the lesson planning competencies of pre-service teachers. In: Faggiano, E., Clark-Wilson, A., Tabach, M., & Weigand, H. G., eds., *Proceedings of the 17th ERME Topic Conference MEDA 4*. Bari: University of Bari Aldo Moro, 129–136.
- Cai, W. X. (2024). *The potential of AI chatbots in enhancing lesson planning in K-12 education-insights from practicum-experienced teaching graduate students*. Thesis for the Master's Degree. Toronto: University of Toronto.
- Cameron, S., & Mesiti, C. (2024). What kind of mathematics teacher is ChatGPT? Identifying the pedagogical practices preferred by generative AI tools when preparing lesson plans. In: *Proceedings of the 46th Annual Conference of the Mathematics Education Research Group of Australasia*. Gold Coast: MERGA, 135–142.
- Clark, C. H., & van Kessel, C. (2024). "I, for one, welcome our new computer overlords": Using artificial intelligence as a lesson planning resource for social studies. *Contemporary Issues in Technology and Teacher Education*, 24(2), 151–183.
- Cooper, P. K. (2026). Music teachers' labeling accuracy and quality ratings of lesson plans by artificial intelligence (AI) and humans. *International Journal of Music Education*, 44(1), 23–36.
- Corp, A., & Revelle, C. (2023). ChatGPT is here to stay: Using ChatGPT with student teachers for lesson planning. *The Texas Forum of Teacher Education*, 14, 116–124.
- Deguara, D., Gatt, T., Said Pullicino, G., & Scerri, D. (2024). UPP: An investigation into the effectiveness of ChatGPT-generated lesson plans in vocational education. *MCAST Journal of Applied Research & Practice*, 8(2), 4–15.
- Demir, G., & Ev Çimen, E. (2024). Lesson plan preparation process with ChatGPT in mathematics teaching: An example of research and practice. *Technology, Innovation and Special Education Research Journal*, 4(1), 83–98.
- Dornburg, A., & Davin, K. (2025). ChatGPT in foreign language lesson plan creation: Trends, variability, and historical biases. *ReCALL*, 37(3), 332–347.
- Fan, H. X., Chen, G. Z., Wang, X. B., & Peng, Z. H. (2024). LessonPlanner: Assisting novice teachers to prepare pedagogy-driven lesson plans with large language models. In: *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. Pittsburgh: ACM, 146.
- Goodman, J., Handa, V., Wilson, R. E., & Bradbury, L. U. (2024). Promises and pitfalls: Using an AI chatbot as a tool in 5E lesson planning. *Innovations in Science Teacher Education*, 9(1).
- Gurl, T. J., Markinson, M. P., & Artzt, A. F. (2025). Using ChatGPT as a lesson planning assistant with preservice secondary mathematics teachers. *Digital Experiences in Mathematics Education*, 11, 114–139.
- Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLoS ONE*, 10(9), e0138237.
- Hu, B. H., Zheng, L. W., Zhu, J. Y., Ding, L. S., Wang, Y. L., &

- Gu, X. Q. (2024). Teaching plan generation and evaluation with GPT-4: Unleashing the potential of LLM in instructional design. *IEEE Transactions on Learning Technologies*, 17, 1445–1459.
- Ivy, C., Busse, A., & Ramos, J. (2024). Lesson planning and assessment: Exploring the integration of AI tools in mathematics teacher preparation. In: Gottfert, M. B., Krug, F., Nerowski, C., & Bruder, R., eds. *Proceedings of the 17th ERME Topic Conference MEDA 4*. Darmstadt: Technische Universität Darmstadt, 207–214.
- Kalenda, P. J., Rath, L., Heidt, M. A., & Wright, A. (2025). Pre-service teacher perceptions of ChatGPT for lesson plan generation. *Journal of Educational Technology Systems*, 53(3), 219–241.
- Kanvaria, V. K., & Ritika. (2024). Exploring the integration of artificial intelligence in lesson planning for pre-service teachers. *IJET (NCERT)*, 6(2), 340–345.
- Karaman, M. R., & Goksu, I. (2024). Are lesson plans created by ChatGPT more effective? An experimental study. *International Journal of Technology in Education*, 7(1), 107–127.
- Karpouzis, K., Pantazatos, D., Taouki, J., & Meli, K. (2024). Tailoring education with GenAI: A new horizon in lesson planning. *arXiv Preprint*, arXiv:2403.12071.
- Kasnezi, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., & et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.
- Kehoe, F. (2023). Leveraging generative AI tools for enhanced lesson planning in initial teacher education at post primary. *Irish Journal of Technology Enhanced Learning*, 7(2), 172–182.
- Kim, D., Jung, J., & Sheumaker, M. F. (2024). Teacher education students' perceptions of ChatGPT for lesson planning: Benefits, challenges, and use. In: *Proceedings of Society for Information Technology & Teacher Education International Conference*. Orlando: AACE, 756–761.
- Kloker, S., Bukoli, H., & Kateete, T. (2025). New curriculum, new chance—retrieval augmented generation for lesson planning in Ugandan secondary schools. Prototype quality evaluation. *Ndejje University Journal of Interdisciplinary Studies*, 1(1), 1–5.
- Koenig, S. H. (2024). *Fine-tuning for lesson planning: Comparing teachers' perceptions and use of a fine-tuned AI assistant for lesson planning*. Thesis for the Master's Degree. Gothenburg: University of Gothenburg.
- Kohnke, L., & Zou, D. (2025). The role of ChatGPT in enhancing English teaching: A paradigm shift in lesson planning and instructional practices. *Educational Technology & Society*, 28(3), 4–20.
- Kussin, H. J., Megat Khalid, P. Z., Chaniago, R. H., Moneyam, S., & Hassim, H. E. (2024). The future of lesson planning: AI integration experiences among TESL teacher trainees in a Malaysian public university. *The Asian Journal of English Language and Pedagogy*, 12(2), 178–189.
- Lammert, C., DeJulio, S., Grote-Garcia, S., & Fraga, L. (2024). Better than nothing? An analysis of AI-generated lesson plans using the universal design for learning & transition frameworks. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 97(5), 168–175.
- Lee, G.-G., & Zhai, X. M. (2024). Using ChatGPT for science learning: A study on pre-service teachers' lesson planning. *IEEE Transactions on Learning Technologies*, 17, 1643–1660.
- Mahtoum, N., & Medjdoub, A. (2024). *Investigating the use of ChatGPT by EFL teachers in lesson planning: Case study*. Thesis for the Master's Degree. Guelma: Université 8 Mai 1945 Guelma.
- Mengist, W., Soromessa, T., & Legese, G. (2020). Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX*, 7(2), 100777.
- Mohammadi, L., Asadi, M., & Taheri, R. (2024). Transforming EFL lesson planning with 'to teach AI': Insights from teachers' perspectives. *Technology Assisted Language Education*, 2(3), 46–73.
- Moundridou, M., Matzakos, N., & Doukakis, S. (2024). Generative AI tools as educators' assistants: Designing and implementing inquiry-based lesson plans. *Computers and Education: Artificial Intelligence*, 7, 100277.
- Ogun, A. O., Otujo, C. O., & Odufeko, G. T. (2024). ChatGPT a veritable tool in lesson planning for sustainable education in the 21st century and beyond: Pre-service teachers' preparedness in the usage of digital tool in science. *Journal of Education in Developing Areas*, 32(2), 91–100.
- Özdemir, S. (2024). Getting support from Microsoft Copilot in lesson plan preparation: Pre-service teachers' experiences and opinions. *Information Technologies and Learning Tools*, 104(6), 180–196.
- Pişkin Tunç, M. (2024). Examining pre-service mathematics teachers' purposes of using ChatGPT in lesson plan development. *Sakarya University Journal of Education*, 14(2), 391–406.
- Povey, E. (2025). Leveraging AI for TESOL: Grammar lesson planning. *Global Scientific and Academic Research Journal of Education and Literature*, 3(1), 11–25.
- Powell, W., & Courchesne, S. (2024). Opportunities and risks involved in using ChatGPT to create first grade science lesson plans. *PLoS ONE*, 19(6), e0305337.
- Sakamoto, M., Tan, S., & Clivaz, S. (2024). Social, cultural and political perspectives of generative AI in teacher education: Lesson planning in Japanese teacher education. In: Searson, M., Langran, E., & Trumble, J., eds. *Exploring new horizons: Generative artificial intelligence and teacher education*. Association for the Advancement of Computing in Education, 178–208.
- Setyaningsih, E., Asrori, M., Ngadiso, Sumardi, Zainnuri, H., & Hariyanti, Y. (2024). Exploring high school EFL teachers' experiences with magic school AI in lesson planning: Benefits and insights. *Voices of English Language Education Society*, 8(3), 685–700.
- Shamsuddin, W. N. F. W., & Mohd Shariff, N. A. (2024). Can AI create better lesson plans than humans? This is what Malaysian English instructors think. *Advance e-Research*

- Magazine*, 5, 62–68.
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339.
- Sun, F. (2024). Unlocking the potential: Exploring student teachers' perceptions of human-generated and ChatGPT lesson plans in education. In: *Proceedings of the 2024 10th International Conference on Humanities and Social Science Research*. Xiamen: Atlantis Press, 1413–1421.
- Sundin, M. (2024). *Unlocking artificial potential: Investigating the effectiveness of Chat-GPT as an operational tool for lesson planning*. Thesis for the Master's Degree. Örebro: Örebro University.
- van den Berg, G., & du Plessis, E. (2023). ChatGPT and Generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences*, 13(10), 998.
- Yılmaz Can, D., & Durmuş, C. (2024). From AI-generated lesson plans to the real-life classes: Explored by pre-service teachers. In: *Proceedings of the 10th International Conference on Higher Education Advances*. Valencia: Universitat Politècnica de València.
- Zheng, Y., Li, X. Y., Huang, Y. Y., Liang, Q. R., Guo, T., Hou, M. L., Gao, B. Y., Tian, M., Liu, Z. T., & Luo, W. Q. (2024). Automatic lesson plan generation via large language models with self-critique prompting. In: *Proceedings of the 25th International Conference on Artificial Intelligence in Education*. Cham: Springer, 163–178.