Journal of Information Technology Education: Research

An Official Publication of the Informing Science Institute
InformingScience.org

JITEResearch.org

Volume 25, 2026

USE OF PROMPT ENGINEERING FOR TEACHING AND LEARNING IN SECONDARY EDUCATION: LITERATURE REVIEW AND RESEARCH PERSPECTIVE

Luca Addiucci*	Sapienza University of Rome, Rome, Italy	luca.addiucci@uniroma1.it
Marco Temperini	Sapienza University of Rome, Rome, Italy	marco.temperini@uniroma1.it

* Corresponding author

ABSTRACT

Aim/Purpose	To address the lack of a clear pedagogical framing of prompt engineering in secondary education and to analyze how it is currently conceptualized in educational research.
Background	While secondary school students increasingly use generative AI tools, prompt engineering is often treated as an implicit technical skill rather than as an explicit educational practice linked to metacognition and AI literacy.
Methodology	This study adopts a systematic literature review following PRISMA guidelines. Twenty-one peer-reviewed studies published between 2021 and 2025 were selected from Scopus, Web of Science, IEEE Xplore, and ACM Digital Library.
Contribution	The paper provides a structured conceptual mapping of how prompt engineering is addressed in secondary education and identifies gaps between research practices, pedagogical frameworks, and AI literacy policies.
Findings	The review shows that prompt engineering is rarely framed as an explicit learning objective, that empirical evidence on cognitive and metacognitive effects is fragmented, and that ethical and reflective dimensions are inconsistently addressed.
Recommendations for Practitioners	Teachers should explicitly scaffold students' prompt design practices, integrate reflective activities on AI use, and align classroom practices with emerging AI literacy frameworks.

Accepting Editor Erda Wati Bakar │ Received: January 7, 2026 │ Revised: March 28. April 9, April 22, 2026 │ Accepted: April 23, 2026.

Cite as: Addiucci, L., & Temperini, M. (2026). Use of prompt engineering for teaching and learning in secondary education: Literature review and research perspective. *Journal of Information Technology Education: Research, 25,* Article 18. https://doi.org/10.28945/5773

Recommendations for Researchers	Future studies should operationalize prompt engineering as a learning objective, develop validated assessment tools, and conduct longitudinal research in secondary education contexts.
Impact on Society	The findings support the development of responsible AI use in education by highlighting the need for pedagogical guidance that fosters students' agency, critical thinking, and ethical awareness.
Future Research	Future research should investigate instructional models for teaching prompt engineering across disciplines and examine its long-term effects on students' learning strategies and epistemic beliefs.
Keywords	prompt engineering, generative AI, secondary education, AI literacy, systematic literature review

INTRODUCTION

This systematic review examines how prompt engineering is conceptualized and implemented in secondary education, positioning it as an emerging component of AI literacy and metacognitive competence. As large language models (LLMs) become widely accessible to adolescents, students increasingly interact with generative AI systems, often without explicit instruction on how to formulate, refine, and critically evaluate prompts. For example, classroom interventions include students iteratively revising prompts to improve argumentative writing or using structured prompting frameworks to guide problem-solving in STEM contexts. Drawing on 21 peer-reviewed studies published between 2021 and 2025, this review maps how prompt engineering is framed pedagogically, the cognitive or metacognitive effects reported, and the instructional approaches that support responsible AI use. Results indicate that 67% of the studies explicitly integrate prompt engineering into educational activities, while ethical and responsible use considerations are fully addressed in only 38% of the reviewed research. The findings reveal a clear misalignment between the rapid diffusion of LLM use in secondary education and the limited development of coherent pedagogical and assessment frameworks. By synthesizing current practices and identifying structural gaps, this review argues that prompt engineering should be treated not merely as a technical interaction skill, but as a core educational competence requiring explicit instructional design, assessment criteria, and alignment with emerging AI literacy policies.

The rapid diffusion of large language models (LLMs) in educational contexts has transformed how students access information, generate content, and engage with academic tasks. Since the public release of ChatGPT in 2022, generative AI systems have become widely accessible to adolescents, reshaping classroom practices and study habits. Within this evolving landscape, prompt engineering – understood as the deliberate formulation, structuring, and iterative refinement of inputs to guide LLM outputs – has emerged as a critical mediating practice in human–AI interaction. In secondary education, students increasingly interact with LLMs to support writing, problem-solving, programming, and conceptual understanding. However, while these interactions are becoming widespread, prompt engineering is rarely framed as an explicit instructional objective. Instead, it often remains an implicit or informal skill embedded in task completion rather than being conceptualized as a competence requiring structured teaching, reflection, and assessment. This absence of pedagogical framing raises concerns for AI literacy, epistemic agency, and responsible AI use.

The challenge is not merely technological but educational. When students use LLMs without explicit guidance on how prompts shape outputs, they may adopt passive or uncritical interaction patterns, rely on opaque system responses, or fail to reflect on the assumptions embedded in their queries. In this sense, prompt engineering can be interpreted not simply as a technical skill but as a metacognitive and socio-technical practice through which learners negotiate control, feedback, and knowledge construction in AI-mediated environments.

Despite the growing body of research on generative AI in education, existing studies remain fragmented in how they conceptualize and operationalize prompt engineering, particularly at the secondary school level. Prior reviews have examined AI in K–12 education, ChatGPT in classrooms, or prompt engineering in higher education contexts, yet none have systematically mapped how prompt engineering is framed, taught, and assessed in secondary education as a distinct educational practice.

This systematic literature review addresses this gap. Focusing on peer-reviewed studies published between 2018 and 2025 (with included studies spanning 2021–2025), the review examines how prompt engineering is integrated into secondary educational contexts, the cognitive or metacognitive effects reported, the pedagogical approaches that sustain its use, and the challenges that emerge regarding responsible AI interaction.

Specifically, the review addresses the following research questions (RQs):

RQ1: In what ways is prompt engineering addressed in formal or informal educational activities targeting high school students?

RQ2: What evidence is reported on the cognitive or metacognitive effects of prompt engineering on student learning in secondary schools?

RQ3: What pedagogical approaches, disciplinary contexts, or tools are used to introduce or sustain the use of LLMs as support for learning in secondary education?

RQ4: What challenges, limitations, or educational needs emerge in training students – and potentially teachers – to use LLMs responsibly in high school settings?

By clarifying how prompt engineering is currently positioned within secondary education research, this review aims to contribute to the development of coherent pedagogical frameworks aligned with emerging AI literacy initiatives and responsible AI policies. We structured these RQs according to the PICOC framework (Kitchenham & Charters, 2007). Table 1 summarizes the PICOC elements, while Table 2 maps each research question to the corresponding PICOC component.

Table 1. Elements of the PICOC framework

Element	Meaning
P	Population: Who is involved?
I	Intervention: What is studied or applied?
C	Comparison: What is the intervention compared to (optional)?
O	Outcome: What effects, results, or measures are considered?
C	Context: the setting in which the study takes place (e.g., educational, disciplinary)?

Table 2. Mapping of research questions to PICOC elements

PICOC	RQ1	RQ2	RQ3	RQ4
P (Population)	High school students	High school students	High school students or teachers	High school students
I (Intervention)	Prompt engineering as a concept or practice	Use of prompt engineering	Introduction or use of LLMs	Training for responsible use of LLMs

PICOC	RQ1	RQ2	RQ3	RQ4
C (Comparison)	Not applicable	Possibly traditional approaches	Not applicable	Not applicable
O (Outcome)	Ways of addressing prompt engineering	Learning outcomes and skills acquisition	Educational strategies, tools, and contexts	Challenges, limitations, and open research issues
C (Context)	Secondary school education	Educational activities with LLMs	Teaching and learning environments	Secondary education

Conceptual Framing of Prompt Engineering as an Educational Practice

In the context of secondary education, prompt engineering can be conceptualized not merely as a technical skill for interacting with large language models but as an educational practice that mediates students' engagement with knowledge, feedback, and epistemic authority. From this perspective, prompt formulation shapes how learners frame problems, articulate intentions, and evaluate the outputs generated by AI systems. Prompt engineering thus operates at the intersection of cognitive, metacognitive, and socio-technical dimensions of learning. This review adopts a pedagogical and conceptual lens, interpreting prompt engineering as a situated learning practice rather than a neutral interface operation. When students interact with generative AI through prompts, they implicitly negotiate assumptions about correctness, explanation, and agency. The quality, structure, and intent of prompts influence not only the relevance of AI-generated responses but also students' opportunities for reflection, iterative reasoning, and self-regulation. Framing prompt engineering as an educational practice enables clearer alignment with emerging constructs such as AI literacy, digital competence, and metacognitive awareness. Within secondary education, where students are still developing stable study habits and epistemic beliefs, making prompt engineering explicit and pedagogically guided becomes essential. This conceptual framing provides the basis for the analysis conducted in this review and informs the interpretation of the selected studies across disciplinary and instructional contexts.

Theoretical Grounding in Learning Sciences

From a learning sciences perspective, the interpretation of prompt engineering as a metacognitive and socio-technical competence can be grounded in established theories of metacognition and self-regulated learning. Metacognition refers to the processes through which learners monitor and regulate their own cognitive activity, including planning strategies, evaluating task performance, and adjusting actions in response to feedback (Flavell, 1979). Within educational research, these processes have been further articulated in models of self-regulated learning, where learners actively manage their learning through cycles of planning, monitoring, and reflection (Zimmerman, 2002). Interactions with large language models can be interpreted within this framework as a form of externally mediated cognitive regulation. When students formulate prompts, evaluate generated outputs, and iteratively refine their queries, they engage in regulatory processes similar to those described in self-regulated learning models. In particular, prompt construction requires learners to articulate task goals, structure information requests, and assess the adequacy of responses, thereby activating planning and monitoring mechanisms.

Iterative prompt revision further resembles the feedback loops described in models of self-regulated learning, where learners adjust strategies in response to outcomes (Winne & Hadwin, 1998). From this perspective, prompt engineering can be understood not merely as a technical interaction skill but as a metacognitive regulation practice embedded within human–AI interaction. In AI-supported learning environments, prompts act as mediating artifacts through which learners structure inquiry, negotiate epistemic authority, and regulate their interaction with algorithmic systems. This interpretation aligns prompt engineering with broader constructs such as epistemic agency and AI literacy, emphasizing the importance of reflective and intentional engagement with generative technologies in educational contexts.

Operationalization of Prompt Engineering as an Educational Construct

To strengthen conceptual clarity in this review, prompt engineering is operationalized as a set of observable practices through which learners regulate their interaction with generative AI systems. Specifically, four core dimensions were considered in interpreting the reviewed studies. First, prompt planning refers to the formulation of structured queries that specify task goals, constraints, and contextual information before interacting with the AI system. Second, iterative refinement is the process of modifying prompts based on system outputs to obtain more relevant or accurate responses. Third, evaluation of AI outputs involves the critical assessment of generated content for correctness, relevance, and completeness. Fourth, reflective awareness of system limitations, which includes recognizing potential issues such as hallucinations, bias, or over-reliance on automated responses. The operational dimensions considered in this review are summarized in Table 3.

Table 3. Operational dimensions of prompt engineering in educational contexts

Dimension	Description	Observable indicators in educational settings
Prompt planning	The learner formulates a structured request that clarifies the task, context, and expected output before interacting with the AI system.	Explicit definition of task goals, specification of constraints, and inclusion of contextual information in prompts.
Iterative refinement	The learner modifies and improves prompts after observing the system's response in order to obtain more relevant or accurate outputs.	Successive prompt adjustments, reformulation of questions, and clarification of instructions following unsatisfactory responses.
Evaluation of AI outputs	The learner critically assesses the quality and reliability of the generated responses.	Checking the correctness of information, comparing AI responses with external sources, and identifying incomplete or misleading outputs.
Reflective awareness of system limitations	The learner demonstrates awareness of potential limitations of AI systems and adjusts interaction strategies accordingly.	Recognition of hallucinations, discussion of bias or uncertainty, and explicit reflection on when AI outputs should be verified or rejected.

These dimensions allow distinguishing between studies in which prompting is merely embedded within system design and those in which prompt construction is explicitly addressed as part of the learning process. By operationalizing prompt engineering in this way, the review provides a clearer analytical lens for examining how prompting practices are integrated into educational interventions and how they relate to cognitive, metacognitive, and ethical dimensions of learning.

PLANNING OF THIS LITERATURE REVIEW

This section outlines the main components of the planning phase for this Systematic Literature Review (SLR):

(1) analysis of existing review studies,
(2) formulation of Research Questions (RQs),
(3) selection of data sources,
(4) definition of search strings, and
(5) specification of inclusion and exclusion criteria.

RELATED REVIEW WORK

The growing integration of generative AI in education has prompted several systematic and bibliometric reviews. However, these studies differ significantly in scope, educational level, and conceptual framing. Existing reviews generally fall into three broad categories: (1) reviews focused on prompt engineering within higher education curricula, (2) reviews examining AI or ChatGPT adoption across K–12 contexts without isolating prompt engineering as an educational construct, and (3) broader bibliometric mappings of AI in science or STEM education. For instance, Lee and Palmer (2025) analyze prompt engineering in higher education, identifying instructional models and curricular implications, yet their scope does not extend to adolescent learners or secondary schooling contexts. Chen et al. (2024) review prompt engineering in K–12 STEM education, but their emphasis remains primarily disciplinary and technical, rather than pedagogical or metacognitive. Similarly, Dimeli and Kostas (2025) synthesize research on ChatGPT in school and university education, highlighting ethical concerns and classroom practices; however, prompt engineering is not examined as a distinct educational competence. Bibliometric reviews, such as those by Kavitha and Joshith (2024), map AI integration in science education but do not specifically address how learners are trained to construct, evaluate, or refine prompts when interacting with LLMs.

Across these reviews, prompt-related practices are typically embedded within broader discussions of AI adoption, system implementation, or disciplinary innovation. What remains underexplored is the explicit pedagogical framing of prompt engineering as a reflective and assessable competence in secondary education. From a learning sciences perspective, prompt formulation can be interpreted as a form of metacognitive regulation, involving planning, monitoring, and revising interactions with AI systems. From a critical pedagogical standpoint, prompts mediate epistemic authority and shape how knowledge is constructed and validated in AI-supported environments. Yet these dimensions are rarely foregrounded in existing syntheses. This review, therefore, addresses a specific and underexamined gap: it systematically maps how prompt engineering is conceptualized, operationalized, and assessed in secondary education research, with particular attention to its cognitive, metacognitive, and ethical implications.

SEARCH STRATEGY AND DATA SOURCE

This systematic literature review was conducted following PRISMA 2020 guidelines (Page et al., 2021) and methodological recommendations for systematic reviews in software engineering and educational technology (Kitchenham & Charters, 2007). The search covered studies published between January 2018 and June 2025. The database search was conducted in June 2025. Four curated academic databases were selected to ensure quality and reproducibility:

- Scopus
- Web of Science
- IEEE Xplore
- ACM Digital Library

Although Google Scholar provides broad coverage of academic literature, previous methodological studies have noted limitations in search transparency, metadata consistency, and the reproducibility of results in systematic review contexts (Gusenbauer, 2024; Martín-Martín et al., 2018; Meho & Yang, 2007). Given the need for controlled indexing criteria and replicable filtering procedures, this review relied exclusively on curated bibliographic databases.

The same search string was used across all databases, with minor syntactic adaptations as required by each database's query format. The search string was:

("prompt engineering" **OR** "prompt design" **OR** "prompt strategies")
AND ("high school" **OR** "secondary education" **OR** "secondary school" **OR** "K–12")
AND ("large language models" **OR** "LLM" **OR** "ChatGPT" **OR** "generative AI")
AND ("teaching" **OR** "education" **OR** "training" **OR** "learning")

Study Selection Process

The initial database search returned 703 records. After automatic deduplication using Zotero, 659 unique records remained. Following the exclusion of non-peer-reviewed proceedings and non-indexed workshop abstracts, 411 records were retained for screening. Title and abstract screening, based on predefined inclusion and exclusion criteria, resulted in 17 potentially relevant studies. Full-text assessment confirmed these 17 studies. A backward snowball search was subsequently conducted in July 2025, leading to the identification of 5 additional relevant studies. The final sample consisted of 21 peer-reviewed studies. All selection decisions were documented in a structured screening spreadsheet, including bibliographic data, inclusion/exclusion decisions, and justification for exclusion.

Inclusion and Exclusion Criteria

Inclusion criteria

- IC1: Peer-reviewed journal articles, conference papers, or book chapters
- IC2: Written in English or Italian
- IC3: Published between 2018 and June 2025
- IC4: Explicitly addressing prompt engineering or responsible LLM use within secondary education contexts

Exclusion criteria

- EC1: Studies not available in full text
- EC2: Studies focused exclusively on higher education or professional training
- EC3: Conceptual or technical papers unrelated to educational implementation

Data Extraction and Coding

Data extraction was conducted using a structured coding template designed to collect the following:

- Bibliographic information
- Educational context
- Study design and methodology
- Type of prompt engineering intervention
- Reported cognitive, metacognitive, or ethical outcomes
- Alignment with RQ1–RQ4

Coding proceeded in two stages. In the first cycle, studies were deductively mapped to the four predefined research questions (RQ1–RQ4). Each study's contribution was classified as full, partial, or marginal according to predefined criteria. To enhance transparency and replicability of the coding process, we defined explicit analytical criteria to distinguish between full, partial, and marginal contributions of each study to the research questions (as shown in Table 4).

Table 4. Criteria used to classify the contribution of each study to the research questions

Contribution level	Definition	Typical indicators in the reviewed studies
Full contribution	The study explicitly investigates the research question and reports empirical findings directly addressing it.	The intervention or analysis focuses on the RQ, and the results provide direct evidence related to it.
Partial contribution	The research question is addressed indirectly or as a secondary aspect of the study.	The study includes relevant elements related to the RQ, but they are not the main focus of the research design.
Marginal contribution	The research question is only mentioned or implicitly related to the study, without direct analysis or empirical evaluation.	Prompt engineering or LLM use appears only as a contextual or technical element rather than a pedagogical focus.

In the second cycle, an inductive thematic synthesis was performed to identify recurring pedagogical patterns, instructional approaches, and reported challenges. Initial coding was conducted by the first author. To enhance conceptual coherence and reduce interpretive bias, coding decisions and thematic categorizations were subsequently reviewed and discussed with the co-author until agreement was reached.

Quality Appraisal

To contextualize the strength of evidence across the included studies, we conducted a qualitative quality appraisal (available at this link) based on four analytical criteria commonly used in systematic reviews, as shown in Table 5. Studies were not excluded based on quality appraisal; however, appraisal informed the interpretive weighting of findings in the discussion section.

Table 5. Quality appraisal criteria used to contextualize the strength of evidence

Criterion	Description	Purpose of the review
Clarity of methodological design	The study clearly describes its research design and methodological approach.	To assess whether the study provides sufficient information to interpret its findings.
Presence of empirical data	The study reports empirical observations, experiments, or classroom implementations rather than purely conceptual discussions.	To distinguish empirical evidence from conceptual or exploratory contributions.
Transparency of data collection	The procedures used to collect and analyze data are clearly described.	To evaluate the reliability and interpretability of the reported results.
Discussion of limitations	The study explicitly acknowledges methodological or contextual limitations.	To assess the degree of critical reflection and transparency in the research.

Overall, most studies met at least three of the four criteria. Empirical designs generally demonstrated clear methodological reporting, whereas exploratory and pilot studies showed greater variability in transparency and limitations reporting. These patterns were considered when interpreting the strength of evidence across RQs.

CONDUCTING THE SYSTEMATIC REVIEW

In this section, we describe the search protocol we developed to identify and select studies.

STUDY SELECTION

To select the relevant studies, we adopted a two-phase search and screening process, in line with the guidelines for conducting systematic reviews proposed by Kitchenham and Charters (2007) and following the PRISMA framework (Page et al., 2021).

Phase 1 – Searching the digital library

In the first stage, we applied the search string to four databases: Scopus, Web of Science, IEEE Xplore, and ACM Digital Library. Initially, this query returned a total of 703 results, distributed as follows:

(1) **439** records from ACM Digital Library
(2) **213** records from IEEE Xplore Digital Library
(3) **30** records from Scopus (Elsevier)
(4) **21** records from Web of Science (Clarivate)

We then removed duplicates using the deduplication function of Zotero, an open-source bibliographic reference manager, and 659 unique records were confirmed. Conference proceedings published as standalone books or unindexed collections (e.g., workshop abstract books) were excluded unless individual contributions were peer-reviewed and indexed in Scopus, Web of Science, or ERIC. This ensured the inclusion of only high-quality, traceable sources.

After the above-described steps, 411 articles remained available. We then applied the predefined inclusion and exclusion criteria to titles and abstracts, obtaining a subset of 17 potentially relevant studies. These were further assessed based on full-text analysis to look for any adherence to the initial research questions.

Phase 2 – Backward snowball search

The second phase consisted of a backward snowball search: we analyzed the bibliographies of the studies selected in Phase 1 to identify additional relevant contributions not found in the initial query. This process led to the inclusion of 5 additional studies, which were subject to the same selection process described earlier. In total, the process led to the selection of 21 studies for data extraction and analysis. All the selection steps we followed were documented in a structured spreadsheet, which included: bibliographic data, source, inclusion/exclusion criteria applied, reason for exclusion and assigned research questions.

Figure 1 shows the study selection process, following the PRISMA 2020 flow diagram guidelines proposed by Page et al. (2021). Eventually, we made the complete dataset available on the web, at the address SLR Excel Dataset, which contains the full list of screened studies, inclusion/exclusion decisions, and coding information used in this review. Table 6 shows all the selected studies.

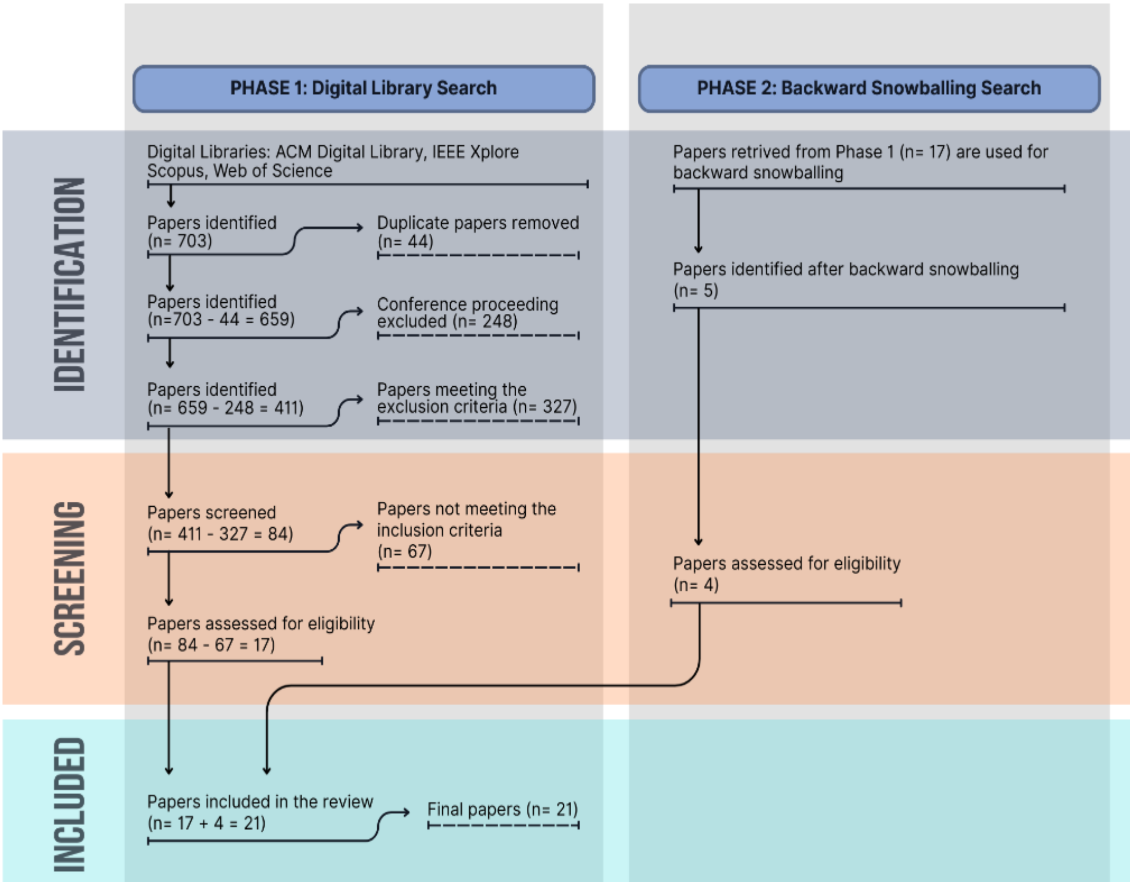


Figure 1. PRISMA 2020 flow diagram of the study selection process

Table 6. List of selected studies included in the review

ID	Reference	ID	Reference
S1	Huffman et al. (2025)	S11	Frazier et al. (2024)
S2	Peng et al. (2024)	S12	Abolnejadian et al. (2024)
S3	Aftabi et al. (2024)	S13	Li et al. (2025)
S4	Zhao et al. (2025)	S14	Doherty et al. (2025)
S5	Verhelst et al. (2024)	S15	Lieb and Goel (2024)
S6	Snyder et al. (2024)	S16	Woo et al. (2024)
S7	Fung et al. (2024)	S17	Shao et al. (2025)
S8	Moreau-Pernet et al. (2024)	S18	Bitzenbauer (2023)
S9	Shin et al. (2025)	S19	Alneyadi and Wardat (2023)
S10	Ali et al. (2024)	S20	Yang et al. (2025)
S21	Javier and Moorhouse (2023)		

DATA EXTRACTION

We conducted the data extraction and synthesis phase using a structured document, designed by us to systematically collect relevant information from each of the selected studies. The data extraction document is structured in a set of general fields, including title, authors, publication year, source, and publication type. It also includes a section summarizing the study's contribution, highlighting methodological strengths and weaknesses, and tracking the inclusion/exclusion criteria applied, as well as the study's correspondence to the RQs. The data extraction document is available via the web at the address Data Extraction Document. To support the data extraction workflow, ChatGPT was used only as an auxiliary tool for language support and to generate preliminary suggestions for possible alignment between studies and the research questions (RQ1–RQ4). Importantly, the model did not perform study selection, coding, or classification decisions. All inclusion decisions and RQ mappings were determined through manual full-text reading conducted by the authors. Any preliminary suggestions generated by the AI system were critically reviewed and corrected when necessary. This procedure was adopted solely to support the organization of the analysis while ensuring that all interpretive decisions remained under direct human control. The annotated versions of the selected contributions, with the highlighted textual citations to support the classification, are available at the following link: Annotated Papers.

RESULTS OF THE SYSTEMATIC LITERATURE REVIEW

The results are organized according to the four predefined research questions (RQ1–RQ4). Themes within each RQ were identified through an inductive grouping process conducted during second-cycle coding. Recurrent patterns across the included studies, such as forms of instructional integration, reported cognitive or metacognitive effects, and identified ethical challenges, were clustered into analytically coherent categories. The classification of each study's contribution (full, partial, or marginal) was based on explicit alignment between the study's objectives, methodological design, and reported findings relative to each RQ. Table 7 summarizes the degree to which each study (S1–S21) contributes to the four RQs. To enhance transparency and facilitate cross-study comparison, Table 8 provides a concise overview of the included studies, summarizing the type of intervention, main reported outcomes, and study design. This overview complements the RQ-based mapping and supports the thematic interpretation presented in the following subsections. For each study, the level of contribution was classified as full (●●●), partial (●●), or marginal (●). Figures 2–5 present the proportional distribution of studies addressing each RQ. Overall, the distribution reveals a stronger emphasis on the technical and contextual integration of LLMs in secondary education (RQ1 and RQ3), while cognitive effects (RQ2) and responsible use considerations (RQ4) remain comparatively less systematically developed.

CROSS-STUDY SYNTHESIS

Across the reviewed studies, several cross-cutting patterns emerge. First, prompt engineering is most often treated as an instrumental interaction technique embedded within disciplinary learning tasks rather than as an explicit learning objective. In many interventions, prompting is used to facilitate content generation, feedback, or tutoring functions, while the process of designing and refining prompts is not itself framed as a competence to be developed. Second, the methodological approaches used to investigate the educational impact of prompt engineering vary considerably. While some quasi-experimental studies report measurable improvements in learning outcomes, particularly in programming or STEM contexts, many classroom-based implementations rely on qualitative observations, self-reported perceptions, or exploratory designs. This methodological heterogeneity limits the possibility of drawing strong causal conclusions regarding the cognitive and metacognitive effects of prompting practices. Third, disciplinary differences can be observed in how prompt engineering is integrated into educational activities. Studies in computer science and programming tend to focus on performance outcomes and support for problem-solving, whereas language-learning studies more often emphasize reflective engagement and students' perceptions of AI-supported learning. Taken together, these patterns suggest that research on prompt engineering in secondary education is currently characterized by strong pedagogical experimentation but limited convergence toward shared instructional models or standardized evaluation approaches.

RQ1: IN WHAT WAYS IS PROMPT ENGINEERING ADDRESSED IN FORMAL OR INFORMAL EDUCATIONAL ACTIVITIES TARGETING HIGH SCHOOL STUDENTS?

Across the 21 selected studies, prompt engineering appears with varying degrees of centrality. In 14 studies, it plays an explicit or structurally significant role within the educational intervention. In these cases, prompt engineering is either treated as a core instructional component or used as a deliberate strategy to personalize learning processes (e.g., S2, S3, S12, S16). In four studies (e.g., S5, S8), prompt engineering is present but remains peripheral, typically embedded within system design or instructional tools without being framed as an explicit learning objective. In three studies (e.g., S1, S19), prompting is either absent from the pedagogical focus or confined to the technical implementation of the system, with no direct student engagement in prompt construction. Only a limited subset of interventions explicitly trained students to formulate, refine, and reflect upon their prompting strategies (notably S12, S16, S21). In contrast, several studies relied on predefined prompts or structured system-generated inputs, as their primary focus concerned content generation, feedback automation, or usability rather than the development of prompting competence itself. Informal learning contexts, such as workshops and creative laboratory activities (e.g., S10), tended to allow greater exploratory engagement with prompt design. Formal classroom settings, by contrast, more frequently employ prompt engineering instrumentally to support other disciplinary learning objectives.

Taken together, these findings indicate that while prompt engineering is increasingly operationalized in secondary education contexts, it is rarely conceptualized as an autonomous pedagogical competence. Its instructional framing remains uneven and often subordinated to broader technological or disciplinary goals.

Table 7. Mapping of the selected studies' contributions across the four research questions

ID	Author	RQ1	RQ2	RQ3	RQ4
S1	Huffman et al. (2025)	●	●	●	●
S2	Peng et al. (2024)	●	●●	●	●
S3	Aftabi et al. (2024)	●	●●	●●	●
S4	Zhao et al. (2025)	●	●	●	●
S5	Verhelst et al. (2024)	●●	●	●	●●
S6	Snyder et al. (2024)	●	●●	●	●●
S7	Fung et al. (2024)	●	●●	●	●●
S8	Moreau-Pernet et al. (2024)	●	●	●	●
S9	Shin et al. (2025)	●	●●	●	●●
S10	Ali et al. (2024)	●	●●	●	●
S11	Frazier et al. (2024)	●	●●	●	●●
S12	Abolnejadian et al. (2024)	●	●	●	●●
S13	Li et al. (2025)	●	●	●	●●
S14	Doherty et al. (2025)	●●	●	●	●●
S15	Lieb & Goel (2024)	●●	●	●	●●
S16	Woo et al. (2024)	●	●	●	●
S17	Shao et al. (2025)	●●	●	●	●
S18	Bitzenbauer (2023)	●	●●	●	●
S19	Alneyadi & Wardat (2023)	●	●	●	●●
S20	Yang et al. (2025)	●	●	●	●●
S21	Javier & Moorhouse (2023)	●	●	●	●

Table 8. Comparative overview of the included studies (S1–S21), summarizing intervention type, main reported outcomes, and study design

Study	Educational context	Type of intervention	Main reported outcomes	Study design
S1 (Huffman et al., 2025)	Deaf and Hard-of-Hearing secondary learners	LLM usage study focused on accessibility and interaction patterns	Identified communication benefits and accessibility challenges for DHH learners	Mixed methods
S2 (Peng et al., 2024)	Secondary CS1 education	RAG-based prompt framework (TILSE) for personalized AI feedback	Improved accuracy and personalization of feedback generation	Quantitative evaluation
S3 (Aftabi et al., 2024)	Secondary-level AI-supported personalized learning	Structured prompt-based framework for customized course design	Enhanced personalization and support for independent learning	Descriptive design
S4 (Zhao et al., 2025)	10th-grade programming course	AI-based human–computer collaborative programming method	Improved computational thinking, learning attitudes, and achievement	Quasi-experimental study
S5 (Verhelst et al., 2024)	High school second language learning (Spanish)	Generative AI tutoring system integrated with a social robot	Significant vocabulary gains; robot presence did not significantly affect outcomes	Quantitative study
S6 (Snyder et al., 2024)	Secondary STEM computational modeling	LLM-supported analysis of collaborative problem-solving behaviors	Identified behavioral differences between high- and low-performing students; implications for adaptive scaffolding	Mixed methods
S7 (Fung et al., 2024)	K-12 data science education	RAG-based automatic feedback generation framework (DSRAG)	Provided scalable, personalized feedback aligned with K-12 pedagogical needs	Qualitative case study
S8 (Moreau-Pernet et al., 2024)	Mathematics tutoring contexts	LLM-based classification of tutor discursive moves	High classification performance for instructional talk moves	Quantitative computational model evaluation
S9 (Shin et al., 2025)	Secondary school formative peer assessment	AI-generated vs peer-generated feedback comparison	Students preferred AI feedback in low-participation contexts; peer feedback preferred by confident students	Mixed methods

Study	Educational context	Type of intervention	Main reported outcomes	Study design
S10 (Ali et al., 2024)	High school workshop on generative AI	Creative workshop using prompt engineering for text-to-image generation	Students developed technical understanding and identified societal benefits and harms	Qualitative exploratory study with informal assessment
S11 (Frazier et al., 2024)	Secondary Computer Science Principles (CSP) course	Customized ChatGPT conversational support for CSP learning	Students preferred customized ChatGPT; improved perceived clarity and relevance of explanations	Mixed methods
S12 (Abolnejadian et al., 2024)	Secondary introductory programming (CS1)	ChatGPT-based personalized prompt-supported learning platform	Increased engagement and hands-on programming experience; shift toward mentor-style teaching	Mixed-methods quasi-experimental design
S13 (Li et al., 2025)	High school text analysis education	Structured GAI prompt framework for analytical skill development	Significant improvements in readability, accuracy, completeness, logicality, and critical thinking	Quasi-experimental mixed-methods study
S14 (Doherty et al., 2025)	Middle and high school collaborative learning	LLM-based agent supporting small-group collaboration	Reduced social loafing and increased respectful collaboration behaviors	Mixed-methods experimental design
S15 (Lieb & Goel, 2024)	Secondary physics education	LLM-based tutoring chatbot (NewtBot) with configurable prompting	Positive user experience; setting-specific tutor configuration rated highest	Mixed methods
S16 (Woo et al., 2024)	Secondary EFL writing instruction	Classroom workshop teaching prompt engineering with ChatGPT	High student satisfaction; increased motivation; high cognitive load during prompt use	Mixed methods
S17 (Shao et al., 2025)	High school biology and physics	LLM-generated analogies for scientific concept learning	Enhanced understanding, especially in biology; required teacher guidance to prevent over-reliance	Mixed methods
S18 (Bitzen-bauer, 2023)	Secondary physics education	Implementation of ChatGPT-based classroom activities	Pilot results indicate feasibility and positive classroom integration	Pilot study

Study	Educational context	Type of intervention	Main reported outcomes	Study design
S19 (Alneyadi & Wardat, 2023)	Eleventh-grade physics (electromagnetism unit)	ChatGPT-supported instructional intervention	Improved student achievement in the electromagnetism unit	Mixed-method quasi-experimental design
S20 (Yang et al., 2025)	High school programming education	ChatGPT-assisted programming learning intervention	Improved programming learning outcomes compared to control group	Quasi-experimental mixed-methods study
S21 (Javier & Moorhouse, 2023)	Secondary English language learning	Instructional activities for productive and critical ChatGPT use	Developed students' productive and critical use of ChatGPT	Mixed methods

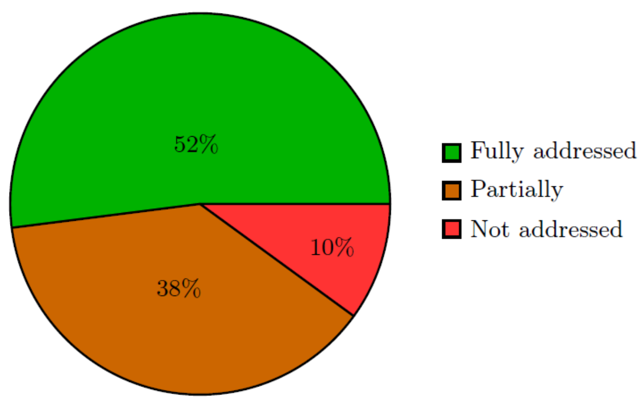


Figure 2. Coverage of Research Question 1 (RQ1) across the 21 selected studies

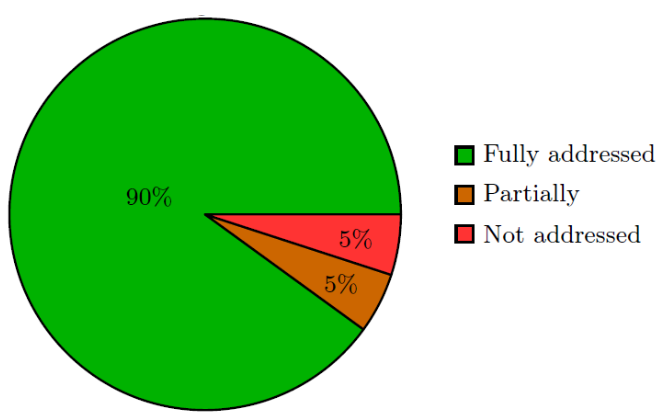


Figure 3. Coverage of Research Question 2 (RQ2) across the 21 selected studies

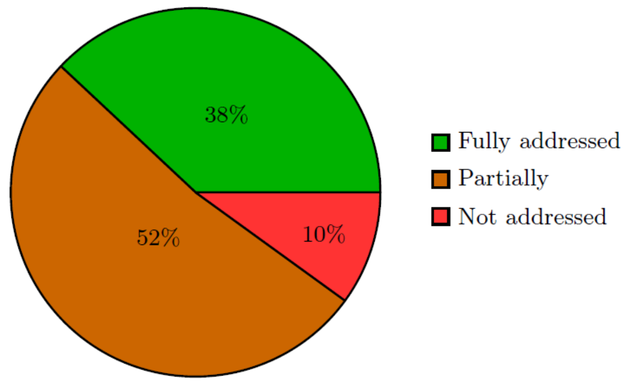


Figure 4. Coverage of Research Question 3 (RQ3) across the 21 selected studies

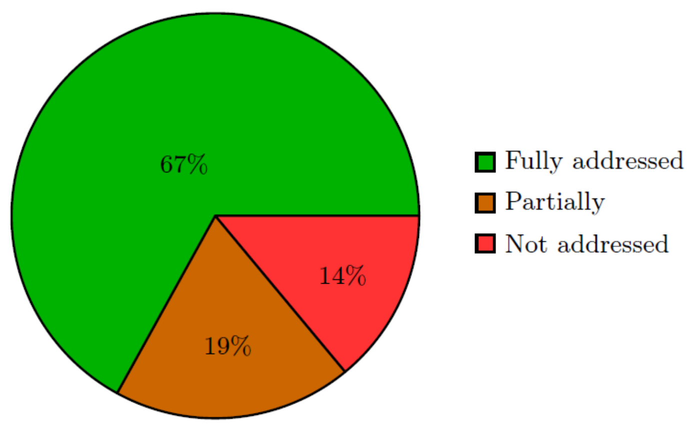


Figure 5. Coverage of Research Question 4 (RQ4) across the 21 selected studies

RQ2: WHAT EVIDENCE IS REPORTED ON THE COGNITIVE OR METACOGNITIVE EFFECTS OF PROMPT ENGINEERING ON STUDENT LEARNING IN SECONDARY SCHOOLS?

Several studies address the cognitive and metacognitive implications of prompt engineering, though with varying methodological rigor. As shown in Figure 3, 52% of the included studies provide full evidence addressing this dimension, 38% offer partial insights, and 10% do not address it explicitly. Some studies report improvements in motivational, affective, or reflective dimensions rather than directly measured academic gains. For example, S16 investigated the use of ChatGPT in an English as a Second Language writing workshop. Although standardized performance measures were not employed, the study documented increased motivation and satisfaction, alongside high cognitive load during prompt construction, based on questionnaire data and think-aloud protocols. These findings suggest that prompt engineering may simultaneously support engagement and introduce additional cognitive demands. Similarly, S21 reported that students developed a more critical and productive understanding of ChatGPT through guided instructional activities. However, evidence relied primarily on post-task reflections rather than controlled measures of learning outcomes, indicating that metacognitive gains were inferred rather than experimentally verified. In S15, data on user experience and AI apprehension were collected, yet learning outcomes, although gathered, were not analyzed, limiting the study's contribution to the empirical evidence base of RQ2. In contrast, quasi-experimental studies such as S4, S13, S19, and S20 provide stronger empirical evidence, reporting measurable improvements in computational thinking, analytical skills, or academic achievement following structured LLM-supported interventions. Overall, while a growing body of research suggests positive cognitive and metacognitive effects associated with prompt engineering, the methodological heterogeneity of the studies and the frequent reliance on indirect measures indicate that robust, longitudinal, and controlled assessments remain limited. This underscores the need for more rigorous evaluation frameworks capable of systematically capturing learning gains associated with structured prompting practices in secondary education.

RQ3: What Pedagogical Approaches, Disciplinary Contexts, or Tools Are Used to Introduce, and/or Sustain, the Use of LLM as a Support to Learning in Secondary Education?

The reviewed studies demonstrate a wide range of pedagogical approaches and technological configurations for integrating large language models (LLMs) into secondary education. As indicated by the high proportion of full coverage (90%), this dimension represents the most extensively developed area in the current literature. In several studies, LLMs are embedded within intelligent tutoring systems or automated feedback frameworks. For example, S7 introduces the DSRAG framework for personalized feedback generation, while S6 and S8 integrate LLM-based analysis tools within authentic classroom contexts to support collaborative problem-solving and discourse evaluation. Other interventions focus on programming education (e.g., S20, S11), physics instruction (e.g., S15), and science concept understanding (e.g., S17), often positioning LLMs as dialogic or scaffolded tutoring agents. A variety of disciplinary contexts are represented, including mathematics (S8), data science (S7), language learning (S16, S21), programming (S12, S20), and broader STEM collaborative environments (S14). Studies also differ in the degree of student agency involved: in some cases, LLMs are used primarily as analytical tools by researchers (e.g., S6), whereas in others, students engage directly in dialogic interaction with AI systems (e.g., S9, S15). Technologically, implementations range from relatively simple prompt-based interfaces to structured frameworks incorporating retrieval-augmented generation (S2, S7) or configurable back-end prompting architectures (S15). Several studies explicitly explore structured prompt frameworks as part of curricular design (e.g., S2, S12, S13), indicating emerging efforts to formalize prompting within instructional sequences. Overall, the findings reveal strong experimentation and technical diversification in how LLMs are introduced into secondary classrooms. However, while pedagogical use cases are numerous and contextually varied, convergence toward standardized instructional models remains limited.

RQ4: What Challenges, Limitations, or Educational Needs Arise in Training Students – and Potentially Teachers – to Use LLMs Responsibly in High School Settings?

Several studies explicitly address challenges related to the responsible and pedagogically sound integration of LLMs in secondary education. As shown in Figure 5, only 38% of the reviewed studies fully engage with this dimension, while the majority address it only partially. Recurring concerns include limited student awareness of model limitations, risks of over-reliance, accessibility constraints, and the need for structured guidance for both students and teachers. For instance, S1 highlights how members of the Deaf and Hard-of-Hearing community engage with LLMs under specific accessibility conditions, underscoring the importance of inclusive prompt design and tailored user support. Similarly, S18 acknowledges that although ChatGPT can serve as a powerful instructional aid, students require explicit instruction to develop critical and responsible interaction practices. Technical limitations are also emphasized. S6 identifies well-known issues such as hallucinations and token constraints, noting that awareness of these limitations should form part of LLM literacy training. S17, while demonstrating the instructional value of LLM-generated analogies, cautions against passive consumption of AI-generated content without adequate teacher mediation. Likewise, S16 reports high levels of student satisfaction but also identifies cognitive overload and the necessity of explicit scaffolding during prompt construction. Structured teacher facilitation is further reinforced in S4, where collaborative AI-supported learning produced positive outcomes under guided instructional conditions. Collectively, these findings suggest that while secondary students are capable of interacting productively with generative AI systems, responsible use does not emerge spontaneously. Instead, it requires intentional pedagogical framing, explicit scaffolding, and teacher preparation. The relatively limited number of studies that fully address this dimension reveals a structural gap between technological adoption and the development of coherent ethical and instructional frameworks for prompt-based interaction.

DISCUSSION

SUMMARY OF RESULTS

A key distinction emerging from this systematic literature review concerns the different educational roles attributed to prompt engineering across the analyzed studies. In particular, RQ1 and RQ2 capture two analytically distinct but often conflated dimensions: while RQ1 addresses whether and how prompt engineering is framed as an explicit educational object in secondary education, RQ2 focuses on the cognitive and metacognitive effects associated with students' use of prompts when interacting with large language models. Overall, the review reveals that prompt engineering is gaining visibility in secondary education research, but its integration remains fragmented and weakly formalized. With respect to RQ1, only a limited number of studies explicitly conceptualize prompt engineering as a learning objective or instructional content (e.g., S12, S16, S21). More frequently, prompt-related practices are embedded implicitly within task-oriented activities, without clear pedagogical framing, curricular positioning, or assessment criteria. As a result, prompt engineering is rarely presented to students as a reflective competence connected to AI literacy or metacognitive regulation.

In contrast, findings related to RQ2 suggest that several studies acknowledge cognitive and metacognitive implications of prompt use (e.g., S4, S13, S20 for measurable learning outcomes; S16, S21 for reflective and affective indicators). These include increased reflection, iterative reasoning, awareness of task formulation, and perceived agency in human–AI interaction. However, such effects are typically inferred from qualitative observations or indirect indicators rather than systematically measured through validated instruments or longitudinal designs. This imbalance indicates that prompt engineering is more often valued for its functional impact on learning processes than intentionally designed as an object of teaching and reflection.

Regarding RQ3, the disciplinary distribution of studies shows that prompt engineering is most frequently explored within specific subject areas, such as computer science, language learning, and STEM education. While this disciplinary grounding provides valuable contextual insights, it reinforces the tendency to treat prompt engineering as a domain-specific technique rather than as a cross-cutting educational competence applicable across curricula. The overlap observed between RQ1 and RQ3 reflects a limitation of the existing literature – where prompt engineering is embedded within disciplinary applications – rather than a weakness of the analytical framework adopted in this review.

Finally, RQ4 highlights that ethical considerations and responsible AI use are commonly mentioned but rarely operationalized. Although concerns related to bias, transparency, over-reliance, and accountability are acknowledged (e.g., S17, S18), few studies propose concrete pedagogical strategies, classroom practices, or institutional policies to address these issues in structured ways. Ethical reflection thus appears more as a peripheral discourse than as an integrated component of prompt engineering education. Taken together, these findings suggest that prompt engineering in secondary education is currently positioned at the intersection of emerging practice and underdeveloped pedagogy. In summary, RQ1 and RQ4 reveal limited and uneven pedagogical formalization, RQ2 shows promising but methodologically heterogeneous evidence of cognitive impact, and RQ3 demonstrates strong disciplinary experimentation without convergence toward shared instructional models. The following sections discuss the broader theoretical and pedagogical implications emerging from these findings.

Alignment with Institutional and Policy Trends

From an educational technology perspective, the observed fragmentation contrasts with the growing international emphasis on AI literacy and responsible AI use in schools. Several institutional initiatives underscore the need for educational approaches that explicitly address students' interactions with AI systems. The Rome Call for AI Ethics, promoted by the Pontifical Academy for Life, emphasizes principles such as transparency, inclusion, accountability, impartiality, and safety (Pontifical Academy for Life, 2020). Similarly, UNESCO's guidance on generative AI in education advocates a human-centered, equity-oriented approach (UNESCO, 2023). More recently, the OECD and the European Commission have proposed an AI literacy framework to equip learners with the knowledge, skills, and attitudes to navigate AI-mediated environments responsibly (OECD and European Commission, 2025). Despite this policy momentum, our analysis reveals a persistent disconnect between institutional aspirations and educational practice. Ethical principles are seldom translated into concrete instructional designs or classroom-level interventions. In particular, the absence of explicit pedagogical framing for prompt engineering (RQ1) limits the ability to systematically address ethical awareness and responsible use (RQ4). This misalignment suggests that without recognizing prompt engineering as an educational object in its own right, ethical considerations are likely to remain marginal and inconsistently addressed.

Implications for Educators and Stakeholders

The implications of this review are primarily conceptual and pedagogical rather than prescriptive. From a theoretical perspective, the findings support interpreting prompt engineering as a metacognitive and socio-technical competence situated within emerging AI literacy frameworks. Prompt construction requires learners to plan, monitor, and revise their interaction strategies when engaging with generative systems, aligning with models of self-regulated learning and epistemic agency. Framing prompt engineering in this way strengthens its conceptual positioning beyond a purely technical interaction skill. First, the findings highlight the need to move beyond treating prompt engineering as a purely technical or instrumental skill. Instead, prompt engineering should be framed as a mediating practice through which students negotiate agency, feedback, and knowledge construction in AI-supported learning environments. For educators, this shift implies the necessity of developing instructional strategies that make prompt design explicit, discussable, and assessable. Teachers require professional development opportunities that support their understanding of how prompts shape AI outputs, influence epistemic authority, and mediate learning processes. More broadly, collaboration between educators, educational technologists, and policymakers is essential to designing coherent curricular frameworks that integrate prompt engineering within AI literacy initiatives. Furthermore, recognizing the socio-technical nature of large language models is crucial. Prompts are not neutral inputs: they encode assumptions, frame problems, and can amplify biases. Supporting students in critically reflecting on their prompts can therefore foster more responsible, transparent, and reflective engagements with generative AI systems.

Limitations of the Review

The limitations of this review largely reflect the current state of research on prompt engineering in secondary education. Although a comprehensive search strategy was employed across multiple academic databases, some relevant studies, particularly recent or unpublished work, may not have been captured. Additionally, the qualitative coding of studies in relation to the research questions involves interpretive judgment and may introduce subjectivity. Moreover, the limited availability of standardized measures to assess prompt engineering skills, metacognitive outcomes, or ethical awareness constrained systematic comparisons across studies. These limitations should be interpreted not as weaknesses of individual contributions, but as indicators of an emerging research area that has not yet reached methodological consolidation.

DIRECTIONS FOR FUTURE RESEARCH

Based on the gaps identified in this review, several directions for future research emerge. First, there is a clear need for empirical and longitudinal studies that investigate how students learn to formulate, evaluate, and revise prompts across different disciplines and educational contexts. Such studies should explicitly operationalize prompt engineering as a learning objective rather than treating it as an implicit practice.

Second, future work should focus on developing validated assessment tools that capture both the quality of prompts and associated cognitive, metacognitive, and ethical dimensions.

Finally, stronger alignment between educational research and policy initiatives is required to produce scalable curricular frameworks that integrate prompt engineering within broader AI literacy and responsible AI education agendas.

CONCLUSIONS

SUMMARY OF FINDINGS

This systematic review examined how prompt engineering is conceptualized, implemented, and evaluated in secondary education. The findings indicate that while prompt engineering is increasingly present in classroom practice, it remains inconsistently framed as an explicit pedagogical objective. Evidence of cognitive and metacognitive benefits is emerging but methodologically heterogeneous, and ethical considerations, though frequently acknowledged, are rarely operationalized through structured instructional models. Overall, the literature reflects strong experimentation with large language models across disciplinary contexts, yet limited convergence toward shared curricular frameworks or assessment strategies.

CONTRIBUTION OF THE STUDY

This review contributes to the field by positioning prompt engineering not merely as a technical interaction skill, but as a metacognitive and socio-technical competence situated within emerging AI literacy discourse. By systematically mapping how prompt engineering is addressed in secondary education, the study clarifies structural gaps between technological adoption, pedagogical formalization, and ethical governance. More broadly, this review reframes prompt engineering as a pedagogical construct that can be explicitly taught, analyzed, and assessed within secondary education. By synthesizing existing studies through this conceptual lens, the review highlights the need to move beyond treating prompting simply as a technical interaction strategy and instead consider it as part of students' metacognitive engagement with AI systems. This perspective suggests new directions for future research, including the development of instructional models, assessment frameworks, and teacher training initiatives specifically addressing prompt-based interaction with generative AI. In this way, prompt engineering can be considered an educational object in its own right, rather than merely a peripheral tool embedded within disciplinary applications. Future research should further consolidate empirical evidence and develop validated assessment frameworks capable of systematically capturing the cognitive, metacognitive, and ethical dimensions of prompt engineering in secondary education.

ACKNOWLEDGEMENTS

The authors acknowledge the use of OpenAI's ChatGPT for support in drafting, editing, and organizing some sections of the manuscript. The AI tool was used exclusively under human supervision, with all final content critically reviewed and validated by the authors, who assume full responsibility

for its accuracy and integrity. The authors also used Zotero for reference management throughout the preparation of this manuscript.

Declaration of Interest

The authors report there are no competing interests to declare.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Abolnejadian, M., Alipour, S., & Taeb, K. (2024). Leveraging ChatGPT for adaptive learning through personalized prompt-based instruction: A CS1 education case study. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24). ACM. https://doi.org/10.1145/3613905.3637148

Aftabi, E., Nategholeslam Shirazi, B., Safavi, A. A., Salimi, G., & Aftabi, H. (2024). A framework for customized course design and personalized learning with AI. *Proceedings of the 11th International and the 17th National Conference on E-Learning and E-Teaching, Isfahan, Islamic Republic of Iran,* 1–6. https://doi.org/10.1109/ICeLET62507.2024.10493063

Ali, S., Ravi, P., Williams, R., DiPaola, D., & Breazeal, C. (2024). Constructing dreams using generative AI. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(21), 23268–23275. https://doi.org/10.1609/aaai.v38i21.30374

Alneyadi, S., & Wardat, Y. (2023). ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools. *Contemporary Educational Technology*, *15*(4), ep448. https://doi.org/10.30935/cedtech/13417

Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to-implement activities. *Contemporary Educational Technology*, *15*(3), ep430. https://doi.org/10.30935/cedtech/13176

Chen, E., Wang, D., Xu, L., Cao, C., Fang, X., & Lin, J. (2024). *A systematic review on prompt engineering in large language models for K–12 STEM education.* PsyArXiv. https://arxiv.org/abs/2410.11123

Dimeli, M., & Kostas, A. (2025). The role of ChatGPT in education: Applications, challenges: Insights from a systematic review. *Journal of Information Technology Education: Research*, *24*, Article 2. https://doi.org/10.28945/5422

Doherty, E., Perkof, E. M., von Bayern, S., Zhang, R., Dey, I., Bodzianowski, M., Puntambekar, S., & Hirshfeld, L. (2025). Piecing together teamwork: A responsible approach to an LLM-based educational jigsaw agent. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Article 19). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713349

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Frazier, M., Damevski, K., & Pollock, L. (2024). Customizing ChatGPT to help computer science principles students learn through conversation. *Proceedings of the ACM Conference on Innovation and Technology in Computer Science Education* (pp. 633-639). Association for Computing Machinery. https://doi.org/10.1145/3649217.3653570

Fung, S. C. E., Wong, M. F., & Tan, C. W. (2024). Automatic feedback generation on K–12 students' data science education by prompting cloud-based large language models. *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (pp. 255-258). Association for Computing Machinery. https://doi.org/10.1145/3657604.3664673

Gusenbauer, M. (2024). Beyond Google Scholar, Scopus, and Web of Science: An evaluation of the backward and forward citation coverage of 59 databases' citation indices. *Research Synthesis Methods*, *15*(5), 802–817. https://doi.org/10.1002/jrsm.1729

Huffman, S., Chen, S., Mack, K. A., Su, H., Wang, Q., & Kushalnagar, R. (2025). "We do use it, but not how hearing people think": How the Deaf and Hard of Hearing community uses large language model tools. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Article 33). Association for Computing Machinery. https://doi.org/10.1145/3706599.3719785

Javier, D. R. C., & Moorhouse, B. L. (2023). Developing secondary school English language learners' productive and critical use of ChatGPT. *TESOL Journal*, *15*(2), e755. https://doi.org/10.1002/tesj.755

Kavitha, K., & Joshith, V. P. (2024). Pedagogical incorporation of artificial intelligence in K–12 science education: A decadal bibliometric mapping and systematic literature review (2013–2023). *Journal of Pedagogical Research*, *8*(4), 437–465. https://doi.org/10.33902/JPR.202429218

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. *Technical Report EBSE 2007-001*. Keele University. https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education*, *22*, Article 7. https://doi.org/10.1186/s41239-025-00503-7

Li, X., Li, T., Wang, M., Tao, S., Zhou, X., Wei, X., & Guan, N. (2025). Navigating the textual maze: Enhancing textual analytical skills through an innovative GAI prompt framework. *IEEE Transactions on Learning Technologies*, *18*, 206–215. https://doi.org/10.1109/TLT.2025.3539104

Lieb, A., & Goel, T. (2024). Student interaction with NewtBot: An LLM-as-tutor chatbot for secondary physics education. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Article 614). Association for Computing Machinery. https://doi.org/10.1145/3613905.3647957

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177. https://doi.org/10.1016/j.joi.2018.09.002

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, *58*(13), 2105–2125. https://doi.org/10.1002/asi.20677

Moreau-Pernet, B., Tian, Y., Sawaya, S., Foltz, P., Cao, J., Milne, B., & Christie, T. (2024). Classifying tutor discursive moves at scale in mathematics classrooms with large language models. *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (pp. 361-365). Association for Computing Machinery. https://doi.org/10.1145/3657604.3664664

OECD and European Commission. (2025). *Empowering learners for the age of AI: Launch of the draft AI literacy framework and stakeholder consultations.* https://education.ec.europa.eu/event/empowering-learners-for-the-age-of-ai-draft-ai-literacy-framework-launch

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, M., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Peng, B., Wang, X., & Xu, L. (2024). "TILSE" framework for RAG-based AIGC feedback prompts: A modular and personalized intelligent feedback generation method. *Proceedings of the International Symposium on Artificial Intelligence for Education* (pp. 398-403). Association for Computing Machinery. https://doi.org/10.1145/3700297.3700365

Pontifical Academy for Life. (2020). *Rome call for AI ethics.* https://www.romecall.org/wp-content/uploads/2022/03/RomeCall_Paper_web.pdf

Shao, Z., Yuan, S., Gao, L., He, Y., Yang, D., & Chen, S. (2025). Unlocking scientific concepts: How effective are LLM-generated analogies for student understanding and classroom practice? *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Article 143). Association for Computing Machinery. https://doi.org/10.1145/3706598.3714313

Shin, I., Hwang, S. B., Yoo, Y. J., Bae, S., & Kim, R. Y. (2025). Comparing student preferences for AI-generated and peer-generated feedback in AI-driven formative peer assessment. *Proceedings of the 15th International Learning Analytics and Knowledge Conference* (pp. 159-169). Association for Computing Machinery. https://doi.org/10.1145/3706468.3706488

Snyder, C., Hutchins, N. M., Cohn, C., Fonteles, J. H., & Biswas, G. (2024). Analyzing students' collaborative problem-solving behaviors in synergistic STEM+C learning. *Proceedings of the 14th International Conference on Learning Analytics and Knowledge Conference* (pp. 540-550). Association for Computing Machinery. https://doi.org/10.1145/3636555.3636912

UNESCO. (2023). *Guidance for generative AI in education and research.* https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

Verhelst, E., Janssens, R., Demeester, T., & Belpaeme, T. (2024). Adaptive second language tutoring using generative AI and a social robot. *Proceedings of the Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 1080-1084). Association for Computing Machinery. https://doi.org/10.1145/3610978.3640559

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum. https://www.researchgate.net/profile/Allyson-Hadwin/publication/247664651_Studying_as_Self-Regulated_Learning/links/663d16347091b94e930f18f1/Studying-as-Self-Regulated-Learning.pdf

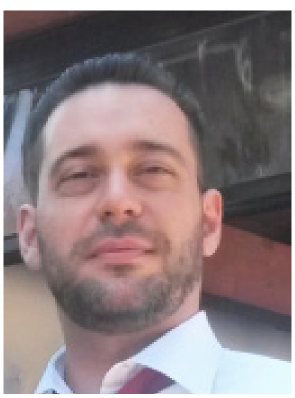
Woo, D. J., Wang, D., Guo, K., & Susanto, H. (2024). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. *Education and Information Technologies*, *29*, 24963–24990. https://doi.org/10.1007/s10639-024-12819-4

Yang, T.-C., Hsu, Y.-C., & Wu, J.-Y. (2025). The effectiveness of ChatGPT in assisting high school students in programming learning: Evidence from a quasi-experimental research. *Interactive Learning Environments*, *33*(6), 3726–3743. https://doi.org/10.1080/10494820.2025.2450659

Zhao, G., Yang, L., Hu, B., & Wang, J. (2025). A generative artificial intelligence (AI)-based human–computer collaborative programming learning method to improve computational thinking, learning attitudes, and learning achievement. *Journal of Educational Computing Research*, *63*(5), 1059–1087. https://doi.org/10.1177/07356331251336154

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

AUTHORS



Luca Addiucci is a PhD student in Computer Science Engineering at Sapienza University of Rome and a secondary school teacher of Mathematics and Physics. His research focuses on artificial intelligence in education, with particular attention to prompt engineering, AI literacy, and the use of large language models to support teaching and learning processes. He is currently involved in research on the integration of generative AI in secondary education and on the development of intelligent tools for educational content generation and assessment.

Marco Temperini is an Associate Professor of Information Processing Systems at Sapienza University of Rome. He holds a degree in Mathematics (1986) and a PhD in Computer Science (1992) from the same institution. His research focuses on Technology-Enhanced Learning, including student and teacher modeling, adaptive e-learning, artificial intelligence, data mining for personalized learning, social and collaborative learning, game-based learning, and automated and peer assessment. He has participated in and coordinated several international research projects and has served on program committees, editorial boards, and in organizational roles for major international conferences and journals in the fields of educational technology and artificial intelligence.